PREDICTING SURGICAL OUTCOME IN PATIENTS WITH ADVANCED OVARIAN CANCER

A thesis submitted to The University of Manchester for the degree of Doctor of Medicine
In the Faculty of Biology, Medicine and Health

2022

AMY HAWARDEN

Division of Cancer Sciences
School of Medical Sciences

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACE-27	Adult comorbidity evaluation-27 index
AFP	Alpha Fetoprotein
AI	Artificial Intelligence
ANN	Artificial neural network
ASA	American Society of Anaesthesiologists
AUC	Area under curve
BER	Base excision repair
BMI	Body mass index
β-HCG	Beta-human chorionic gonadotrophin
BRCA	BReast CAncer gene
CA 125	Cancer antigen 125
CA 72-4	Cancer antigen 72-4
CA 19-9	Cancer antigen 19-9
CEA	Carcinoembryonic antigen
CHARMS	Checklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies
CI	Confidence interval
COCP	Combined oral contraceptive pill
COSHH	Control of Substances Hazardous to Health
CRUK	Cancer Research UK
CT	Computerised tomography
DNA	Deoxyribonucleic acid
ECOG-PS	Eastern Cooperative Oncology Group performance status
EORTC	European Organisation for Research and Treatment of Cancer
EOC	Epithelial ovarian cancer
EPP	Event per predictor
FIGO	International Federation of Gynaecology and Obstetrics
Hb	Haemoglobin
hCG	Human chorionic gonadotropin
HE4	Human epididymis protein 4
HGSOC	High grade serous ovarian cancer
HIPEC	Hyperthermic intraperitoneal chemotherapy
HR	Homologous recombination
HRC	Homologous recombination competent
HRD	Homologous recombination defective
HRH	Homologous recombination heterogenous
HRT	Hormone replacement therapy
ICON	International Collaborative Ovarian Neoplasm

IDS	Interval debulking surgery
IHC	Immunohistochemistry
IMD	Indices of multiple deprivation
ITT	Intention to treat
IUD	Intra uterine device
LDH	Lactate dehydrogenase
LGSOC	Low grade serous ovarian cancer
MDT	Multidisciplinary team
MMR	Mismatch repair
MOCHR	Manchester ovarian cancer homologous recombination
MHGSOC	Manchester high grade serous ovarian cancer
MRC CTU	Medical Research Council Clinical Trials Unit
MRI	Magnetic resonance imaging
NACT	Neo-adjuvant chemotherapy
NCIN	National Cancer Intelligence Network
NER	Nucleotide excision repair
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
NLR	Neutrophil-to-lymphocyte ratio
NPV	Negative predictive value
OC	Ovarian cancer
OS	Overall survival
PARP	Poly ADP ribose polymerase
PCI	Peritoneal cancer index
PCOS	Polycystic ovarian syndrome
PDS	Primary debulking surgery
PET CT	Positron emission tomography/computed tomography
PFS	Progression free survival
PICOS	Population, Intervention, Comparator, Outcome, Study design
PID	Pelvic inflammatory disease
PLCO	The American Prostate, lung, colorectal and Ovarian cancer screening
PPV	Positive predictive value
PROBAST	Prediction model risk of Bias Assessment Tool
QUIPS	Quality in prognostic studies tool
RCOG	Royal College of Obstetrics and Gynaecology
RCT	Randomised controlled Trial
RMI	Risk of malignancy index
RNA	Ribonucleic acid
ROB	Risk of bias
ROBINS-1	Risk of Bias in Nonrandomised Studies of Interventions tool

ROC	Receiver operator characteristic
STIC	Serous tubular intra-epithelial carcinomas
TCGA	The cancer genome atlas
TVUSS	Transvaginal ultrasound scan
UKCTOCS	UK Collaborative Trial of Ovarian Cancer Screening
UK	United Kingdom
VUS	Variant of uncertain significance
WCC	White cell count
WHO	World Health Organisation

Abstract

Introduction: Epithelial ovarian cancer affects more than 7,000 women each year in the UK, and approximately 70% of patients present with advanced disease. Treatment involves one of the following options: initial primary debulking surgery (PDS) followed by chemotherapy; initial chemotherapy followed by interval debulking surgery (IDS); or palliative management. Removing all visible disease at the time of surgery (complete cytoreduction) is the most important prognostic marker in these patients, and it is important that the treatment option that is most likely to result in complete cytoreduction is chosen. Complete cytoreduction at the time of PDS holds a survival advantage over IDS. Failing to achieve complete cytoreduction equates to reduced overall survival, increased morbidity and delayed chemotherapy start. Under the current decision-making process 9 - 67% of patients suffer residual disease, which highlights a significant area for improvement. Many clinical, biochemical, genomic and radiological predictors have been linked with surgical outcome. Despite this, there is no accepted tool or guideline to aid clinicians in this decision-making process. This thesis aims to review currently published models predicting surgical outcome, externally validate a pre-existing model, explore new predictors and create a new prognostic model combining all available predictors.

Methods: A systematic review of all published multimodal prognostic models predicting outcome of PDS in patients with stage II-IV epithelial ovarian cancer was performed. Data extraction was performed using the checklist for critical appraisal method (CHARMS) and risk of bias was assessed using the prediction model study assessment tool (PROBAST). The external validation of a three-protein signature was performed via immunohistochemistry (IHC) on a validation cohort from the ICON5 trial. The association between homologous recombination (HR) and surgical outcome was assessed in two separate cohorts. The cancer genome atlas (TCGA) cohort had HR status assessed using an established gene panel, and the Manchester database cohort via a functional assay. Both logistic regression and random forest models were developed combining multiple predictors including operating surgeon to predict surgical outcome at the time of PDS.

Results: The systematic review included 26 publications describing 27 prognostic models. Predictors included clinical, biochemical, genomic and radiological features. All but one model was developed by logistic regression. Validated performance measured by AUC ranged between 0.50 and 0.89, with low levels of external validation. The majority of models showed high risk of risk of bias. The three-protein signature was validated via IHC on a cohort of 238 HGSOC tumour samples. Staining intensity scores from each protein were combined to create a combined prognostic model. Validation failed, with AUC dropping from 0.866 in the original cohort to 0.593 in the validation cohort. Two patient cohorts were used to assess association between HR status and surgical outcome. The TCGA cohort (n=258) assessed via a 14 gene panel demonstrated association between HR status and surgical outcome (p=0.033). The Manchester cohort (n=38) assessed via a functional assay did not show any association (p=0.5205). Finally, the developed prognostic prediction models developed on a cohort of stage III-IV HGSOC patients (n=100) incorporated 18 predictors. When internally validated, they performed with AUC values of 0.688 and 0.734 for logistic regression and Random Forest models, respectively.

Conclusions: Models incorporating single modalities rarely show accurate prediction upon external validation and there are currently no prognostic models validated successfully enough for use in clinical practice. Models combining multiple predictors including surgeon heterogeneity show the most promise, and further validation would be the next step to progress these models with a view to apply them in clinical use.

Declaration

This is a declaration to state that that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

Statement regarding published work

Data in chapter 5 has been accepted for publication pending corrections in PLOS ONE- see appendix A for paper in full. This paper was co-authored by Marcus Price, Bryn Russell, Godfrey Wilson, Laura Farrelly, Andrew Embleton, Mahesh Parmar and Richard J Edmondson. I am the first author and performed all work described in the paper. The project was directly supervised by Professor Richard Edmondson.

Acknowledgements

First and foremost, I would like to thank my supervisors, Professor Richard Edmondson and Professor Emma Crosbie, for giving me the opportunity to pursue this MD. Your feedback, guidance and support have been invaluable throughout the project, especially so in recent months. I would also like to thank Dr Godfrey Wilson for his invaluable teaching of histology.

My MD experience would not have been the same without my wonderful lab group. A particular thank you to Marcus Price for his endless patience and teaching in the lab and for volunteering to score a mountain of slides. A special mention to Bryn Russell and Tom Walker, without whom my knowledge of statistics would still be sorely lacking. And finally thank you to Caitlin Waddell, for always being there to offer me an ear and cheer me up when things were tough.

Finally, I am forever grateful to my family and friends. To my wife Raj for always supporting and encouraging me with endless love and champagne. To my mum, Annette, dad, Geoff and brother Andrew, thank you for always being my biggest fans. A special thank you to my sister Lucy, without your support and guidance this thesis would never have been finished, and I can never thank you enough.

1 Introduction

1.1 Ovarian cancer

1.1.1 Background

Ovarian cancer accounts for an estimated 239,000 new cases and 152,000 deaths annually worldwide (Reid et al., 2017). It is the cause of more deaths in women in the developed world than any other gynaecological cancer (Colombo et al., 2019), with over 7,000 women diagnosed per year in the United Kingdom (UK) (CRUK, 2018). These statistics highlight that ovarian cancer is a significant cause of both morbidity and mortality to women globally.

The incidence of ovarian cancer varies with geographical area, the highest age-adjusted incidence being observed in the developed world. In North America and Europe, rates exceed 8 per 100,000, with lower rates seen in Asia and Africa (Webb & Jordan, 2017).

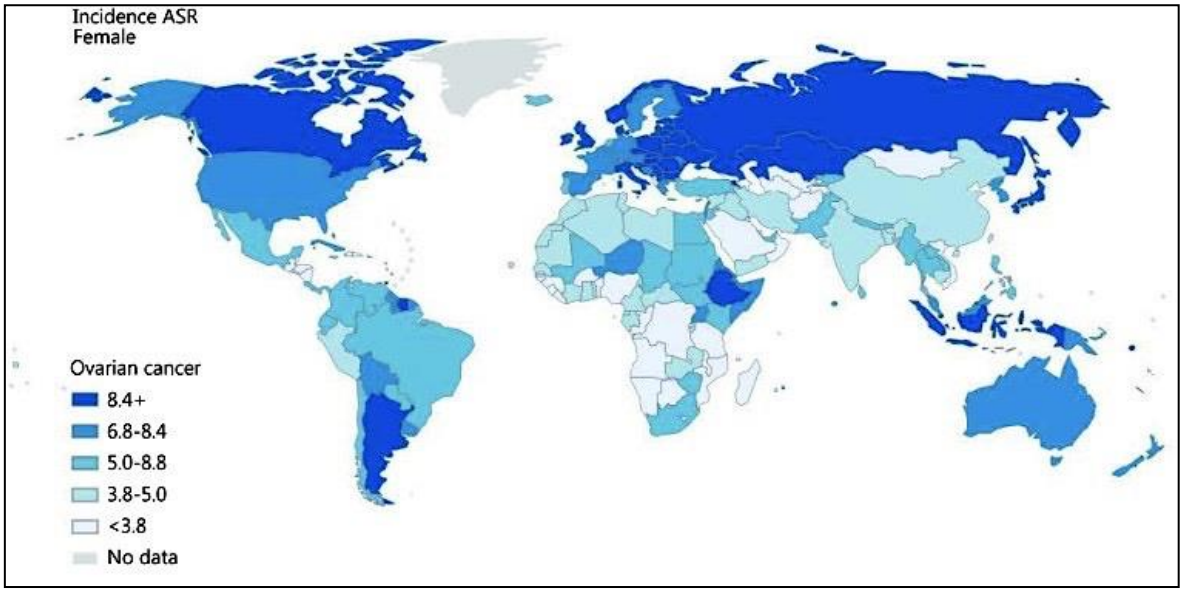


Figure 1.1 *Map of the world demonstrating the wide geographical variation of OC incidence.*
Figure source (Reid et al., 2017).

A woman's lifetime risk of ovarian cancer in the UK stands between 1 in 50 and 1 in 75 with risk increasing with age. The majority of cases present in women over 55 years of age, peaking between 75 – 79 years (CRUK, 2018; Reid et al., 2017). Survival of women with ovarian cancer has improved only slightly since the 1980s (Lisio et al., 2019). The overall ten year survival for

all histological subtypes and stages of ovarian cancer currently stands at just 35% (CRUK, 2018). Survival rates are significantly decreased by a more advanced stage at the time of presentation (CRUK, 2018).

Table 1.1 *Breakdown of the five-year survival rates by FIGO stage for ovarian cancer*

Stage at time of presentation	Five-year survival rate
Stage I	95%
Stage II	70%
Stage III	25%
Stage IV	15%

(CRUK, 2018), see table 1.5 for breakdown of FIGO stage

As approximately 65 – 70% of ovarian cancer cases present at FIGO stage III or IV (CRUK, 2018), Table **1.5**, the majority of patients face very poor survival rates.

The most significant risk factor for developing ovarian cancer is family history of the disease. Women with a first-degree relative with a history of ovarian cancer have an increased risk of 3- to 7-fold. Mutations in the *BRCA1* and *BRCA2* genes account for approximately 10 – 15% of all ovarian cancer cases and represent a lifetime risk of 44% and 27%, respectively. This mutation is mainly associated with high grade serous ovarian cancer (HGSOC). Other genetic associations with ovarian cancer include mutations in *BRIP1* and *RAD51D* genes (Reid et al., 2017) as well as mismatch repair genes (MMR) as is seen in Lynch syndrome (Fotopoulou et al., 2017).

Nulliparity, infertility (Reid et al., 2017), polycystic ovarian syndrome (PCOS) (Schildkraut et al., 1996), endometriosis (Sayasneh et al., 2011), pelvic inflammatory disease (PID) (Lin et al., 2011), first birth after 35 years, early menarche and late menopause (Fotopoulou et al., 2017) all carry an increased risk of developing ovarian cancer. There is conflicting evidence surrounding an increase risk with the use of an intrauterine device (IUD) (Huang et al., 2015; Tworoger et al., 2007), hormone replacement therapy (HRT) (Danforth et al., 2007), obesity and poor diet (Reid et al., 2017). The combined oral contraceptive pill (COCP) is protective, with this effect increasing with longer duration of use (Beral et al., 2008). Pregnancy,

sterilisation/tubal ligation and hysterectomy are also recognised protective factors (Fotopoulou et al., 2017).

1.1.2 Histopathology

Ovarian cancer is represented by multiple distinct histological entities, with tumours most commonly originating histologically from epithelial, stromal or germ cells. 90% of malignant tumours originate from epithelial cells (Reid et al., 2017). Epithelial ovarian cancer (EOC) is in itself a very heterogenous disease, made up of several histological subtypes all with differing origins, namely: high grade serous ovarian cancer (HGSOC), low grade serous ovarian cancer (LGSOC), mucinous, clear cell and endometrioid (Reid et al., 2017).

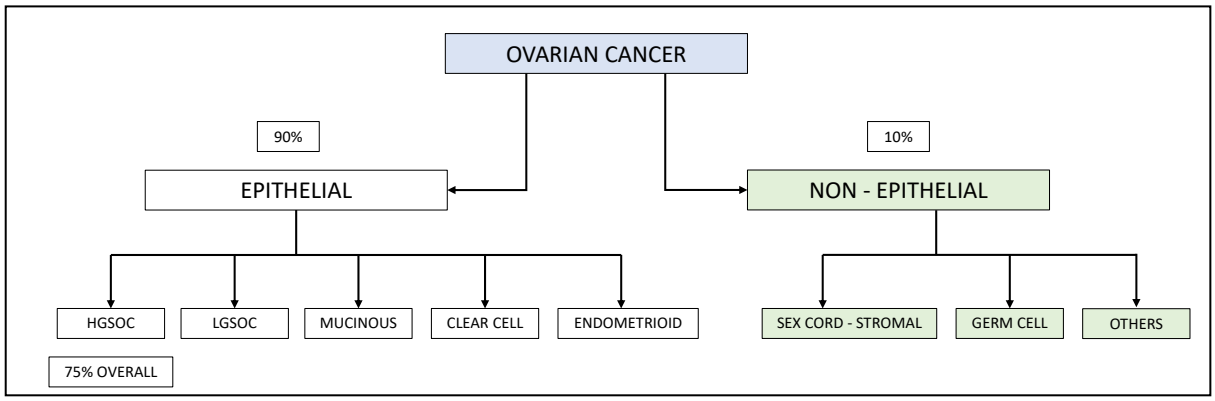


Figure 1.2 *Breakdown of the histological subtypes of ovarian cancer*

Demonstrating its wide histological heterogeneity, with epithelial cells being the most common cellular origin and HGSOC accounting for 75% of all new cases overall.

In an attempt to more clearly categorise epithelial ovarian cancer, and in turn to study the disease from a molecular and genetic basis, two distinct groups have been introduced and accepted by the World Health Organisation (WHO). The histological subtypes are grouped based upon their clinical behaviour alongside frequently observed genetic mutations (Kurman & Shih Ie, 2010). Type one tumours describe those that tend to grow locally, metastasize late, and behave in a more indolent fashion. Type two tumours describe a more aggressive tumour type that often present at a more advanced stage (Terada et al., 2016). Type two tumours can be characterised by P53 mutations and often display genomic instability due to defects in pathways in deoxyribonucleic acid (DNA) repair such as homologous recombination (HR) (Kurman & Shih Ie, 2016; Reid et al., 2017).

Table 1.2 *Breakdown of OC histological subtype by group*

	Histological subtype	Associated mutations
Type one tumour	LGSOC, endometrioid, clear cell, clear cell	KRAS, ARID1A, PIK3CA, PTEN, BRAF
Type two tumour	HGSOC, carcinosarcoma, undifferentiated carcinomas	TP53, defects in HR

Adapted from (Kurman & Shih Ie, 2010, 2011)

HGSOC is the most prevalent histological subtype of ovarian cancer, with the highest mortality rate. It accounts for 75% of all ovarian cancer diagnoses (Reid et al., 2017). Histologically, HGSOC can be defined by one or more of five main histological features (Table 1.3/Figure 1.3).

Table 1.3 *Breakdown of the heterogeneity of HGSOC histopathological architecture*

Histological feature of HGSOC	Corresponding image in figure 1.3
Solid architecture	A
Glandular architecture 'slit like' fenestrations	B
Papillary architecture	C
Solid architecture with 'geographical' necrosis	D
Solid architecture with tumour infiltrating lymphocytes	E

Adapted from (Reid et al., 2017)

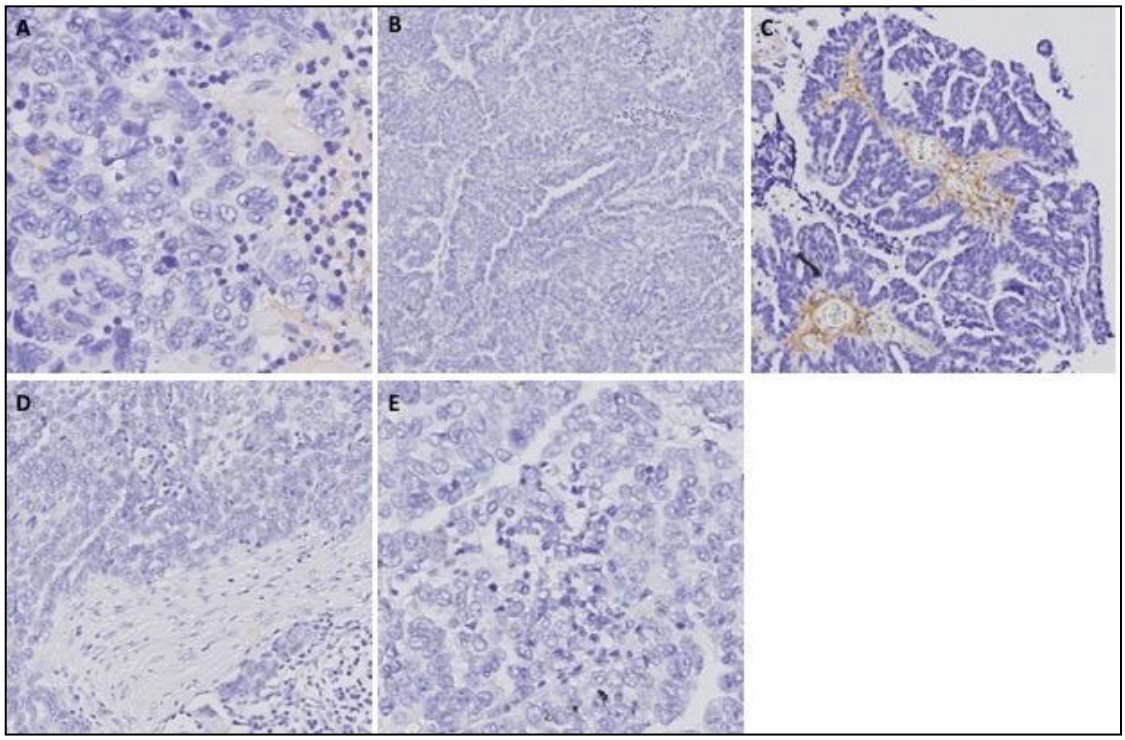


Figure 1.3 *Corresponding images to illustrate table 1.3*

Original magnification A x 20, B x5, C x10, D x5, E x20

From a cytological perspective, HGSOC can be characterised by high-grade nuclear atypia, and can be differentiated from other subtypes of ovarian cancer by the positive staining for several immunological markers, such as P53, WT1, P16, Ki-67, CK7, and PAX 8 (Reid et al., 2017)

1.1.3 Natural history

Most epithelial ovarian cancers have disease localised to the ovary in the early stages. It is therefore unsurprising that the disease was originally assumed to be ovarian in origin. More recently this theory has been questioned. Histological sub types such as endometrioid and clear cell cancers develop upon a background of endometriosis, which would suggest that the

origin is in fact endometrial (Terada et al., 2016), whilst mucinous tumours often derive from metastatic intestinal tumours (Riopel et al., 1999).

For many years HGSOCs did not have a well understood pre-cancerous lesion (Webb & Jordan, 2017). It was originally believed that HGSOC originated from the surface epithelium of the ovary (Lengyel, 2010). The ovarian epithelial cells originate embryonically from the coelomic mesoderm and are closely related to the peritoneum covering the peritoneal cavity. As early as 1971, Fathalla described the 'incessant ovulation' theory; continued ovulation throughout a woman's reproductive life resulted in the continual damage of surface epithelial cells. This cycle of damage and repair repeated until the damage became irreparable, resulting in the formation of a malignant process (Fathalla, 1971). According to this theory, the more ovulatory cycles a woman was exposed to, the greater her risk of developing HGSOC. This theory supported the protective nature of the COCP and multiple pregnancies and therefore maintained widespread acceptance for many years. However, it did not explain the fact that histologically, HGSOC closely resembles tissues derived of Mullerian duct origin, and raises the question how ovarian cells of coelomic origin could have differentiated as such.

An alternative argument was proposed; that HGSOC originated from cells of Mullerian duct origin found in structures such as the fallopian tubes (Dubeau, 1999; Piek et al., 2001). This explained the origin of primary fallopian tube and primary peritoneal cancers, both of which arise with no ovarian involvement but behave as, and are treated in parallel with, HGSOC (Dubeau, 1999).

There is now widespread acceptance for the presence of HGSOC pre-cursors known as serous tubular intra-epithelial carcinomas (STIC lesions), first described in the fallopian tubes of patients with inherited BRCA mutations (Kuhn et al., 2012). These small dysplastic lesions made up of secretory cells are found in the distal ciliated end of the fallopian tubes and have led to the belief that HGSOC originates as STIC lesions in the fallopian tubes. From the fallopian tube epithelium, the STIC lesions appear to seed the ovary and progress to spread throughout the peritoneal cavity (Reade et al., 2014).

1.1.4 Screening programmes in ovarian cancer

There are currently no successful screening programmes for ovarian cancer. Both the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) and The American Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening have undertaken large randomised controlled trials (RCTs) to establish the viability of screening programmes. UKCTOCS randomised 202,000 women to either observation alone or screening, however showed no reduction in mortality in the screened women in their primary analysis (Jacobs et al., 2016). PLCO screened asymptomatic postmenopausal women, and despite 13 years follow up did not show a reduction in mortality in their screened group (Buys et al., 2011).

1.1.5 Dissemination

In many cases, HGSOC does not utilise blood or lymphatics for its spread, instead it either undergoes local spread to organs in its proximity, or through the detachment of cells from the primary tumour migrating throughout the peritoneal cavity and to the surface of the visceral organs within this cavity, the transcoelomic route. Although HGSOC can metastasise to any organ within the peritoneal cavity, it shows a particular affinity for the omentum (Lengyel, 2010). Generally, the spread of HGSOC is made up of surface lesions of distal organs, and often stays within the peritoneal cavity. Extraperitoneal spread and deeper organ involvement is more rarely seen (Lengyel, 2010). These lesions can be thought of as disseminated primary disease, rather than true metastasis, which may explain the role for surgery in advanced disease.

1.1.6. Symptoms of HGSOC

Presentation with advanced disease has frequently been attributed to a lack of symptoms. HGSOC is often termed ‘the silent killer’ or described as ‘insidious’ in nature (Jasen, 2009). It is now recognised that symptoms do occur in all stages, with as few as 5 – 10% of women being asymptomatic at the time of diagnosis (Bankhead et al., 2005). A large case control study in England reported 85% of women were suffering symptoms that had been reported to their GP up to several months before a diagnosis was achieved (Bankhead et al., 2005).

This would suggest that education of common red flag symptoms for both patients and physicians would be of benefit.

Unfortunately, even the most commonly reported symptoms of HGSOC such as abdominal distension, loss of appetite and urinary frequency do not display high positive predictive values (PPV) for a diagnosis (Bankhead et al., 2005).

Advice regarding concerning symptoms varies greatly throughout guidelines. A recent systematic review assessing the variation in the initial assessment and investigation for ovarian cancer in symptomatic women described that in eighteen different guidelines, between four and fourteen symptoms were described as cause for concern, however only one symptom (abdominal distension) was consistent across all (Funston et al., 2019).

The National Institute for Health and Clinical Excellence (NICE) guidelines advise further investigation for any women, but especially over the age of 50, presenting with the red flag symptoms listed in table 1.4. These symptoms are concerning if persistent, particularly more than twelve times per month (National Collaborating Centre for, 2011). Once a red flag has been identified, the patient is then commenced on the ovarian cancer investigation pathway (figure 1.5), and further treatment will be dependent upon individual results.

Table 1.4 *Red flag symptoms prompting further investigation by NICE*

Red flag symptoms
Persistent abdominal distension (bloating)
Feeling full (early satiety) and/or loss of appetite
Pelvic or abdominal pain
Increased urinary urgency and/or frequency
New onset Symptoms of IBS (if over 50 years)
Unexplained weight loss, fatigue or changes in bowel habit
Concerns of pelvic malignancy on physical examination

Adapted from (National Collaborating Centre for, 2011)

1.2 Initial referral and diagnostic pathway

Once a patient presents to a clinician with red flag symptoms suggestive of ovarian cancer, table 1.4, UK guidelines recommend an initial serum test for Cancer Antigen 125 (CA 125) levels. CA 125 is widely distributed in adult tissues and is well established as part of initial investigations (Jacobs et al., 2016). The upper level cut off of 35 IU/ml is based on the distribution of values in 99.7% of 888 healthy men and women (Bast et al., 1983). CA 125 is raised in over 80% of epithelial ovarian cancer cases (Bast et al., 1983) and has a diagnostic sensitivity of 81% and specificity of 75% when a cut off of 35 IU/ml is used (Jacobs et al., 1990). Other tumour markers such as human epididymis protein 4 (HE4), carcinoembryonic antigen (CEA), CDX2, cancer antigen 72-4 (CA72-4), cancer antigen 19-9 (CA 19-9), alphafetoprotein (AFP), lactate dehydrogenase (LDH) and beta-human chorionic gonadotrophin (β-HCG) have been suggested as biomarkers, and have shown varying promise in the initial assessment for the diagnosis of ovarian cancer. However, the evidence does not currently support their routine use (Abdel-Azeez et al., 2010; Huhtinen et al., 2009; Montagnana et al., 2009; Moore et al., 2008; Moore et al., 2009; National Collaborating Centre for, 2011; Nolen et al., 2010; Shah et al., 2009; Urban et al., 2012).

If the CA 125 level is <35 IU/ml then the patient should be reassessed by the clinician. If symptoms persist, other causes should be considered. If symptoms resolve without intervention, no further investigation is required (National Collaborating Centre for, 2011). If CA 125 result is $\geq$ 35 IU/ml, the patient should proceed through the pathway and undergo a transvaginal ultrasound scan of the abdomen and pelvis (TVUSS abdomen pelvis). Patients under the age of 40 years should have additional serum levels of AFP and hCG to identify any non-epithelial lesions (National Collaborating Centre for, 2011).

If initial physical examination identifies ascites and/or an pelvic or abdominal mass which is not obviously uterine fibroids, the patient should not undergo initial tests but instead be referred urgently to secondary care (National Collaborating Centre for, 2011).

Ultrasound is the preferred modality for first line investigation of ovarian cancer, with TVUSS being preferable over abdominal ultrasound scanning. If a large pelvic mass is present both

scans should be undertaken ("ACOG Practice Bulletin. Management of adnexal masses," 2007; Leibman et al., 1988).

Once a TVUSS has been undertaken, a risk of malignancy index (RMI) should be calculated. There are currently four published RMI scores (RMI I, RMI II, RMI III and RMI IV), however RMI I remains the most utilised, widely available and validated scoring system, with a sensitivity of 70% and specificity for malignancy of 90% with an RMI cut off of 250 (Geomini et al., 2009).

The RMI I combines three features: CA 125 level; the patient's menopausal status; and a score calculated from the TVUSS (Jacobs et al., 1990), figure 1.4. If the RMI score is <250, the patient can be treated in a secondary care setting. However, if the RMI is >250, the patient must be referred for discussion at a specialised tertiary level multidisciplinary team (MDT) meeting (National Collaborating Centre for, 2011).

If the RMI and clinical status suggest malignancy, further radiological staging by way of a computerised tomography (CT) scan of the abdomen and pelvis should be performed to establish the extent of the disease, with the thorax included if clinically indicated (National Collaborating Centre for, 2011). Neither magnetic resonance imaging (MRI), nor positron emission tomography/computed tomography (PET CT) are currently recommended as part of the standard diagnostic or staging pathway in the NHS (Fotopoulou et al., 2017).

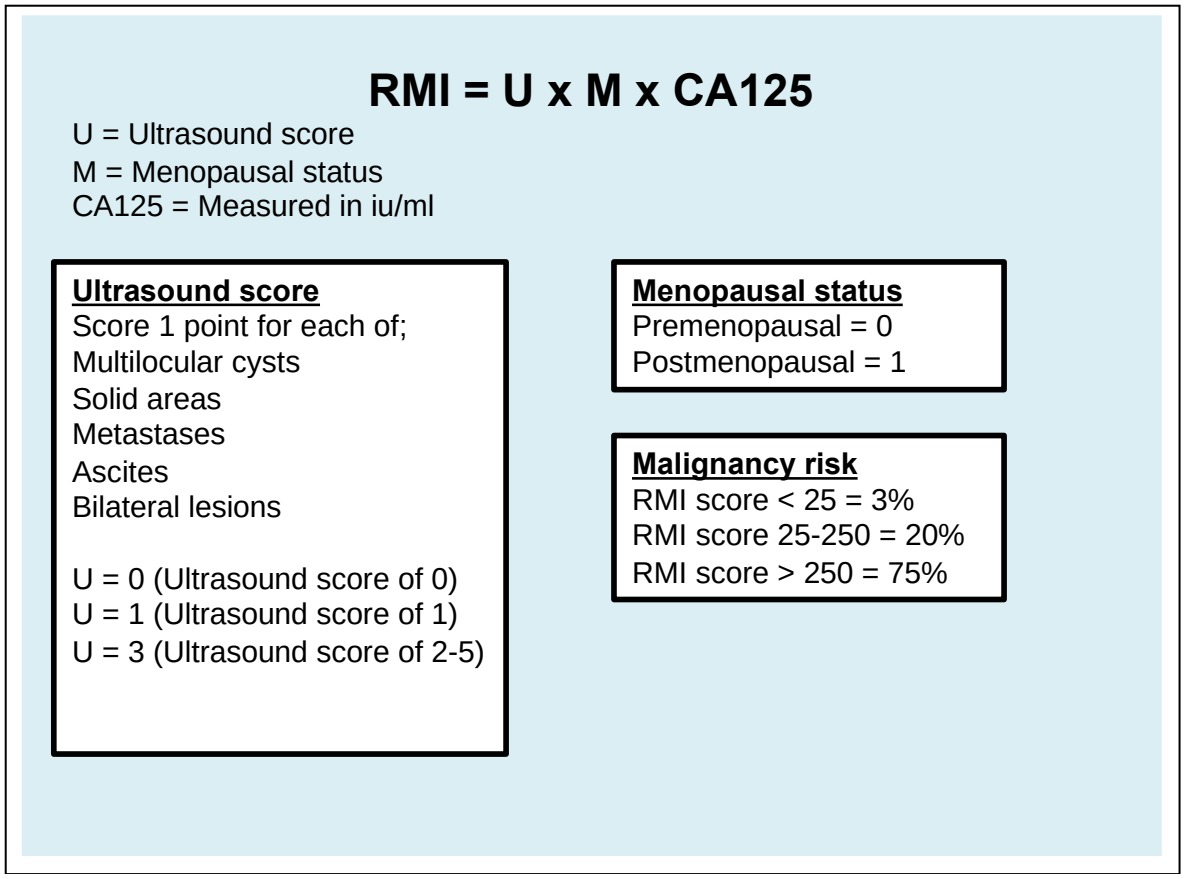


Figure 1.4 *Components and methods of calculating the risk of malignancy index*

Adapted from (Jacobs et al., 1990).

If malignancy is suspected following all preliminary investigations, the patient should proceed for treatment. Pre-operative staging is undertaken radiologically via CT scan, however histological diagnosis is not mandatory, and often not achieved prior to upfront surgical treatment (Fotopoulou et al., 2017). As such, staging is often not fully completed until surgical treatment has been undertaken.

Histological diagnosis should be confirmed before the commencement of chemotherapy, unless it would be inappropriate to do so. If a histological diagnosis is not achievable, chemotherapy can be commenced based upon a cytological diagnosis as an alternative (Fotopoulou et al., 2017).

Overall summary of treatment pathway can be seen in figure 1.5.

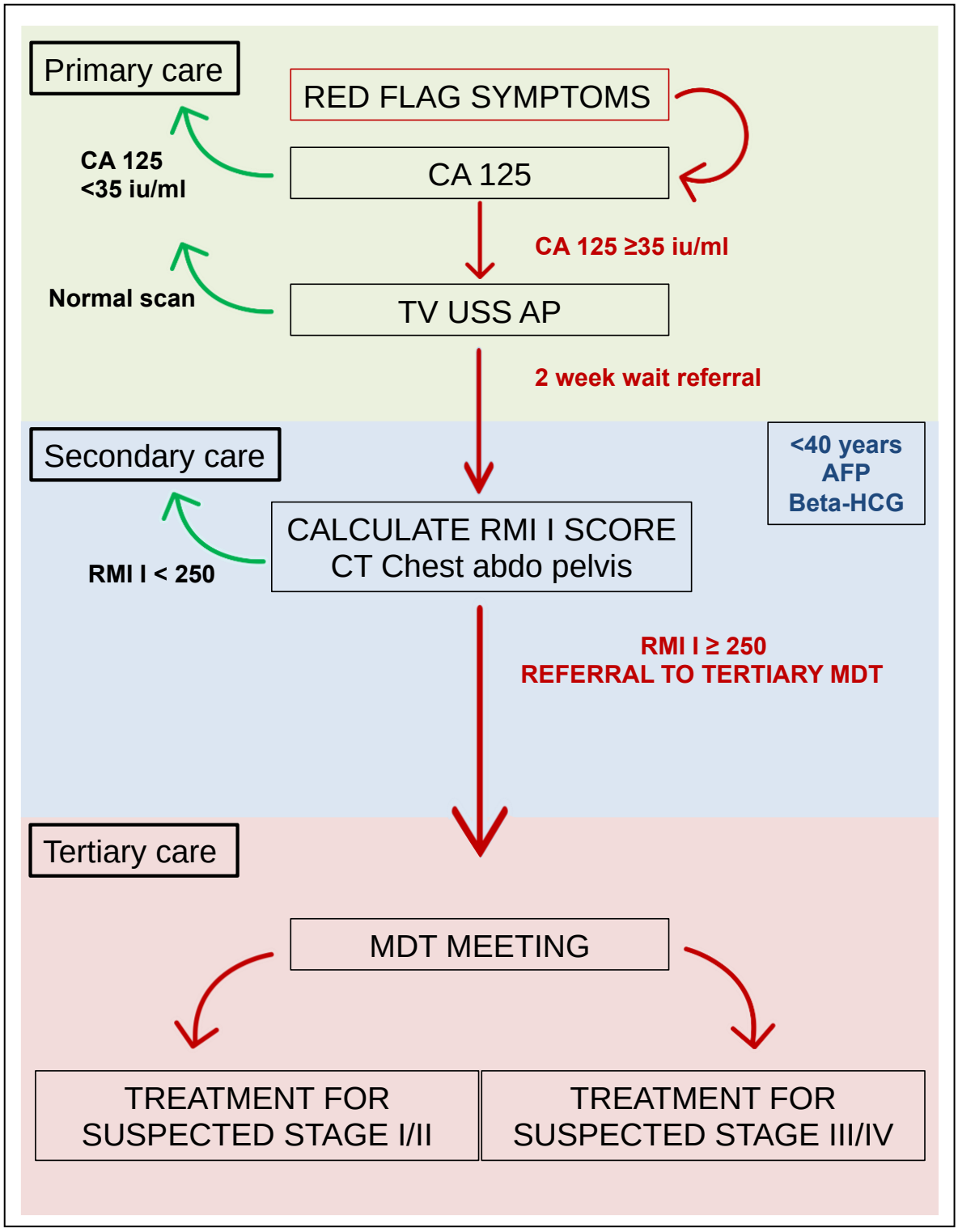


Figure 1.5 *Summary of the initial treatment pathways for investigation and diagnosis of ovarian cancer*

Adapted from (National Collaborating Centre for, 2011).

1.3 Staging of ovarian cancer

The final staging of ovarian cancer is achieved via a combination of radiological, histological, cytological and surgical investigation and management. For this reason, the final staging and diagnosis is often not known until treatment has been commenced. The International Federation of Gynaecology and Obstetrics (FIGO) staging system is routinely used in the UK, table 1.5. First published in 1973, the system was updated in 1988 and 2014 (Prat, 2015).

Table 1.5 *Summary of FIGO staging for ovarian cancer*

FIGO Stage	Disease distribution
IA	Limited to one ovary/fallopian tube
IB	Limited to both ovaries/fallopian tube
IC1/2	Surgical spill/capsule rupture
IC3	Malignant cells in ascites/peritoneal washings
II	Tumour involves one or both ovaries or fallopian tubes with pelvic extension (below pelvic brim/primary peritoneal cancer)
III	Tumour involves one or both ovaries or fallopian tubes or primary peritoneal cancer, with cytologically or histologically confirmed spread to the peritoneum outside of the pelvis
IIIA	+ microscopic extra-pelvic lymph node involvement
IIIB	+ macroscopic peritoneal metastasis beyond pelvis >2cm
IIIC	+ extension of tumour to capsule of liver or spleen, with no parenchymal involvement
IV	Distant metastasis excluding peritoneal metastasis
IVA	Pleural effusion with positive cytology
IVB	All other distant metastasis

Adapted from (Prat, 2015)

1.4 The role of the Multidisciplinary Team Meeting

Multidisciplinary team (MDT) working was first introduced in the NHS in 1995 following the publication of the Calman-Hine report (Haward, 2006), and is defined as a 'group of professionals from one or more clinical disciplines who together make decisions regarding recommended treatment of individual patients' (Scott et al., 2020). MDT meetings provide a platform for discussion of patients' results of radiological, histological and cytological investigations and decide upon best care with the benefit of all available information that might influence any decisions. MDT working is associated with increased survival rates, as well as increased recruitment into clinical trials (Fleissig et al., 2006).

The make-up of a typical gynaecological oncology MDT, alongside patient information usually available for discussion, is shown in figure 1.6. Despite the plethora of patient and tumour

factors available to aid this decision, the final treatment pathway is often decided based upon CT images and input from the radiologist (Scott et al., 2020). National guidance stresses that CT scans are not reliable markers for the prediction of surgical outcome and therefore should not be used in isolation (Fotopoulou et al., 2017). MDT discussion should be undertaken and documented prior to decision to operate for patients with suspected cancer in all but exceptional circumstances (Fotopoulou et al., 2017).

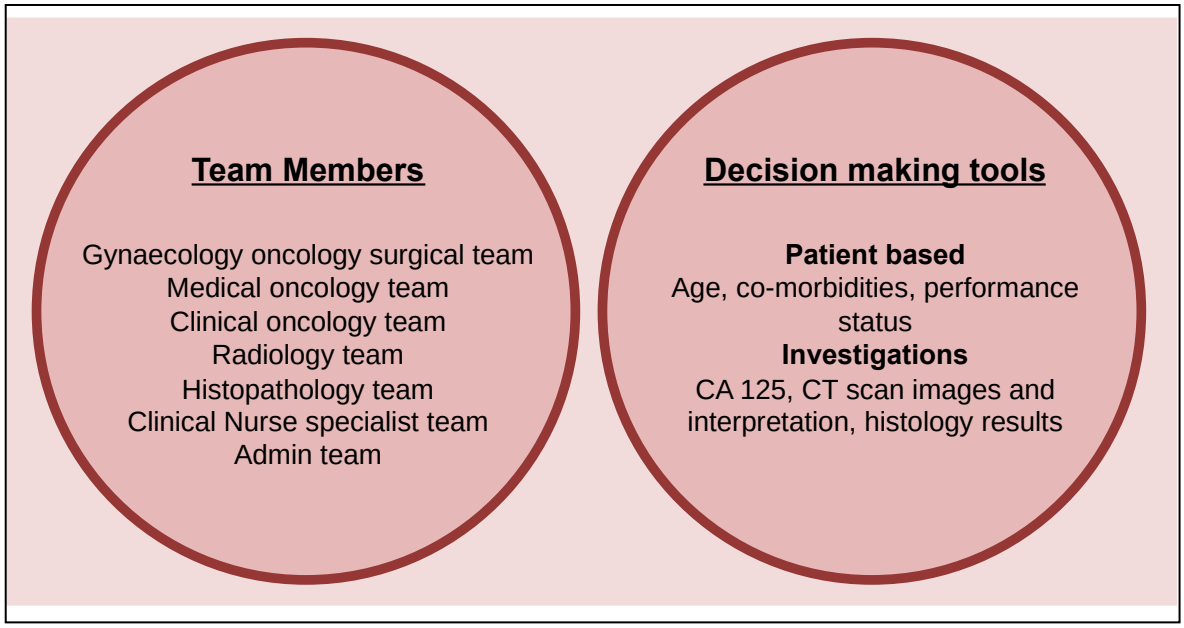


Figure 1.6 *Members of the MDT and available information at time of treatment decision.*

Left circle demonstrates commonly frequent members of the MDT and right circle the common patient and tumour information available to guide treatment decision making.

1.5 Treatment of HGSOC

The treatment for HGSOC is now a well-established combination of both surgical management, platinum-based and taxane chemotherapy, and targeted therapies. These treatments vary depending on the stage of the cancer at the time of presentation.

1.5.1 A historic perspective of surgical management in ovarian cancer

In 1685, the first oophorectomy to remove a necrosed ovary was described, the first time this procedure was used as a medical intervention (Bristow et al., 2016). In Glasgow in 1710, Surgeon Robert Houstoun described the case of a patient with a distended abdomen and his

subsequent removal of a large ovarian mass through an abdominal incision. However, Houstoun claimed to have only removed the mass, not the ovary itself. Inspired by Houstoun, surgeons Hunter and Bell, also in Glasgow, described the principles of oophorectomy for ovarian disease in the 1700s, although there is no evidence either actually performed the procedure (Bristow et al., 2016). In 1809, a 46-year-old mother of four, Jane Todd Crawford, sought help from American-born Glasgow-trained surgeon Ephraim McDowell. Jane was suffering with abdominal distension and was diagnosed with a large ovarian mass (Bristow et al., 2016). In a time before anaesthetic and much awareness of aseptic technique, McDowell performed what was known as an 'ovariotomy' to remove the mass on his own kitchen table with little more than a dose of opium for analgesia and sedation (Barnett, 2016).

At this time any surgery, not least intra-abdominal surgery, came with very high mortality rates. However, Lister's discovery of carbolic acid as an antiseptic, alongside the development of anaesthetic in the late 1800s (Robinson & Toledo, 2012; Worboys, 2013) opened many doors for progress in surgery, with the 100 years following known as 'the century of the surgeon' (*american cancer society: evolution of cancer treatments:surgery.*; Robinson & Toledo, 2012).

By the 1870s, surgical technique had developed and procedures entering the abdominal cavity were significantly less perilous. However, at this time, surgery was not the mainstay of treatment for ovarian cancer, as it was not believed to confer any survival benefit to the patient. Tait challenged this resistance, and championed the surgical approach proposing the 'exploratory laparotomy' as a method to determine if an ovarian mass was cancerous, an idea that was met with opposition.

This resistance to surgical management of ovarian cancer continued until the 1900s, when with the establishment of chemotherapy, Meig proposed that there may be benefits to removing as much tumour as possible in order to enhance the benefits of post-operative chemotherapy, a concept that became known as cytoreduction (Bristow et al., 2016).

The principles of cytoreductive surgery that we know today, to remove all residual disease, often with surrounding lymph nodes and other tissue, was originally developed by

Halsted(*american cancer society: evolution of cancer treatments:surgery.*)(*american cancer society: evolution of cancer treatments:surgery.*) with his work into the development of the radical mastectomy for breast cancer (Newmark, 2016). The transfer of these principles to the care of ovarian cancer was developed and by the 1970s, the established treatment for advanced ovarian cancer included a total abdominal hysterectomy, bilateral salpingo-ophrectomy and omentectomy, and additional procedures such as bowel surgery to achieve cytoreduction (surgical resection). However, this treatment combination was not universally accepted (FA., 1949; Munnell, 1968).

Radical and extensive surgery pre-chemotherapy to increase survival continued to be developed. In 1968, Munnell released the first study reporting the beneficial effects of cytoreduction and tumour volume in ovarian cancer patients (Munnell, 1968). This was echoed in 1975, when Griffiths published his landmark study which convincingly demonstrated an inverse relationship between residual tumour mass and patient survival (Bristow et al., 2016; Griffiths, 1975). This was further supported by Hunter's meta-analysis in 1992, showing that whilst removal of tumour mass is related to survival, this benefit is significantly maximised when followed by platinum-based chemotherapy (Hunter et al., 1992), something also demonstrated by Bristow in 2002 (Bristow et al., 2002). These landmark studies paved the way for the goal of surgical treatment for HGSOC to be to leave no disease behind at the time of surgical intervention.

Summary of development of surgical management over time can be seen in figure 1.7.

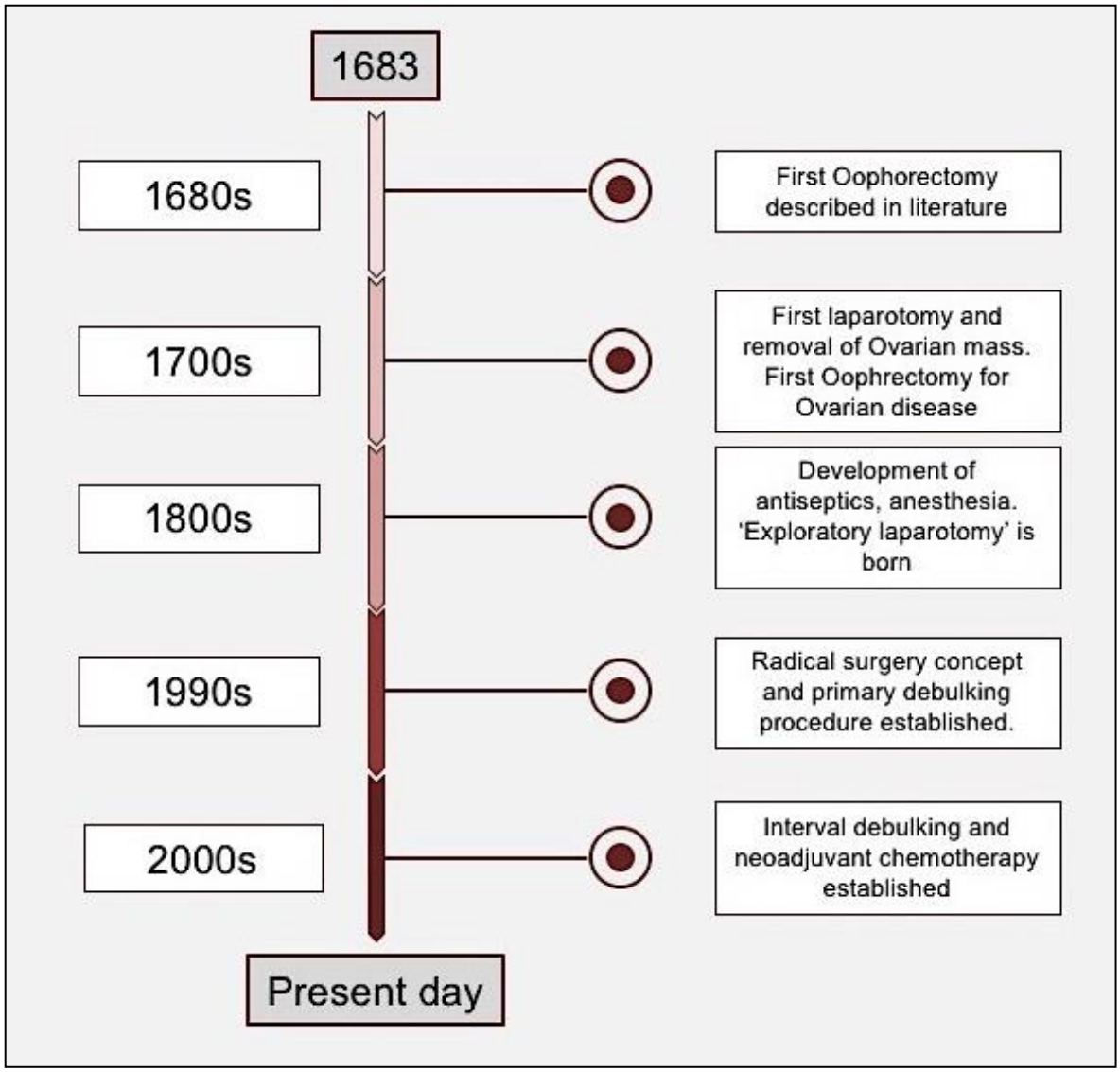


Figure 1.7 *Timeline of the development of the surgical management of ovarian cancer*

1.5.2 A historic perspective of medical management in ovarian cancer

Ovarian cancer was one of the first malignancies to be successfully treated with cytotoxic chemotherapy, reviewed in (Markman, 2008). Prior to 1970, alkylating agents, anthracyclines and antimetabolites such as melphalan, thiotepa and cyclophosphamide were the mainstay of treatment for ovarian cancer (Markman, 2003). It was during this time that the concept of multiple drugs being administered alongside each other and work synergistically originated (Markman, 2008).

The 1970s saw the development of platinum-based therapy, with the discovery of cisplatin (Markman, 2008). During the 1980s, carboplatin was developed and proved itself to be not

only equally effective to cisplatin, but with a more favourable side-effect profile (Alberts et al., 1992).

Alongside the development of carboplatin came the discovery of taxanes, with paclitaxel being the prototype. Early studies hoped that taxanes could play a role in women whose disease displayed platinum resistance (Markman, 2003). In 1996, a landmark trial reported cisplatin-paclitaxel combination therapy was superior to cisplatin-cyclophosphamide (McGuire et al., 1996). Further trials have confirmed this benefit with carboplatin-paclitaxel ("Paclitaxel plus carboplatin versus standard chemotherapy with either single-agent carboplatin or cyclophosphamide, doxorubicin, and cisplatin in women with ovarian cancer: the ICON3 randomised trial," 2002). This combination is still used in the majority of patients with HGSOC today.

More recent years have seen the development of the administration of carboplatin via the intraperitoneal route (IP) as well as hyperthermic intraperitoneal chemotherapy (HIPEC), which shows promise alongside surgical management for disease affecting the peritoneal cavity (Riggs et al., 2020). The RCT GOG 172 (Armstrong et al., 2006) led to the recommendation that IP chemotherapy can be considered in the treatment of FIGO stage II-III epithelial ovarian cancer following surgical treatment in some instances (Ledermann et al., 2013). Despite promising results, IP chemotherapy and HIPEC are still considered experimental treatments by many, and further RCTs are needed. Their use is exceptional in the UK.

Most recently, cancer treatments have become more focussed towards the development of targeted therapies. These therapies aim to specifically target the pathways abnormal in a particular cancer cell, and in turn spare the healthy cells from the cytotoxic effects. One of the most developed targeted therapies for HGSOCs are poly ADP ribose polymerase (PARP) inhibitors. The mode of action of PARP inhibitors relies upon the fact that approximately 50% of HGSOCs have a fault in the DNA repair pathway, homologous recombination (HR). Half of these tumours will present in patients with BRCA 1 and BRCA 2 mutations. As these tumours cannot repair via HR, and the drug inhibits their alternative method of repair, tumour cell death results.

Several RCTs have demonstrated the efficacy of PARP inhibitors as first-line maintenance in stage III/IV HGSOC in patients with germline or somatic BRCA mutations (Mahmood et al., 2020; Moore et al., 2018). The SOLO-1 trial led to the recommendation that all patients presenting with HGSOC undergo germline and somatic BRCA testing, with maintenance Olaparib recommended for those whose tumour harboured mutant BRCA1/2 ((NICE), August 2019).

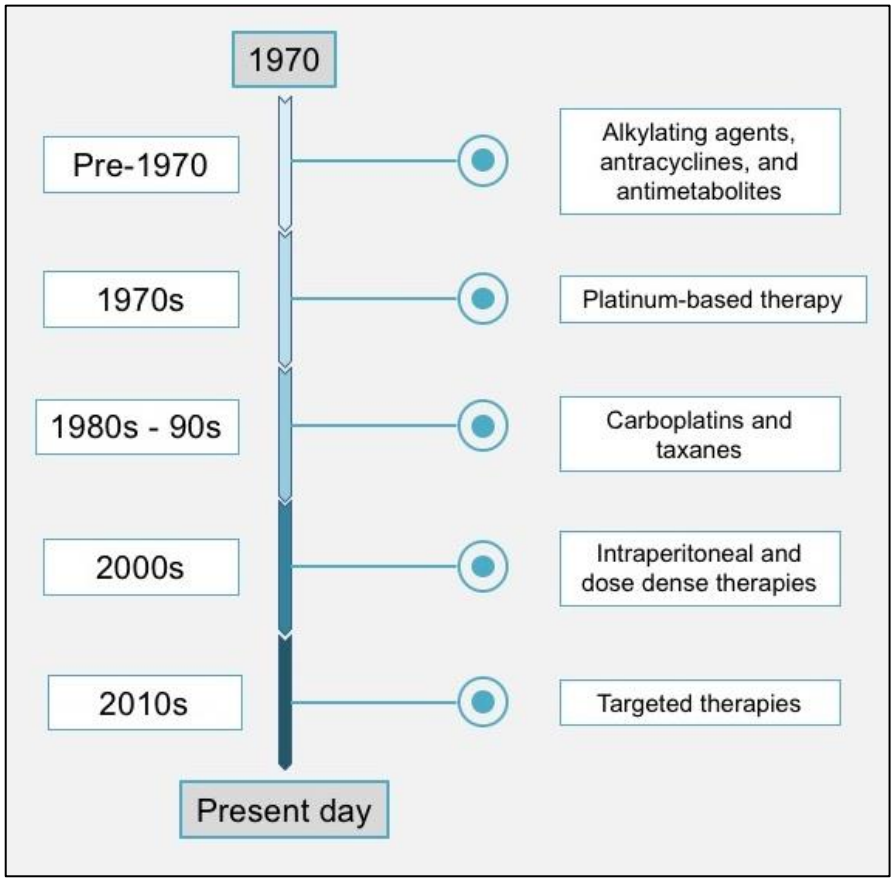


Figure 1.8 *Timeline of the development of chemotherapy and its use in ovarian cancer*

Adapted from (Reid et al., 2017).

These studies lead us to the standard of care for the treatment stage II-IV HGSOC recognised today; cytoreductive (disease debulking) surgery followed by chemotherapy (Wright et al., 2016). Observational studies spanning over the last 50 years have consistently confirmed that overall survival in stage II-IV HGSOC is inversely proportional to residual disease at the end of a surgical procedure (Bristow et al., 2002; Griffiths, 1975; Hunter et al., 1992; Winter et al.,

2008). Each 10% increase in complete surgical cytoreduction, confers an estimated increase of 5.5% in median survival (Bristow et al., 2002).

The most desirable outcome from a cytoreductive procedure is to leave no visual disease, as this translates to maximum survival (Horowitz et al., 2015). Following cytoreductive surgery for HGSOC, the lead surgeon records the outcome of the surgery. This outcome is recorded as one of the following; complete cytoreduction (no visible/macroscopic disease remains), optimal cytoreduction (visible disease remains <1cm), or suboptimal cytoreduction (visible disease remains >1cm) (Horowitz et al., 2015; Vergote et al., 2010; Wright et al., 2016), table 1.6. This decision is made by visual observation, rather than objective measurement or photographic evidence, leading to variation in assessment of residual tumour size between clinicians (Chi et al., 2007).

Table 1.6 *Outcome definitions depending on residual disease at the time of cytoreductive procedure*

Cytoreductive outcome	Remaining disease
Complete cytoreduction	No macroscopic visible disease
Optimal cytoreduction	Macroscopic disease <1cm
Suboptimal cytoreduction	Macroscopic disease ≥1cm

The outcome of surgery depends on many factors, including patient selection, distribution of tumour deposits and surgical expertise and practice (Schorge et al., 2010). There is often debate regarding whether achieving a complete cytoreduction is reliant mainly on tumour biology or surgical technique (Eisenkop et al., 2003). The clear survival advantage of complete cytoreduction has led to increasingly aggressive surgical techniques in many units, with the rates of bowel resections, splenectomies, liver resections and diaphragmic stripping increasing significantly over the years (Jones et al., 2018). This increase in trends towards extensive surgery has led to concerns as to whether the associated morbidity is warranted. Increased morbidity in patients undergoing extensive procedures is reported to be as high as 54.9%, compared with 22.8% in patients undergoing standard cytoreductive surgery (Rausei et al., 2019). Despite this, overall survival is increased in patients achieving complete cytoreduction, despite their increased morbidity (Rausei et al., 2019).

1.5.3 Treatment of stage I HGSOC

Patients with disease confined to the ovaries, see table 1.5., should undergo a full staging procedure, including a midline laparotomy to allow thorough assessment of the abdomen and pelvis, a total abdominal hysterectomy, bilateral salpingo-oophrectomy and infra-colic omentectomy, random biopsies of the pelvic and abdominal peritoneum and a retroperitoneal lymph node assessment (National Collaborating Centre for, 2011). Women with stage I disease that is deemed high risk (high grade or stage Ic) should be offered adjuvant chemotherapy consisting of six cycles of carboplatin (Colombo et al., 2003; National Collaborating Centre for, 2011).

Depending upon histological stage and subtype, up to 30% of the patients with apparently early epithelial ovarian cancer will be upstaged following full surgical staging (Fotopoulou et al., 2017). Despite this, recommendations suggest that in young women, fertility sparing surgery can be considered following thorough discussion with the patient regarding the risk of recurrence (Fotopoulou et al., 2017).

1.5.4 Treatment of stage II-IV HGSOC

Currently, cytoreductive surgery for stage II-IV HGSOC should be performed as a midline laparotomy, and include the removal of the uterus and cervix (total hysterectomy), both fallopian tubes and ovaries (bilateral salpingo-oophrectomy), removal of the infra-colic omentum (omentectomy), removal of any other visible disease that is safe to remove, including bowel related disease, peritoneal disease and upper abdominal disease, and a full assessment of pelvic and para-aortic lymph nodes, figure 1.9. The routine removal of lymph nodes is not currently recommended, however any clinically enlarged or abnormal nodes should be removed in order to achieve complete cytoreduction (Fotopoulou et al., 2017).

Cytoreductive surgery for stage II-IV HGSOC

Midline laparotomy
Total abdominal hysterectomy (cervix, uterus)
Bilateral salpino-oophorectomy (tubes, ovaries)
Omentectomy
Pelvic and paraaortic lymph node assessment +/- removal
+/- Removal of any other disease including peritoneal stripping, upper abdominal disease, bowel disease

Figure 1.9 *Standard procedure for cytoreductive surgery in stage II-IV HGSOC*

The combination of primary debulking surgery (PDS), defined as cytoreductive surgery as the first line treatment, followed by 6 cycles of platinum based ± taxane chemotherapy ± maintenance therapy, remained the mainstay of treatment for many years. However, despite best surgical effort, a proportion of cases remain in which complete cytoreduction is deemed impossible. This often occurs when the perceived morbidity and mortality risks of the surgery are thought to outweigh the benefits. In response to these cases, the concept of neoadjuvant chemotherapy (NACT) with interval debulking surgery (IDS) was developed. This treatment pathway sees the patient first undergoing 3 cycles of platinum based ± taxane chemotherapy, followed by a cytoreductive surgery. When the surgery occurs following 3 cycles of chemotherapy it is termed interval debulking surgery (IDS). The IDS consists of the same procedures and objectives as PDS: complete cytoreduction. In patients with large disease bulk, chemotherapy is potentially not always effective due to limited tumour blood supply, and therefore interval surgery is essential (van der Burg et al., 1995). Following IDS, the patient completes the remaining 3 cycles of chemotherapy ± maintenance therapies. A small proportion of patients undergo delayed IDS, that being surgery following ≥5 cycles of chemotherapy. However, these patients appear to have worse outcomes than with standard IDS, and for that reason surgery is not currently recommended following >4 cycles of chemotherapy(Thomas et al., 2022).

The introduction of NACT/IDS as an alternative treatment pathway to PDS raised questions as to which pathway was superior. A meta-analysis of 835 patients in 2006 to compare outcomes between treatments concluded that patients undergoing NACT/IDS were

associated with worse outcomes (Bristow & Chi, 2006). This study then prompted two multicentre randomised phase III trials to adequately compare the two treatment options (Kehoe et al., 2015; Vergote et al., 2010).

The first of these trials (carried out by the 'European Organisation for Research and Treatment of Cancer' (EORTC) group) recruited 670 late stage HGSOC patients between 1998 – 2006 and randomised between PDS and NACT/IDS, with the primary endpoint of overall survival (OS). The OS in the two groups were largely equivalent (hazard ratio 0.98 CI 90% and 1.01 CI 90% for NACT/IDS and PDS, respectively), although there did appear to be a slight (but not statistically significant) increase in post-operative morbidity in the PDS group (Vergote et al., 2010).

The second RCT (the CHORUS trial) recruited 550 women with late stage HGSOC between 2004 – 2010, randomising between PDS and NACT with the primary end point of OS. They again concluded that OS in patients undergoing NACT was non-inferior to primary surgery (Kehoe et al., 2015). However, it was also reported that treatment related morbidity and mortality was significantly higher in the PDS patients (Kehoe et al., 2015).

Both trials concluded that NACT/IDS is an acceptable alternative to PDS in women with advanced ovarian cancer. It is worth noting that any difference seen between groups in both these trials was very small, with both having very low ten year survival rates, as well as low complete cytoreduction rates (Narod, 2016). Inclusion into these trials was dependent upon the treating clinicians and patients were likely only included when treatment management was unclear. Therefore, patients who would very obviously been suited to one group rather than the other may not have been included, potentially leading to a skewed patient population. Due to this limited patient selection, extrapolation of the results of these trials across the entire ovarian cancer population warrants caution.

Further analysis of EORTC data suggested that if patients have stage IIIc disease with less extensive metastasis, then they were likely to have better outcomes with primary surgery. However, those with stage IV disease or more extensive metastasis gained more survival benefit from NACT/IDS. Patients for whom a complete cytoreduction was not possible at the

time of PDS may live longer undergoing IDS (Gill et al., 2017; Horowitz et al., 2015). These findings led to the established treatment pathway; PDS should be the standard treatment of care, however if complete cytoreduction is thought to be unlikely, NACT/IDS should be considered as an alternative (Fagotti et al., 2013; Gill et al., 2017; Horowitz et al., 2015; van Meurs et al., 2013).

In the UK, the decision as to the most appropriate first line treatment is currently made at the time of tertiary referral at the MDT meeting. NICE state that histological confirmation is not required if the patient is to undergo PDS (National Collaborating Centre for, 2011). If NACT is being considered as first line treatment, histological diagnosis should be sought, with biopsy the most common method. Tissue is obtained by way of radiologically guided, laparoscopic omental or peritoneal disease biopsy before the commencement of NACT (National Collaborating Centre for, 2011). If histological biopsy is not obtainable, then cytological diagnosis can be used, and is most often achieved through ascitic drainage or pleural effusion drainage.

There remains no set guidance to aid this decision-making progress in the UK. A recent national audit comparing the treatment of stage II-V Ovarian cancer of over 13,000 patients in England between 2016 - 2018 highlighted significant discrepancies between cancer alliances. It revealed surgical rates varied between alliances (61.8% - 73.6%), that between 29.6% and 20.7% of patients received no treatment at all. Five-year net survival rates varied between 28.6% and 49.6%, with patients who were statistically less likely to undergo surgery demonstrating lower than average survival. This audit highlights the heterogeneity in the UK of treatment for ovarian cancer, the impact this heterogeneity has upon patient outcomes, and the need for more uniform care.(Sundar et al., 2020)

The summarised treatment pathway options adapted from current UK guidance can be seen in figure 1.10.

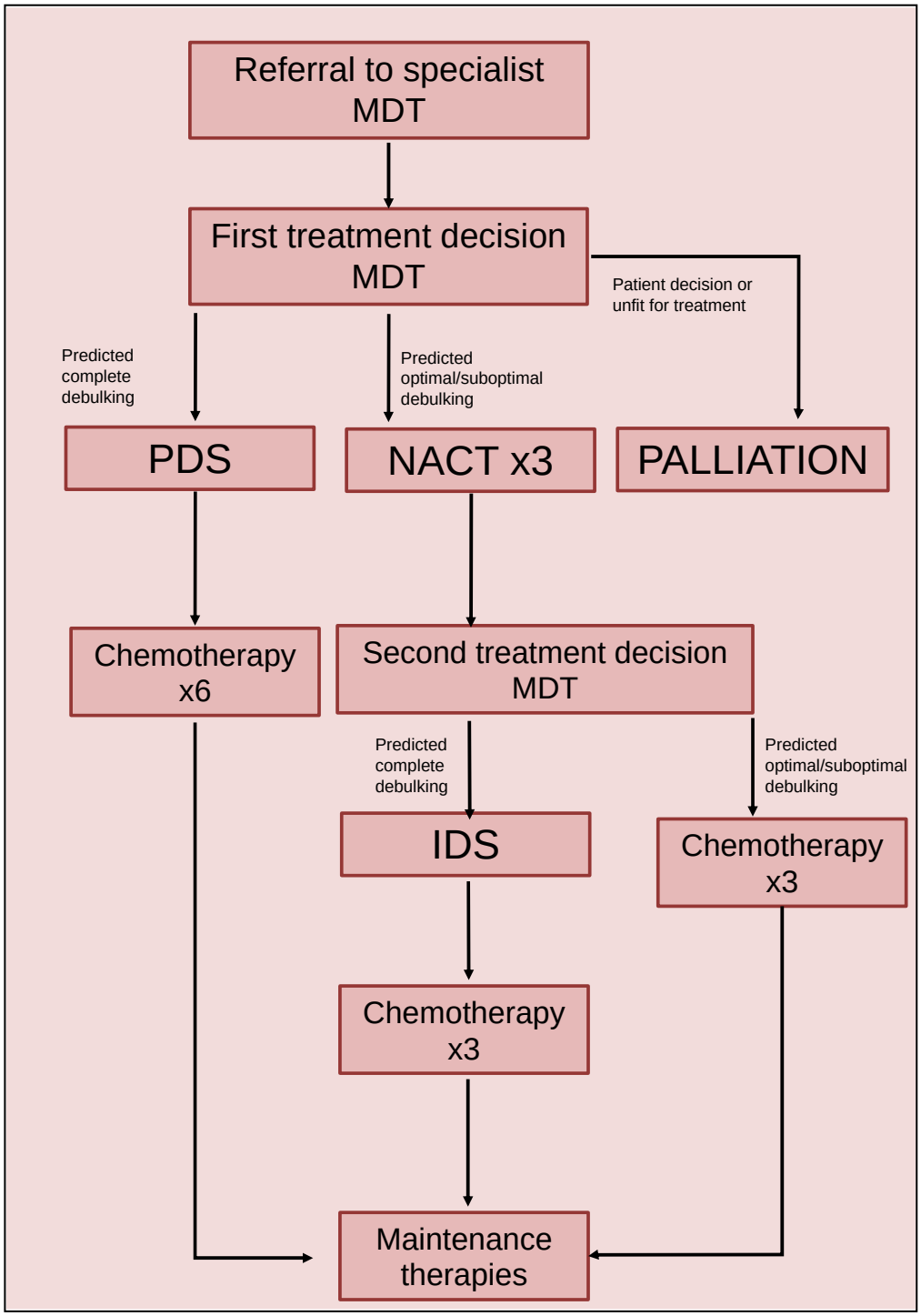


Figure 1.10 *Flowchart outlining possible treatment pathways for stage II-IV HGSOC*

1.6 Factors affecting overall survival in HGSOC stage III/IV

High grade serous ovarian cancer carries the poorest survival rates for all the subtypes of ovarian cancer, with an estimated 36% of patients dying of their disease within the first year of diagnosis (Barclay et al., 2016). Although 60% of HGSOC patients will have advanced disease that has extended outside of the pelvis at the time of diagnosis (CRUK, 2018), the

presentation of each patient is different. The disease biology and distribution will differ, as will the patient's symptoms, and the physiological effect the tumour will demonstrate, biochemically, haematologically and physically.

Outcome at the time of PDS and IDS has been well established as the most important independent prognostic markers of survival in late stage OC since Griffiths' ground-breaking paper in 1975 (Griffiths, 1975; Romanidis et al., 2014; Marianne Jetske Rutten et al., 2015). A surgery resulting in minimal or no visible tumour equates to improved overall outcomes, with an increase in overall survival of over 60 months being achieved with a complete (no visible disease) cytoreduction compared to suboptimal outcomes (Bast et al., 1983; Delgado et al., 1984; Griffiths, 1975; Griffiths et al., 1979; Hacker et al., 1983; Hacker & Rao, 2017; Piver et al., 1988).

More radical surgery, including more extensive upper abdominal surgery, is associated with increased debulking rates, and in turn increased overall survival (Aletti, Dowdy, et al., 2006; Chi et al., 2009). It is well established that surgical heterogeneity exists, not only on an individual level within units, but also between units, and even countries (Aletti, Gostout, et al., 2006; J. M. Janco et al., 2015; Jones et al., 2018). The tendency of the surgeon towards performing radical procedures has been found to be associated with optimal cytoreduction (Aletti, Gostout, et al., 2006). Range of debulking rates between surgeons can be marked (42-67%) depending upon their surgical tendencies (Aletti, Dowdy, et al., 2006). Over time, the radicality of surgery for ovarian cancer has increased, and surgery including bowel surgery, peritoneal and diaphragmatic stripping being described as ultra-radical (Baxter et al., 2019; Lheureux et al., 2019). Ultra-radical surgery performed by sub-speciality trained gynaecological oncology surgeons increases mean survival and progression free survival, however can be offset by an increase in post-operative morbidity (Baxter et al., 2019).

Aside from surgical outcome, there are many other factors that have demonstrated an effect on survival or predict treatment response. These factors can be helpful to guide treatment pathway decisions and design an individualised treatment plan for each patient.

1.6.1 Geographical and socioeconomic factors

Survival in HGSOC varies greatly on a worldwide scale. However, when compared with other countries, patients treated in the UK have lower survival (Doufekas & Olaitan, 2014). This geographical variation could reflect the UK's lower optimal surgical cytoreductive rates when compared to other European countries (Doufekas & Olaitan, 2014), or could also be attributed to the increasing financial pressure within the NHS.

Socioeconomic background has also been linked with ovarian cancer survival. A recent UK study assessing survival in relation to a patients' Index of Multiple Deprivation (IMD), defined by postcode and addressing factors such as access to healthcare, education, financial state, reported that patients from a more deprived background had on average poorer survival rates. Patients from this group were also more likely to refuse intervention for their disease (Phillips et al., 2019). These findings were echoed by the National Cancer Intelligence Network in the UK, who showed that patients who presented as an emergency and who had poor socioeconomic status suffered poorer survival (NCIN, cited 2018). As well as emergency presentation, access to healthcare has also been linked to poor survival. Patients who were unable to access specialist care had worse outcomes in the United States of America (USA) (Urban et al., 2016). However as this was in the USA, this observation was not seen within a comprehensive universal care system such as the NHS. As well as access to care, treating patients in hospitals with National Cancer Centre status leads to better adherence to guidelines and is also reported to improve outcomes (Bristow et al., 2015; Khoja et al., 2016).

Despite a widespread belief that rapid decline and poor outcome is associated with delays in diagnosis or commencement of treatment, there is sparse evidence in the literature to support this, with only a few small studies reporting outcomes reflecting this viewpoint (Kirwan et al., 2002; Urban et al., 2016). A case control cohort study recently disputed this, demonstrating that treatment delays were not associated with poor short-term survival (Hawarden et al., 2021). Prompt referral to specialist care from primary care following the presentation of a patient experiencing red flag symptoms is linked to survival (Rose et al., 2015).

1.6.2 Patient pre-morbid state

In common with many cancers, survival for patients with ovarian cancer is inversely associated with age; women aged 75 – 99 years at the time of diagnosis display the lowest 3 year age specific survival of between 20 – 34% (Cabasag et al., 2020; Chang et al., 2018; Macnab, 2018; NCIN, cited 2018; O'Malley et al., 2012; Urban et al., 2016). Older women are also more likely to present with advanced disease (Cabasag et al., 2020), and this, alongside the increasing co-morbidities seen in older patients (Janssen-Heijnen et al., 2005) could be further compounding this effect.

Increasing age is associated with increasing co-morbidities, however the presence of significant co-morbidities in women, when age, stage and socioeconomic factors are adjusted for, can increase this risk of death by as much as 40% (O'Malley et al., 2012). There is conflicting evidence as to the association between increased body mass index (BMI) and mortality in ovarian cancer, with a large systematic review and meta-analysis suggesting a link between obesity and increased mortality (Yang et al., 2011). However, Zhou et al reported that although increased activity levels pre-diagnosis could decrease ovarian cancer specific mortality by 26%, BMI itself had no effect on mortality (Zhou et al., 2014).

1.6.3 Distribution of disease at time of presentation

Histological type of ovarian cancer plays a large role in a patient's survival, with the majority of type I tumours, excluding clear cell cancer, displaying overall significantly better survival than type II tumours. Of all the histological subtypes, HGSOC and clear cell tumours appear to have the worst outcomes. Clear cell tumours affect a proportionately younger cohort, usually at an earlier stage, and HGSOC an older cohort and at a later stage. Tumour cell type is the most useful histological prognostic marker for ovarian cancer (Ezzati et al., 2014).

It is well recognised that increasing stage of disease at the time of presentation has a negative effect on survival. The FIGO staging system used in ovarian cancer has been extensively studied (Benedet et al., 2000) and is well recognised as a surrogate marker for survival (Ezzati et al., 2014), with five year survival falling from 95% for stage I down to just 15% for stage IV disease (CRUK, 2018). The presence of ascites at the time of diagnosis, even with early stage

disease, is associated with significantly worse mortality than those where ascites is not present (Ezzati et al., 2014). This effect remains when early stage cancers are excluded from analysis, with presence of ascites alongside stage III and IV disease remaining as an independent marker of poor survival (Puls et al., 1996).

Whilst FIGO staging describes the anatomical position of the tumour distribution, it does not address volume of disease, termed 'disease burden'. A high pre-operative disease burden is an independent negative prognostic indicator, even when complete cytoreduction is achieved at the time of surgery (Horowitz et al., 2015).

1.6.4 Pre-treatment haematological markers

Many pre-treatment serum blood parameters have been associated with survival in ovarian cancer. Low serum albumin at the time of presentation is a surrogate marker for poor nutritional status, and has been associated with poor OS and progression free survival (PFS), as well as increased rates of surgical site infection and prolonged post-operative hospital stay (Ge & Wang, 2018; Koirala et al., 2020). A raised pre-operative fibrinogen level and platelet count are associated with worse outcomes, with a pre-operative platelet count above 289 $x10_9$/L demonstrating a significantly shorter OS (37.3 VS 46.1 months, HR 1.14 95% CI 0.89 – 1.46, P=0.306) (Feng et al., 2016; Wahner Hendrickson et al., 2015). A high neutrophil-lymphocyte ratio (NLR) has also been associated with increased survival (Ashrafganjoei et al., 2016; Feng et al., 2016; Zhou et al., 2017). A low haemoglobin (Hb) (<120g/L) pre-operatively has also been linked to worse overall survival and a higher risk of recurrence (Obermair et al., 2000; Warner et al., 2013). Correction of this anaemia with a blood transfusion has been shown to worsen rather than improve OS, which may be a reflection of the initial low Hb being a marker of more advanced disease (Pergialiotis et al., 2020).

CA 125 has been extensively studied in ovarian cancer and plays a role in the diagnostic pathway, monitoring of chemotherapy response as well as detection of relapse (Chang et al., 2016; Chen et al., 2019). A review assessing the role of CA 125 in predicting survival in ovarian cancer described an inverse relationship between level at diagnosis and OS. The prognostic

value was stronger for levels in response to treatment than pre-treatment levels (Gupta & Lis, 2009). The same findings are echoed for HE4, which displays some prognostic value at pre-treatment levels, the change to levels in response to treatment have a higher predictive value (Chudecka-Glaz et al., 2014).

1.7 Genomics in ovarian cancer

Following advances in and decreasing costs of sequencing technologies, understanding of the molecular basis of ovarian cancer has improved. While surgery and platinum-based chemotherapy remain the backbone of ovarian cancer treatment, platinum resistant cancers recur in approximately 25% of patients within six months, creating the urgent need for alternative therapies (Miller et al., 2009).

The Cancer Genome Atlas Programme (TCGA) analysed messenger ribonucleic acid (RNA) expression, microRNA expression, promoter methylation and DNA copy number in 489 HGSOC tumour samples, and performed exon sequencing on 316 of the tumours (TCGA, 2011). Almost all tumours contained TP53 mutations as well as extensive copy number variation (TCGA, 2011). Approximately half of HGSOC tumours are associated with a deficiency in the homologous recombination (HR) pathway, with approximately half of these tumours (20% overall) containing a mutation in the BRCA 1 and BRCA 2 genes (Hudson et al., 2010; "Integrated genomic analyses of ovarian carcinoma," 2011; Mukhopadhyay et al., 2012; TCGA, 2011). These mutations can be germline (identified in the patient serum sample), and therefore contribute to an inherited pre-disposition, or are somatic (identified in the tumour sample only) in origin.

Chromosomal instability is the hallmark of a HGSOC cell. Cell DNA damage occurs as either a single strand or double strand break, and mammalian cells utilise one of five mechanisms to identify and repair DNA damage. Single strand breaks are repaired by mismatch repair (MMR), base excision repair (BER) and/or nucleotide excision repair (NER). Double strand breaks are repaired by HR and/or non-homologous end joining (NHEJ). It is likely that all HGSOC contain a defect in at least one of these pathways (Gee et al., 2018).

At present only a defect in the HR pathway has been utilised as a predictor of response to

therapies (Gee et al., 2018). Patients who harbour a defect in their HR pathway (HRD) have a better overall survival, and better response to treatment with platinum based chemotherapy, as well as PARP inhibitors (Macintyre et al., 2018). This survival advantage was at first thought to be limited to patients with a BRCA mutations, but it is now accepted that this advantage holds true for all HRD tumours (González-Martín et al., 2019).

In order to establish whether a patient or tumour has a deficiency in the HR pathway, one of three methods is used:

1. Gene panels - detection of germline or somatic mutations in genes in the HR pathway;
2. Surrogate tests - detecting genomic scars or mutational signatures representing patterns of genomic instability; or
3. Functional tests - checking the function of RAD51 localisation to sites of DNA damage (Chiang et al., 2021).

The most commonly included genes in panels for HRD are BRCA 1 and BRCA 2, with 17% of HGSOC patients carrying a germline mutation in these genes. It is important to test for both germline and somatic BRCA mutations, as a further approximately 6 - 7% of patients with a negative germline test will have a somatic mutation (Chiang et al., 2021). Although an isolated somatic mutation does not have the inherited implications for the patient and their family, these patients would be suitable for PARP inhibitor therapy, and it is therefore important to identify them. Although BRCA 1 and BRCA 2 are the most common gene defects, approximately 28% of patients will have aberrations in other genes involved in the HR pathway (Chiang et al., 2021). The TCGA devised a gene signature of 16 genes involved in the HR pathway, to be assessed to determine HR status (BRCA1 BRCA2 C11of30 PTEN RAD51 ATM ATR PALB2 FANCA FACCI FANCL FANCD2 FANCE FANCG FANCM). The inclusion of PTEN was noted to be contentious as its exact role in the HR pathway remains controversial, a belief supported by several other recent studies (Bian et al., 2018; Huang et al., 2018; Hunt et al., 2012; TCGA, 2011). A major challenge in using gene panels to test for HR status is the annotation of variants of uncertain significance (VUS). In the broader gene panel tests, the functional and clinical impacts of most individual mutations in the genomic loci have not been well characterized (Chiang et al., 2021). For this reason, there is currently no single agreed gene panel for HRD.

Surrogate tests identify permanent genomic scars or mutational signatures caused by the genetic variations as a result of defects in the HR pathway (Lord & Ashworth, 2012). These genetic variations generally consist of either copy number variants, single nucleotide variants, and small insertions and deletions (Chiang et al., 2021). The myChoice CDx (Myriad Genetics) and Foundation Focus CDx BRCA LOH (Foundation Medicine) are the two most commonly available next-generation sequencing (NGS) assays (Frampton et al., 2013; Telli et al., 2016).

With ever-increasing evidence highlighting the importance of HRD status in predicting a patient's response to treatment, a faster, cheaper and more widely available functional assay would be beneficial. There are currently two commercially available functional HRD tests (Pellegrino et al., 2019), with many groups publishing their own methods. The majority of these tests include inducing DNA-damage *ex vivo* and detecting the nuclear localisation of RAD51 to assess HR status (Graeser et al., 2010; Mukhopadhyay et al., 2010; M. Tumiati et al., 2018).

1.8 Existing predictors of surgical outcome in ovarian cancer

Complete cytoreduction is the overriding goal of surgical treatment, with OS being compromised if this is not achieved (Horowitz et al., 2015; Rose et al., 2004). If complete cytoreduction is not likely to happen at the time of PDS then IDS holds a survival advantage (Chern & Curtin, 2016; van Meurs et al., 2013). Therefore, the clinical need exists to develop a method or tool with the ability to predict the cytoreductive outcome of surgery. In current clinical practice, the rate of suboptimal cytoreduction ranges broadly between 9 – 65% (Horowitz et al., 2014; J. M. Janco et al., 2015). The patients suffering suboptimal cytoreduction at the time of PDS may have had improved survival and lower morbidity if they had undergone IDS and may have not benefitted from a surgery at all. Any pre-operative prognostic tool developed would need to reduce this percentage.

There have been many studies investigating the predictive ability of multiple modalities, both individually, and combined to create prediction models. Many have sought to predict those patients who will have suboptimal outcomes at the time of surgery, in the hope that if these

patients are identified pre-operatively, they can be instead redirected towards NACT/IDS. Other models have attempted to predict those patients who will have complete outcomes, as a way to direct those patients strongly towards PDS. The models incorporate a variety of modalities including patient characteristics, biochemical markers, radiological findings and disease distribution at the time of laparoscopy.

1.8.1 Biochemical predictors

The ability of serum biological predictors, taken via a relatively non-invasive low risk blood test, as a tool for prediction of optimal cytoreduction would be an attractive concept. CA 125 is a marker routinely tested as part of the standard diagnostic pathway in patients with suspected ovarian cancer (Ledermann et al., 2013) and therefore is a marker widely available for assessment. CA 125 is therefore the most commonly investigated marker in the literature.

A meta-analysis performed by Kang et al (2010) summarised findings of CA 125 as a marker for OC, identifying 14 studies incorporating 2,192 patients. They found that none of the investigated CA 125 cut off levels were able to predict the success of optimal cytoreduction and concluded that the sole use of CA 125 to aid the clinical decision of treatment route was inappropriate (Kang et al., 2010).

There have been multiple small studies investigating many other clinical variables, and their use as predictors of surgical cytoreductive status, with limited success. These include gene expression data (Abdallah, Chon, et al., 2015; Berchuck et al., 2004), biological markers such as YKL-40, bcl-2, cathepsin L (Chudecka-Glaz et al., 2014), ADH 1B and FABP4 (Tucker et al., 2014), peritoneal vascular endothelial growth factor burden (Diniz Bizzo et al., 2010), the use of proteomic panelling (Risum et al., 2009), the neutrophil-lymphocyte ratio (Wang et al., 2015), as well as simple haematological markers such as albumin (de Jong, Eijkemans, Lie Fong, et al., 2007). Of the above, albumin and the biological markers ADH 1B and FABP4 show the most promise, although both studies incorporated only very small numbers, and are yet to be validated.

1.8.2 Radiological predictors

The most frequently investigated modality in the prediction of surgical outcome is radiology. This is likely owing to the fact that patients have a CT scan as routine pre-operative management. There have been 30 studies investigating imaging in the last 20 years, with all but two focussed upon computerised tomography (CT) as the modality of choice.

A systematic review of models created to predict cytoreductive surgery outcomes in patients with FIGO stages III and IV ovarian cancer using CT imaging variables (M. J. Rutten et al., 2015) identified 11 models that had been developed before this date. Only one model maintained its predictive accuracy when internally validated (C. G. Gerestein et al., 2011). None were successfully externally validated.

1.8.3 Diagnostic laparoscopic prediction

The principle of diagnostic laparoscopy as a predictor of surgical outcome combines the advantage of real-time assessment with the ability to obtain tissue biopsy to establish a definitive histological diagnosis. It is not currently part of the treatment pathway used in the UK, however patients do on occasion have a pre-debulking surgery laparoscopy to obtain a histological biopsy, although only usually if treatment is leaning towards NACT and IDS.

There have been significant concerns raised regarding the safety of laparoscopies in gynaecological cancer surrounding the occurrence of port site metastasis (PSM) (Manvelyan et al., 2016). There are several hypotheses that attempt to explain the cause of PSM, including immune response, spread by pneumoperitoneum, and wound contamination (Manvelyan et al., 2016). It is accepted that in general, oncology patients who develop PSMs are associated with poor outcomes (Manvelyan et al., 2016). Although a meta-analysis of 11,027 patients with a malignancy found the rate of PSM to be low (<2%) (Manvelyan et al., 2016), it is suggested that patients who are at highest risk are those with disease of ovarian or peritoneal origin, the presence of ascites, and biologically aggressive

disease (Manvelyan et al., 2016). Vergote et al studied rates of PSM in 173 patients with ovarian cancer undergoing laparoscopy in 2005. They found an increased rate of PSM in this cohort, although did not find this to be associated with poorer outcomes (Vergote et al., 2005).

An economic analysis to evaluate the cost-effectiveness of diagnostic laparoscopies was conducted in 2017. This measured both direct medical costs alongside health outcomes by way of quality-adjusted life-years. They concluded that the rates of futile laparotomies could be reduced, without increasing direct medical health care costs or adversely affecting complication rates or quality of life (van de Vrie et al., 2017).

1.8.4 Inclusion of surgeon heterogeneity in prediction

Several studies have demonstrated that more radical surgery, including more extensive upper abdominal surgery, is associated with improved cytoreduction rates, and in turn increased overall survival (Aletti, Dowdy, et al., 2006; Chi et al., 2009). It is well established that surgical heterogeneity exists, not only on an individual level within units, but also between units, and even between countries (Aletti, Gostout, et al., 2006; J. M. Janco et al., 2015; Jones et al., 2018). Multiple prediction models have been developed, incorporating a wide variety of clinical data. However only two studies incorporate and address the issue of surgeon heterogeneity. It has been suggested that in excluding this factor, all prediction models will be unsuccessful, as cytoreduction rates rely so heavily on surgical practice (Aletti, Dowdy, et al., 2006).

1.9 Summary

Patients diagnosed with HGSOC face bleak survival outcomes. Their cytoreductive outcome is an important prognostic marker of survival. There are two well-established treatment pathways in the treatment of advanced stage HGSOC; PDS and NACT/IDS. Both pathways are a combination of both surgery and chemotherapy, and it is widely accepted that patients who achieve complete surgical cytoreduction benefit from an improved OS, especially if this is achieved at primary surgery.

For this reason, there have been a plethora of prediction scores and models developed aiming to predict both good and bad surgical outcomes in order to correctly triage each patient towards the treatment pathway that will give them the best survival chance. However, despite being widely researched there have to date been no robustly, externally validated models that would be of use in clinical practice. Therefore, at present clinical acumen remains the mainstay of treatment decision-making in UK practice. Despite this, there are a consistent percentage of patients that undergo only optimal or suboptimal cytoreduction, and therefore potentially have their survival outcome compromised due to the course of treatment they received.

A limitation of existing prediction models is the failure to be successfully externally validated on a cohort that differs to the one on which they were developed. It is possible that developing a model that has been built using very locally specific patient demographics and surgical practice heterogeneity renders it non-transferrable to a different population. There has been promising work utilising immunohistochemistry as a predictor, as well as laparoscopic scoring systems. Additionally, models appear to increase in their accuracy as more modalities are introduced. Despite dramatic advancements in genomic medicine over recent years including the recent completion of the 100K Genome Project in the UK, there have been very few prediction models utilising genomic data.

An ideal surgical outcome prediction tool should have the ability to accurately predict surgical outcome, whilst being transferable between different patient populations, regardless of surgical practice or patient demographic. It should include data that is readily available to clinicians, with minimal increased procedures. It should be quick to perform, to avoid the delay of initiating treatment.

Such tool could be further enhanced by the incorporation of artificial intelligence (AI). AI would allow a prediction tool to be adaptive, and learn and develop from its own dataset, as patient demographics change and surgical trends develop. Therefore, increasing in accuracy over time as the dataset increases in number. By developing a tool that is multi-

modal and self-adapting depending upon individual patient characteristics and current surgical practice, emphasis shifts from a ‘one tool fits all’ towards ‘individualised’ medicine.

2 Hypothesis, aims and objectives

2.1 Hypotheses

1. There is currently a dearth of surgical prognostic models which correctly and accurately predict cytoreduction in PDS in HGSOC patients. Models which have been validated are unlikely to further validate when applied to populations with differing surgical practice.

2. The improved survival of patients with a defective homologous recombination pathway may be further accounted for by improved surgical outcome.

3. By combining all available patient, tumour and surgeon data, a prognostic surgical tool will improve predictive accuracy and allow successful validation.

2.2 Aims

1. Comprehensively assess all currently existing surgical prediction tools in high grade serous ovarian cancer.
2. Develop biological predictors to predict surgical outcome in high grade serous ovarian cancer.
3. Develop a novel surgical prediction tool in HGSOC.

2.3 Objectives

1. Perform a systematic review of all pre-existing published surgical prediction models.
2. Externally validate a published three protein surgical prediction gene panel using immunohistochemistry using a separate large cohort of tumour samples.
3. Explore the surgical predictive value of HR status in high grade serous ovarian cancer tumours.
4. Develop novel multimodality surgical prediction model for patients with stage III and IV HGSOC.

3 Materials and methods

3.1 General laboratory practice

All experiments were performed to university standards for safe working with chemical substances in laboratories, which comply with the Control of Substances Hazardous to Health (COSHH) Regulations 2002.

3.2 Systematic review

Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines were followed to ensure transparency and quality of reporting in the systematic review (Liberati et al., 2009). The research question was formulated and defined using the following PICOS strategy (Eriksen & Frandsen, 2018):

- Population: Patients of any age, with suspected/confirmed epithelial ovarian cancer FIGO stage I-IV who are being considered for PDS;
- Intervention/Exposure: Primary debulking surgery;
- Comparator: Not applicable;
- Outcome: Prediction of surgical debulking status using definitions as set out in table 1.5 (complete - no macroscopic disease, optimal - $\leq$1cm disease, suboptimal - >1cm disease);
- Study design: Pre-surgical prognostic prediction models incorporating more than one predictor.

The goal of the review was to systematically search and summarise all available multi-predictor prognostic models aimed at pre-operatively predicting cytoreduction status in stage II-IV epithelial cancer in primary debulking surgery.

3.2.1 Literature search strategy

The original literature search was undertaken in February 2020 and included the databases MEDLINE, EMBASE and Web Of Science. Full search strategy is outlined in figure 3.1.

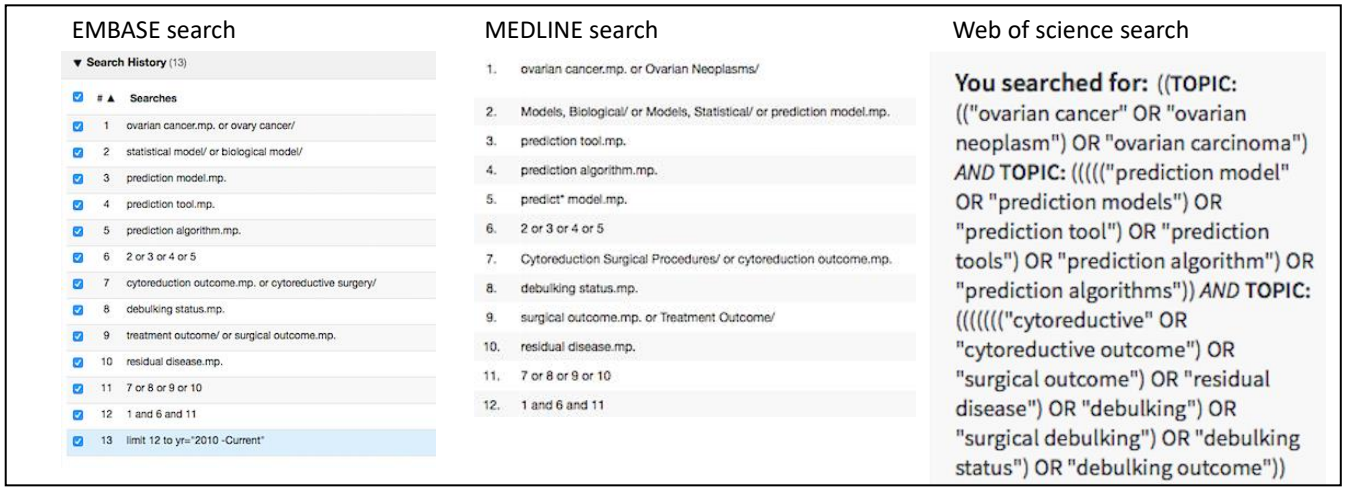


Figure 3.1 *Search strategy to identify relevant studies*

Performed February 2020

3.2.2 Inclusion and exclusion criteria

Pre-specified inclusion criteria were set out to determine the inclusion of abstracts and articles. Inclusion criteria were peer-reviewed original studies presenting models for the pre-operative prediction of surgical outcome (complete, optimal or suboptimal) at the time of primary debulking surgery in newly diagnosed epithelial ovarian cancer stage II-IV. Models were required to include more than one predictor. Models or studies evaluating the predictive ability of individual predictors were excluded. English language was required, and the publication date considered was from 1 January 2010 to 1 February 2020. Models included could be at the development stage and internally or externally validated, either by the original developers or external groups. Studies reporting amendments or improvements to previously reported studies were also included.

3.2.3 Selection of studies

Titles and abstracts of references identified by the search strategy above were independently reviewed by two reviewers, A. Hawarden (AH), thesis author and B. Russell (BR), post-doctorate researcher, according to the aforementioned eligibility criteria. This was followed by a full-text review by AH.

3.2.4 Data extraction

A standardised data extraction spreadsheet was developed, and data were collated in excel. Data extraction of the first reference was first performed to assess the relevance of the data extraction spreadsheet. Following this, relevant data were extracted by AH. The extraction was performed according to the Checklist for critical Appraisal and data extraction for systematic Reviews of prediction modelling Studies (CHARMS) (Moons et al., 2014). Data collected therefore included manuscript general information, population description, characterisation of predictors and outcome, model development, model performance, results and conclusions.

3.2.5 Risk of Bias

Quality and risk of bias of each study was assessed by AH using the Prediction Model study Risk Of Bias Assessment Tool (PROBAST) (Wolff et al., 2019). This tool is designed specifically for systematic reviews of prediction models. Risk of bias was assessed through questions regarding several domains: participants, predictors, outcome and analysis. The questions were answered as either 'high risk of bias', 'low risk of bias', or 'unclear risk of bias'.

3.2.6 Data collation

Data were collated in excel, and all analysis performed using Prism 9 for macOS version 0.2.0.

3.3 External validation of a three-protein signature

Permissions for the replication of data from the original paper were sought and secured. Ethics approval for the use of tissue samples and corresponding clinical data from the ICON5 study were also in place at the start of the study appendix B.

Examples of patient consent forms allowing the procurement of patient samples included in the Manchester ovarian cancer homologous recombination (MOCHR) database, as well as collection of clinical data can be found in appendix C.

All lab practice including immunohistochemistry (IHC) protocol, antibody concentrations and scoring were performed in keeping with the original paper.

We applied to gain access to the biorepository associated with the ICON5 study, a large number of original tumour slides from UK recruiting centres, along with a limited amount of demographic data and surgical outcomes. The tissue samples made available to our group consisted of pre-cut plated and paraffin fixed samples that were collected at the time of surgery within the trial. These samples were stored at room temperature.

Clinical data corresponding to tissue samples were provided by the Medical Research Council Clinical Trials Unit (MRC CTU) as an excel spreadsheet. All clinical data including surgical outcomes remained unknown to the author until the time of data analysis.

3.3.1 Immunohistochemistry

All antibodies used were in line with working concentrations used in the original paper by Riester et al (Riester et al., 2014) whenever possible and were as follows; Anti- POSTN 1.25μg/mL Oxford biosystems (RD18104050), Anti- CXCL 14 2.5μg/mL ab46010, abcam, Cambridge, UK, Anti- phosphor- Smad2, cell signalling Tech (3108S). All primary antibodies were raised in rabbit.

The initial antibody concentration optimisation was performed by hand by AH. The IHC staining of slides for the final validation was performed by the lab group led by Garry Ashton in the Cancer Research UK Manchester institution using the BOND-III automated IHC stainer. The automation resulted in greater standardisation and is the preferred method for large sample numbers.

Slides were stored at room temperature, pre- cut and paraffin fixed since the time of inclusion in ICON5 study (2001 - 2004).

The three-layer avidin–biotin technique was used. Each slide was stripped of paraffin by immersion in Histoclear for 30 minutes and rehydrated through graded concentrations of

ethanol for 3 minutes at 100%, 100% and 70%. Endogenous peroxidase was quenched with 3% hydrogen peroxide. The slides were washed in phosphate-buffered saline prior to heat-induced epitope retrieval by boiling covered samples in 400ml 0.01% sodium citrate buffer pH 6.0 in a microwave for 2 x 5 minutes, cooled for 20 minutes and washed for 5 minutes. To reduce background staining, the slides were blocked with 5% bovine serum albumin in Tris buffered saline for 30 minutes at room temperature. Primary antibodies, diluted in phosphate-buffered saline and used at optimisation concentrations were applied overnight at 4°C. As a negative control, primary antibody was substituted with non-immune rabbit IgG.

The slides were washed in phosphate-buffered saline to remove excess primary antibody and a concentration 1μg/ml of secondary antibody applied to the slides for 30 minutes at room temperature. After washing off secondary antibody, slides were incubated with avidin biotin complex (50μl avidin, 50μl biotinylated horseradish peroxidase, 2.5ml phosphate buffered saline) for 30 minutes before being washed with phosphate-buffered saline. Slides were then further incubated with 1-3 drops of peroxidase substrate (1.6ml distilled H_2O, 5 drops 10x substrate buffer, 1 drop 50 x 3,3'–Diaminobezidine Tetrachloride, 1 drop 50x peroxidase substrate) for 30 seconds–10 minutes during which time a brown colour change was observed. The slides were counterstained with Harris' formulation #2 hematoxylin for 5 -10 seconds, de-stained with acid alcohol and bluing reagent and washed with tap water. Finally, slides were dehydrated through graded alcohol 70%, 95%, 100% to xylene, agitated in Histoclear and coverslips were fixed using p-xylene bis-pyridinium bromide.

Hand stained slides were examined with a microscope and a single power field of each core examined at 200 - 400 times magnification.

Slides stained on BOND-III were digitally scanned and images captured using Leica SCN 400. All scanned images were accessed remotely for scoring using QuPath–0.2.0–m8 software. All images were scored at 20 times magnification.

Table 3.1 *Antibody concentrations and suppliers*

Antibody	Working concentrations	Supplier
Anti- POSTN	1.25μg/mL	Oxford biosystems (RD18104050)
Anti- CXCL	14 25μg/mL	ab46010, abcam, Cambridge, UK
Anti- phosphor- Smad2	Not specified	cell signalling Tech (3108S)

3.3.2 Immunohistochemistry slide scoring system

The slide scoring system followed that described by Riester et al (Riester et al., 2014). All slides were scored by two independent scorers AH, M. Price (MP) lab technician, and any discrepancies between scores settled by an independent scorer G. Wilson (GW), Consultant histopathologist specialising in gynaecologic oncology. Before scoring, AH, MP and GW determined examples for varying percentages of positive cells, as well as examples of the staining intensities for each of the three proteins stained, for reference. As in the original study, three separate area regions of interest (ROIs) were scored. These ROIs were pre-selected by AH and MP to be representative of the staining for each slide and both scorers scored the same ROIs. The scoring system is illustrated in figure 3.2. The total score (0 - 12) was calculated by multiplying the average percentage of positive cells score (0 - 4) by the average intensity of staining score (1 - 3). Consistency of scores was checked for quality control following the first 50 slides scored before proceeding with the remaining slides.

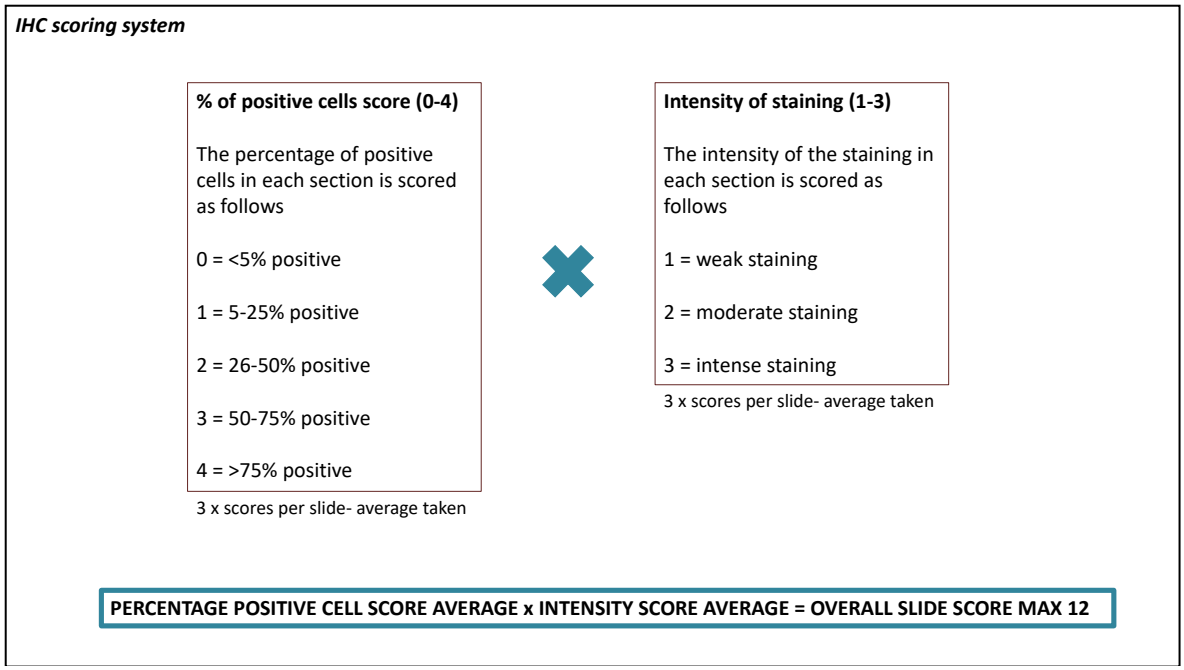


Figure 3.2 *Slide scoring system used for IHC stained slides*

3.3.3 Sample size calculation

The sample number required in order to gain statistical significance for the independent validation of a pre-existing prediction model is only sparsely described in the literature. However, general consensus dictates that sample size can be calculated based on number of events and that a number of events > 100 was optimal. In the case of the three-protein signature, an event is defined as a suboptimal cytoreductive outcome (Collins et al., 2014). However in a review of 78 studies, 45% of studies had < 100 events, 21% of studies did not report the number of events, and the range of events was between 6 – 42,408 (Collins et al., 2014). Although a number of events greater than 100 is optimal, this may not always be possible in practice.

An event in this external validation is defined as a suboptimal cytoreduction outcome. The ICON5 study reported a suboptimal debulk rate of 30% overall. Therefore, in order to achieve 100 suboptimal outcome events, a sample size of 333 patient samples would be required to achieve statistical significance.

3.3.4 Statistical analysis

A p value of <0.05 was used to determine significance for all statistics.

Data were collated using excel, and for description of demographic data a mean and range was used for normalised data, and a median and range for non-normalised data. Inter-variability was determined by calculating Spearman's Rank Coefficient, and differences in cohorts when data was not paired was determined using Mann-Whitney-U test, both performed using Graphpad Prism version 8.4.3 (471).

Validation of both individual proteins predictive value, and combined score value was firstly calculated via simple logistic regression in Graphpad Prism version 8.4.3 (471). The original paper contained access to the original R code. The code was run in R (BR), inputting the validation dataset to ensure consistency of results.

The multivariable prediction model was created using logistic regression in WEKA, an open source machine learning software.

3.4 Homologous recombination functional assay

All tissue samples included for HR status testing were collected from theatres at Saint Mary's Hospital, Manchester following resection, or diagnostic biopsy. Suitable patients were identified at the time of the MDT. Characterisation and functional HR assay was undertaken by MP. Samples categorised as either homologous recombination repair competent (HRC), homologous recombination repair deficient (HRD), or homologous recombination repair heterogenous (HRH). HRH was reserved for those patients who harvested samples from multiple sites at the time of collection, and there was discordance between the HR status of the samples.

Cells were cultured from resected tumour tissue over a period of two weeks in a 175cm^2 adherent tissue culture flask (Sarstedt AG & Co), until fully confluent. 80,000 cells were isolated and immunofluorescently stained for CA 125, PAX8, Vimentin, and a Pan-Cytokeratin marker (Abcam PLC), in order to ensure cultured cells were epithelial and likely representative of the patient's cancer.

40,000 cells from the above culture were then divided into UV treated and Control / non-UV treated slides. The UV treated slides were subjected to UV type C irradiation equivalent to 200J/m^2.

Slides were returned to the incubator for 2 hours under fresh complete media. At +2 hours all slides were fixed in ice cold methanol for 20 minutes. Cellular and nuclear membranes were permeabilized using 0.5% (v/v) Triton 100 in PBS for 5 minutes and a 5% (v/v) Goat serum (Sigma Aldrich) block applied for 1 hour.

Cells were stained with mouse anti-gH2AX antibody (merk-millipore) and rabbit anti- RAD51 antibody (abcam PLC), both at 2µg/ml. Goat anti-mouse, Alexafluor 546 conjugate, and Goat anti-rabbit, Alexafluor 488 conjugate (Invitrogen) antibodies were applied at 1µg/ml. Slides were then mounted with Vectashield H100 with DAPI. Slides were imaged using a Zeiss Axio Observer microscope at x40 magnification and Zen Software.

Images were imported into imageJ (FIJI distribution) and an automated script was used to count the number of nuclei and measure the average number of Alexafluor 546 / gH2AX foci and the number of Alexafluor 488 / RAD51 foci across both the UV treated and Control slides. A greater than 2-fold increase in the number of gH2AX foci from control to UV treatment represented successful damage induction from irradiation.

A greater than 2-fold increase in RAD51 foci from control to UV treatment was taken to represent a Homologous recombination repair competent sample (HRC), and a less than 2-fold increase was taken to represent a Homologous recombination repair deficient (HRD) sample.

3.5 Manchester data base composition

3.5.1 Patient identification

All patient data were pulled from the Manchester database, which holds data for all patients diagnosed with ovarian cancer between 2013 – 2018 who were discussed at the MDT meeting at St Mary's Hospital, Manchester.

The database was compiled via retrospective data collection over several years. Patient NHS numbers (unique identifiers) were identified by searching the SOMERSET database for a diagnosis of 'ovarian cancer', 'fallopian tube cancer', or 'primary peritoneal cancer'. The original histology reports were checked for all included patients to confirm the diagnosis of ovarian/fallopian/primary peritoneal cancer. All cytological and histological samples were assessed by two independent consultant histopathologists, one being a specialist in gynaecological cancers. All histological samples were assessed macroscopically, microscopically and underwent immunohistochemistry staining for p53, WT-1, oestrogen receptor, PAX-8, CK7 and CK20. In patients where histology was not available, cytology was used to make a diagnosis (from ascites or pleural effusions) and diagnostic methods included immunocytochemistry with the panel outlined above. In cases where the patient died before any samples were obtained, the diagnosis was made on clinical grounds by consensus between consultant radiologists, gynaecological-oncologists and clinical oncologists, based upon imaging.

Patients included in the MOCHR cohort were identified and recruited prospectively at the time of the MDT if they fulfilled the criteria set out in figure 3.3. Tumours that were harvested at the time of both PDS and IDS were included, and were categorised as either HRD, HRC or HRH using the functional assay. Tumours with a histology other than HGSOC, as well as patients with an HRH status were excluded.

3.5.2 Data collection

Data were collected at a basic level for all patients, and then with additional more in-depth fields for the Manchester high grade serous ovarian cancer (MHGSOC) patients and the Manchester ovarian cancer homologous recombination (MOCHR) patients, figure 3.3. All fields were collected manually, by mining several electronic patient record systems including: SOMERSET (MDT decision making documentation, date of diagnosis, performance score, dates of MDTs, CT scan expert radiologist review); PACS (CT scan general radiologist reporting); ICE (blood parameters, histology reports); and Christie web portal (CWP)

(chemotherapy and medical treatment information, BRCA status, co-morbidities, BMI, survival data).

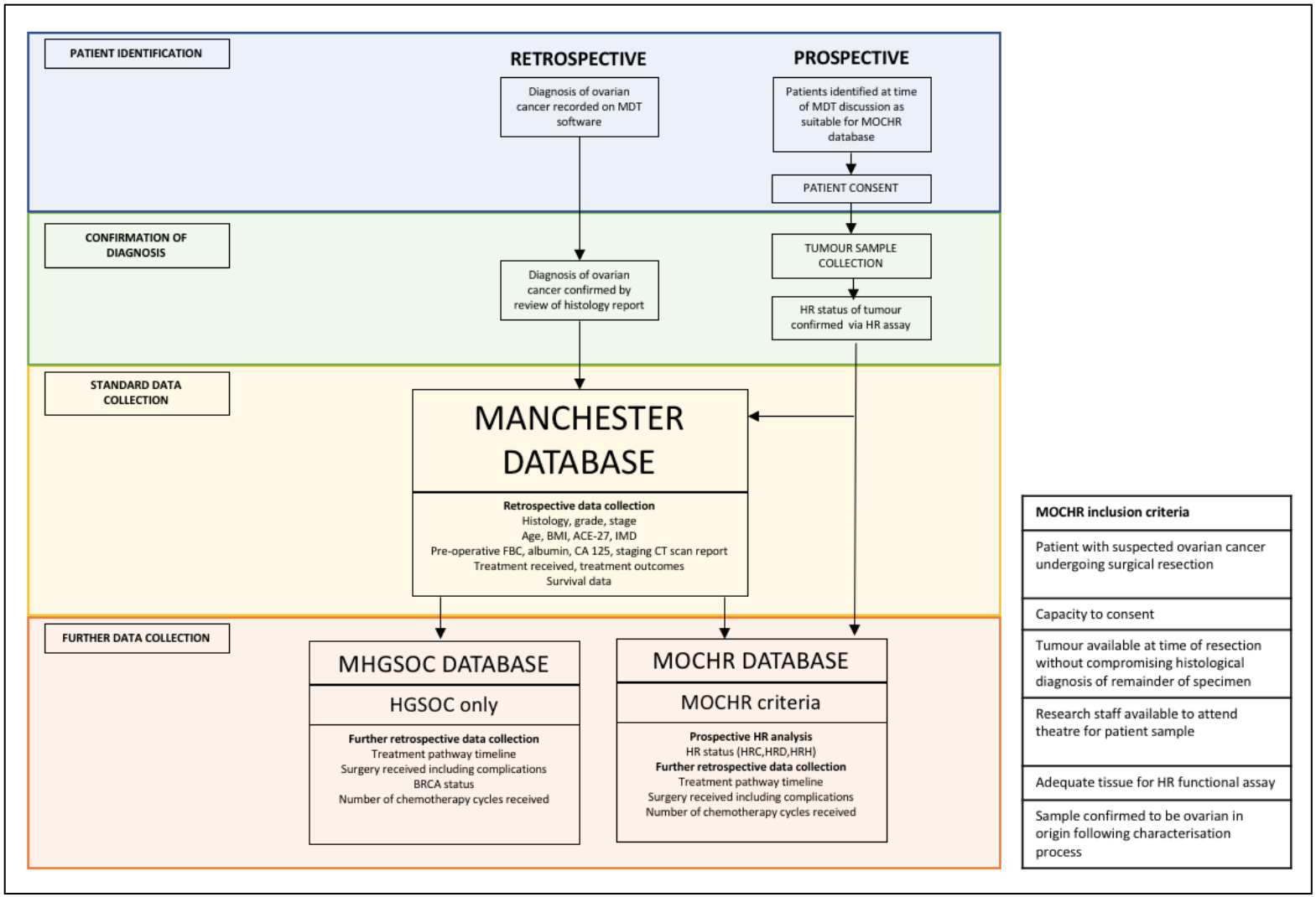


Figure 3.3 *Development of Manchester database and its sub-cohorts*

Key*- MOCHR- Manchester ovarian cancer homologous recombination database, MHGSOC- Manchester high grade serous ovarian cancer database, ACE-27- score of comorbidities, BMI- body mass index, FBC- full blood count*

3.5.2.1 Patient co-morbidities and baseline clinical characteristics

Patient's age, body mass index (BMI), Adult comorbidity evaluation-27 index (ACE-27) and index of multiple deprivation score (IMD), were collected as correct at the time of presentation and first discussion at the MDT. These factors were chosen to represent surrogate markers of general health before onset of disease, generally defined as "individual effects''. Although BMI can be effected by the presence of ascites or cachexia, it remains a useful marker, especially for extremes of weight categories. The ACE-27 score quantifies co-morbidities present at the time of diagnosis. The score ranges from grade 0 (no comorbidities) to 3 (severe comorbidities) (Kallogjeri et al., 2014). This score does not take into account the current acute state of the patient, but instead acts as a background marker of fitness. The IMD score provides a decile ranking of deprivation for each geographical area of 1,500 residents in the UK, where 1 is the most deprived and 10 is the least deprived. The score encompasses income, employment, education, health, including access to healthcare, crime, barriers to housing and services, and living environment to give an overall marker of deprivation (Ministry of housing, 2015).

3.5.2.2 Treatment received

The treatment decision as an intention to treat (ITT) at the time of the MDT was recorded. This was defined as either primary debulking surgery (PDS) with adjuvant chemotherapy, neoadjuvant chemotherapy (NACT) with interval debulking surgery (IDS), or no treatment. For PDS and IDS, the date of operation, operation received, extra procedures deviating from standard staging procedure, intra operative complications and cytoreductive outcomes were recorded. For medical therapy, date of commencing treatment, chemotherapy regime, number of cycles completed, and any additional medical management such as PARP inhibitors were recorded.

In order to create a 'patient timeline', the dates of first MDT, ITT decision MDT, date treatment commenced, date diagnosis confirmed, and date of death were recorded.

3.5.2.3 Disease effect

To ascertain the stage of disease, the disease burden at the time of presentation, and the effect this disease manifested on the clinical state of the patient, we recorded the FIGO stage of disease (Prat, 2015), patient blood parameters, and Eastern Cooperative Oncology Group performance status (ECOG-PS). These factors were chosen to represent surrogate markers of disease burden generally defined as "tumour effects".

ECOG-PS is a WHO recognised tool widely used as a measure of fitness for treatment in oncology patients. It is useful to assess the acute fitness of a patient but does not take into account co-existing co-morbidities. It is graded between 0 – 5, 0 being fully active and 5 being deceased (Su et al., 2015).

The blood parameters (haemoglobin, platelet, lymphocyte, neutrophil, albumin and CA 125) were recorded for both groups at initial presentation, to avoid any bias created by clinical intervention, such as blood transfusion. Although median albumin and haemoglobin levels decrease in an aging population (Vásárhelyi, 2017), these effects are small. Given that both groups had very similar age ranges, no adjustment was made. These bloods were also selected as they are a routine part of the established treatment pathway.

There remains a lack of consensus upon an accurate way to assess tumour volume or distribution pre-operatively. Therefore, the diagnostic CT scan reports, generated by specialist radiologists at the time of the MDT, were mined to generate a radiology score, based on presence or absence of disease in up to 30 anatomical sites, adapted from (M. J. Rutten et al., 2015; Son et al., 2017; Suidan et al., 2017).

3.5.3 TCGA patient cohort development

The cBio cancer genomics portal (cBioportal) is an open access web resource developed for exploring, visualising and analysing multidimensional cancer genomics data. It contains genomics information on over 5,000 tumour samples from over 20 cancer studies, including

316 HGSOC patients from the cancer genome atlas database. (Cerami et al., 2012; Gao et al., 2013; "Integrated genomic analyses of ovarian carcinoma," 2011).

In 2011, Nature published a comprehensive overview of the genomics of ovarian cancer, which included full exome DNA sequencing data for 316 HGSOC patients, the same patients that are contained within the cBioportal ("Integrated genomic analyses of ovarian carcinoma," 2011). All patients were assigned a unique TCGA patient identifier, and surgical outcome for a large percentage of the 316 patients was included in the supplementary data ("Integrated genomic analyses of ovarian carcinoma," 2011).

The cBioportal was searched to identify the unique identifiers for the 316 patients who demonstrated a defect in one of the 13 genes identified in the HR panel (BRCA1, BRCA2, RAD51, ATM, ATR, PALB2, FANCA, FACCI, FANCL, FANCD2, FANCE, FANCG, FANCM). The inclusion of PTEN was noted to be contentious as its exact role in the HR pathway remains controversial, a belief supported by several other recent studies (Bian et al., 2018; Huang et al., 2018; Hunt et al., 2012; "Integrated genomic analyses of ovarian carcinoma," 2011). For this reason, PTEN was excluded from the HR panel. Patients who showed a germline or somatic defect in one of the 13 genes were defined as being HRD, and patients who showed no defected in the genes included in the panel were defined as HRC.

The unique identifiers for all 316 patients were cross referenced with the clinical data provided in the supplementary materials to access patient demographics, surgical outcome and survival data for included patients. Patients for whom surgical data was not available were excluded from the analysis.

3.6 Prognostic Model to predict suboptimal surgical outcome

3.6.1 Patient and disease data collection

Patient and disease data were drawn from the MHGSOC database as per methods section, figure 3.3. All data were collected retrospectively using pre-defined definitions and data collection methods, as set out in the data collection guide, see appendix D. Data included patients operated on in a single centre tertiary unit between 2013 – 2018 (inclusive) by sub-

specialty trained gynaecological oncology surgeons. All CT scans were interpreted by consultant radiologists specialising in gynaecological oncology.

All predictors previously shown to have association with debulking rate or patient survival in the literature as demonstrated in chapter 4, that were also available in the MHGSOC database (figure 3.3) were collated for inclusion in the analysis. Predictor data fields were collected as per data collection guide in appendix D.

All continuous predictors were kept as continuous for analysis. Categorical predictors were kept as per their categories apart from IMD and ECOG-PS and ACE-27 which were grouped as per table 3.2.

Surgical outcome was defined by the lead surgeon at the end of the procedure as either complete (no macroscopic disease), optimal (disease <1cm remains), or suboptimal (disease ≥1cm remains). For the purpose of model development, the binary outcome of either good (complete cytoreduction) or bad (optimal and suboptimal cytoreduction) was used. All surgical procedures were performed as primary debulking procedures before the administration of any chemotherapy.

All included patients had HGSOC diagnosed post operatively by histopathologists specialising in gynaecological oncology after samples were assessed macroscopically, microscopically and had undergone immunohistochemistry staining for p53, WT-1, oestrogen receptor, PAX-8, CK7 and CK20.

Table 3.2 *Predictor data fields collected for use in model*

Predictor	Definition	Data type	Data fields	Grouped variables
Age	Age of patient at time of presentation	Continuous	NA	NA
ECOG-PS (Eastern co-operative Oncology Group Performance Status)	Patient fitness at time of presentation	Categorical	0 (No fitness restrictions)-5 (patient deceased)	Fit 0,1 Unfit 2,3,4
BMI (Body mass Index)	BMI at time of presentation (Height, weight and BMI all calculated)	Continuous	NA	NA
ACE-27 (Adult co-morbidity evaluation-27 index)	ACE-27 at time of presentation excluding diagnosis of HGSOC	Categorical	0 (no comorbidities)-3 (severe comorbidities)	NA
IMD (Indices of Deprivation)	IMD based upon patient postcode at time of presentation	Categorical	1 (most deprived)- 10 (least deprived)	Low 1-5 High 6-10
Stage of disease	FIGO Stage confirmed following completed staging taken from histology report and MDT summary	Categorical	IIIa, IIIb, IIIc, IVa, IVb	III- IIIa, IIIb, IIIc IV- Iva, IVb
CA125 (kU/L) Hemoglobin count (g/L) Platelet count (10^9/L) Lymphocyte count (10^9/L) Neutrophil count (10^9/L) Albumin level (g/L)	Level of CA125 first recorded in notes before any active treatment All blood parameter levels taken from time of pre-operative bloods i.e. blood test closest to, but before any active treatment commences	Continuous	NA	NA
Operating surgeon	The lead surgeon as per the operation notes	Categorical	1-5 one number assigned to each of the five consultant surgeons operating over this time scale	NA
CT scan disease distribution	CT scan taken from time of diagnostic work up closest to, but before any active treatment commences	Categorical	Chest disease (Yes,No) Includes all sites of disease above the diaphragm including thrombus Unremovable disease (Yes, No) Includes hydroureter and disease sites; gallbladder, root of SMA, adrenal, renal, pancreatic, liver, gastrosplenic ligament, lesser sac Nodal disease (Yes, No) Includes pelvic lymph nodes, para-aortic lymph nodes, inguinal lymph nodes Ascites	NA

ECOG-PS (Oken et al., 1982; Su et al., 2015), ACE-27 (Kallogjeri et al., 2014), IMD (Ministry of housing, 2015)

3.6.2 Statistical analysis

3.6.2.1 General statistics

All statistical analyses were performed using Prism 9 for Mac OS (version 9.2.0), WEKA (open source machine learning software), and R software package. Continuous variables with normal distribution are presented as mean and standard deviation, and non-normally distributed as median and inter-quartile (IQ) ranges. Comparisons between groups were performed using Students-t test for continuous data with normal distribution, Mann Whitney for non-normalised data and Fishers exact test for categorical data. All missing data were accounted for using multiple imputation. Univariable logistic regression was used to determine individual predictor association with surgical outcome. A p-value <0.05 was deemed significant throughout.

All models were generated with the aim of predicting bad outcome. Thus, specificity was defined as the number of patients achieving bad surgical outcome who were correctly identified divided by the total number of good surgical outcome patients. Positive predictive value corresponded to the number of true positives (bad surgical outcome) divided by the total number of patients predicted to have good surgical outcome, and Negative predictive value corresponded to the number of true negatives (good surgical outcome) divided by the total number of patients predicted to have bad surgical outcome. Accuracy was calculated as the sum of the true positives and true negatives divided by the total number of patients in the study.

3.6.2.2 Multivariable logistic regression model development

Logistic regression model was developed using multivariable logistic regression using Prism and WEKA with the inclusion of all pre-determined predictors, table 3.2. Following development, internal validation was performed using leave-one-out cross validation to account for overfitting of data. The area under the curve (AUC) for the ROC curve and 95% CI were used as model discrimination method. The Hosmer-Lemeshow (H-L) goodness of fit test was used to evaluate the calibration performance. The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy were calculated as above.

3.6.2.3 Machine learning model development

The Random Forest model was developed using WEKA with the inclusion of all pre-determined predictors, table 3.2. Both development and internal validation via leave-one-out cross validation were performed simultaneously to account for overfitting of data. The AUC was used as model discrimination method. The sensitivity, specificity, PPV, NPV, and accuracy were calculated as above to measure model calibration performance.

4 A systematic review of models predicting cytoreductive outcome of primary debulking surgery in HGSOC.

4.1 Introduction

The volume of remaining residual disease at the end of debulking surgery is well established as the most important independent prognostic marker of survival in advanced stage OC (Griffiths, 1975; Romanidis et al., 2014; Marianne Jetske Rutten et al., 2015). Surgery that is performed before chemotherapy (PDS) with no residual disease remaining translates to the best rates of overall survival, and this is reflected in the national guidance and treatment pathway. PDS should be the standard of care, with NACT/IDS reserved for cases where complete cytoreduction is unlikely (Fagotti et al., 2013; Gill et al., 2017; Horowitz et al., 2015; van Meurs et al., 2013).

Conversely, it is well recognised that patients who undergo PDS whose surgical outcome is suboptimal suffer significantly worse overall survival and increased morbidity when compared to those undergoing NACT/IDS (Gill et al., 2017; Horowitz et al., 2015). This means that not only does a bad surgical outcome negatively impact survival, it also increases morbidity and becomes a detrimental procedure. In order to reduce the morbidity and mortality of unsuccessful surgery, it is of utmost importance to achieve correct pathway selection for all patients.

The treatment pathway decision is currently made following discussion at the MDT. Current available data in the UK when following the national diagnostic pathway include patient co-morbidities, clinical state, biochemistry and CA 125 alongside TVUSS pelvis and CT of chest abdomen and pelvis. Although a large amount of data is available, decision-making relies heavily upon radiology, with patient factors often underrepresented in discussions (Scott et al., 2020). Despite this, tertiary led decision-making process, rates of suboptimal surgical outcome are very varied worldwide, and a large number of women undergo potentially non-beneficial major surgery with its associated morbidity.

Therefore, it is unsurprising that many prediction models have been proposed to help improve this decision-making process and reduce the increased morbidity and mortality associated with suboptimal surgical outcomes at the time of primary surgery. However, to date, none have achieved adequate external validation to justify their use in routine clinical practice.

Prediction models are increasingly gaining a place for routine use in many branches of clinical medicine (Bernard, 2017). Clinical prediction models act as adjuncts to help better inform decision-making. They are in common usage in many branches of medicine, but the most commonly used risk prediction model in ovarian cancer is the RMI, see figure 1.4. Their uses are varied, from predicting likelihood a patient may develop a disease in the future, or predicting the current likelihood they have the disease currently, to helping decide between multiple treatments based on predicted survival, or predict specific outcomes of treatment, such as surgical outcome (Chen, 2020). In healthcare, prediction models use predictors to estimate for an individual the probability that a condition or disease is already present (diagnostic model) or will occur in the future (prognostic model) (Moons, Kengne, Woodward, et al., 2012).

Steyerberg proposed a checklist for the development of prediction models, including three domains: general considerations, modelling steps, and validation (Chen, 2020). The prediction must have adequate clinical value, target a clear patient population and be easy to apply using ideally readily available clinical data in order to avoid increasing morbidity or delaying treatment with further testing. Prediction models in medicine can be developed using a variety of both traditional statistics and machine learning. For prognosis and prediction, regression models are the most commonly used and include linear, logistic and cox regression. A systematic review and meta-analysis comparing the use of machine learning with traditional statistics in prediction models concluded that one technique was not superior above the other (Christodoulou et al., 2019).

Before being considered for use in clinical practice, a model must be successfully validated. A true external validation should involve a fully separate patient cohort, to ensure the

predictive ability of the model translates across different populations and was not a result of overfitting of data in the original dataset (Moons, Kengne, Grobbee, et al., 2012).

Powering of studies in prediction models varies between the development and validation stages. When developing a prediction model, the minimum number of events per predictor parameter (EPP) is suggested to decide whether the sample size is enough. For models utilising logistic regression an EPP of ideally >20 has been suggested, but should as an absolute minimum be >10 (Moons, Kengne, Woodward, et al., 2012; Peduzzi et al., 1996). The sample number required in order to gain statistical significance for the independent validation of a pre-existing prediction model can be calculated based on number of events. The event is defined as the prediction outcome. A number of events >100 is considered optimal (Collins et al., 2014). However, in a review of 78 studies, 45% of studies had <100 events, 21% of studies did not report the number of events, and the range of events was between 6 – 42,408 (Collins et al., 2014). Although a number of events greater than 100 is optimal, this may not always be possible in practice.

There are three main frameworks that exist to determine the risk of bias in risk prediction in medicine. If the study is predictor finding i.e. identifying individual predictive or prognostic factors, then the Quality in Prognosis Studies (QUIPS) tool should be used (Hayden et al., 2013). If the study is prediction model development, then either the revised Risk of Bias in Nonrandomised Studies of Interventions tool (ROBINS-1) (Sterne et al., 2016) or the Prediction model risk of Bias Assessment Tool (PROBAST) should be used (Wolff et al., 2019). The PROBAST was developed specifically for assessing the risk of bias in models that combine multiple predictors to estimate risk for the presence of a particular condition, or the occurrence of a certain event in the future (Wolff et al., 2019). The tool includes the domains participants, predictors, outcome and analysis, as well as 20 signalling questions in order to comprehensively assess all steps in the development and validation of prediction models (Moons et al., 2019).

There are many published prognostic models in the literature to predict surgical outcomes in primary debulking surgery in advanced stage ovarian cancer. These models utilise a variety of modalities as predictors including clinical (age, BMI, ASA grade, ECOG-PS), biochemical (Hb,

platelet count, WCC, albumin CA 125), genomic and histological factors, as well as radiological (volume or distribution of disease on CT scan).

Despite the number of published models, none are routinely in usage in clinical practice. This review serves to systematically collate and compare all currently published prognostic models and attempt to identify limitations to these models, in order to make recommendations for future research.

4.2 Results

4.2.1 Study identification

Of 214 abstracts reviewed, 57 publications were identified for full text screening. Subsequently, 26 publications met eligibility criteria and were included in the final review (Abdallah, Chon, et al., 2015; Arab et al., 2018; Borley et al., 2015; Chesnais et al., 2017; Chong et al., 2019; Enshaei et al., 2015; Fago-Olsen et al., 2014; Feng et al., 2020; Feng et al., 2018; Fujwara et al., 2011; C. G. Gerestein et al., 2011; B. Gu et al., 2020; Y. Gu et al., 2020; Heitz et al., 2020; Horowitz et al., 2018; J. M. T. Janco et al., 2015; Karlsen et al., 2016; Kumar et al., 2019; Llueca et al., 2018; MacKintosh et al., 2014; Petrillo et al., 2015; Riester et al., 2014; Rutten et al., 2016; Son et al., 2017; Stashwick et al., 2011; Suidan et al., 2014).

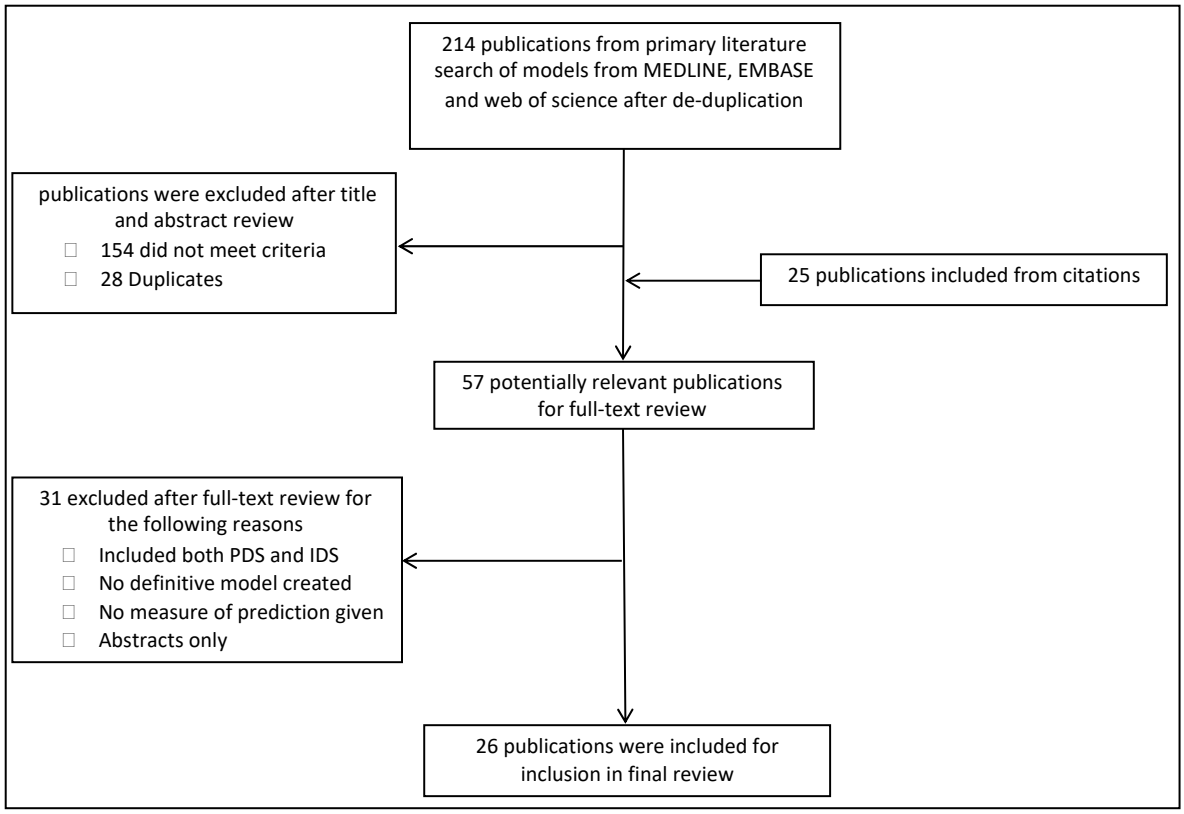


Figure 4.1 *PRISMA flow diagram showing selection of included studies*

The 26 publications all included the development and internal validation, as well as external validation of models predicting the debulking status at the time of primary surgery in patients with epithelial ovarian cancer. The publications describe the initial development of 27 prediction models, with 15 of these including internal validation and five external validation, as well as the external validation of 14 pre-existing models.

4.2.2 Patient and disease characteristics

The characteristics of patients and their disease included in the models can be seen in table 4.1. The majority of models were developed using data collected retrospectively in single centre cohort studies, with patient numbers varying vastly between 31 and 668. Fago-Olsen et al and Heitz et al used data from pre-existing RCTs, and Abdallah et al from data registries. Patient ages were well reported in the studies, ranging between 15 and 92 years. All patients included in all models had a diagnosis of primary FIGO stage I-IV epithelial ovarian cancer, with HGSOC being the most common histological subtype, and the majority of patients having stage III disease. All surgeries were performed by specialist gynae-oncology surgeons and

dates of procedures ranged between 1995 and 2020. Suboptimal cytoreduction rates varied widely between models with Janco et al reporting rates as low as 9% and Horrowitz et al as high as 76%.

Table 4.1 *General characteristics of included patient cohorts in systematic review*

Study	n	Age			Stage (FIGO)			Number of centers	Dates of surgery	Rates of suboptimal debulking
		Median	Range	Mean (SD)	Stages included	Stage III (%)	Stage IV (%)			
Abdallah 2015	124			62 (10)	III/IV	51%	49%	2		48%
Abdallah 2015	468				III/IV					29%
Abdallah 2015	190				III/IV					41%
Arab 2018	129		20-80	50 (12)	I-IV	54%	6%	1	2007-2017	28%
Borley 2014	181	63	22-92		III/IV	66%	34%	1	2001-2012	43%
Chesnais 2017	247			62 (12)	I-IV	49%	31%	1	2008-2013	62%
Chesnais 2017	47			59 (13)	I-IV	27%	7%	1	2009-2015	
Chong 2019	51	58	37-74		III/IV	72%	28%	1		33%
Enshaei 2015	668	67			I-IV	56%	10%	1	1995 - 2005	24%
Fago-Olsen 2014	238	65	55-74		III/IV	80%	20%	1	2004-2012	42%
Feng 2018 (Suidan)	161	57	27-77		III/IV	81%	19%	1	2015-2017	18%
Feng 2018 (Aletti)	110	57	27-78		III/IV	81%	19%	1	2015-2018	18%
Feng 2018 (PCI)	110	57	27-79		III/IV	81%	19%	1	2015-2019	18%
Feng 2018 (Eisenkop)	110	57	27-80		III/IV	81%	19%	1	2015-2020	18%
Feng 2018 (Fagotti)	39	57	27-81		III/IV	81%	19%	1	2015-2021	18%
Feng 2020	83			53 (10)	III/IV	89%	11%	1	2012-2018	38%
Fujwara 2011 Model 1	98		33-79	54	I-IV	30%	5%	1		12%
Fujwara 2011 Model 2	98		33-80	54	I-IV	30%	5%	1		12%
Gerestein 2011	115	62.4	15-83		III/IV	81%	18%	6	2005-2008	54%
Gu 2020 (Suidan)	31	57	38-76		III/IV	70%	30%	1	2016-2017	65%
Gu 2020	296			54	III/IV			9	2016-2019	15%
Heitz 2020 (Reister)	266		39-73	59	III/IV	78%	22%	multiple		61%
Heitz 2020 (Liu)	266		39-74	59	III/IV	78%	22%	multiple		61%
Heitz 2020 (Tucker)	266		39-75	59	III/IV	78%	22%	multiple		61%
Heitz 2020	266		39-76	59	III/IV	78%	22%	multiple		61%
Horowitz 2018	1480		24-87	58	III/IV			73	2001-2004	76%
Janco 2015	279			64	III/IV	75%	25%	1	2003-2008	9%
Janco 2015	279			64	III/IV	75%	25%	1	2003-2009	9%
Karlsen 2016	150	65	41-89		III/IV	79%	21%	1	2004-2010	59%
Kumar 2019 (Suidan)	276	64	21-91		I-IV	93%	0%	1	2003-2011	51%
Llueca 2018 (PCI)	80	59	30-84		III/IV	69%	31%	1	2013-2016	13%
Mackintosh 2014	91	62	25-83		III/IV	79%	21%	1	1995-2003	65%
Mackintosh 2014	35	62	38-84		III/IV	85%	14%	1	2005-2007	65%
Petrillo 2015	234	57	25-84		III/IV	82%	18%	1	2007-2014	43%
Reister 2015	179				III/IV			1		24%
Rutten 2016 (Ferrandina A)	151	63	30-88		III/IV	79%	21%	7	2000-2009	53%
Rutten 2016 (Ferrandina B)	151	63	30-89		III/IV	79%	21%	7	2000-2009	53%
Rutten 2016 (Gerestein)	151	63	30-90		III/IV	79%	21%	7	2000-2009	53%
Son 2017	327	57	25-70		III/IV	80%	20%	1	2007-2015	35%
Stashwick 2011	106		39-82	61	III/IV	73%	27%	1	2005-2010	43%
Suidan 2014	350	61	34-86		III/IV	73%	27%	multiple	2001-2012	25%

4.2.3 Modalities and predictors included in models

The models included in the review utilised different modalities as predictors, outlined in detail in table 4.2. The majority of predictors included would be available pre-operatively as part of the UK diagnostic pathway: patient age, BMI, ECOG PS, serum platelet count, albumin and CA 125 levels and sites of disease seen on CT scan. Other predictors used would not be routinely available in the UK under current guidelines but could be achieved without invasive testing: HE4, transferrin and β2-macroglobulin serum levels, specific germline gene panels and PET CT scanning. Some models utilised predictors that would require pre-operative biopsy of tumour: tumour protein panels, histology type and somatic gene panels. Petrillo et al developed their model based upon a scoring system at the time of laparoscopy to be performed as a triaging procedure before a debulking laparotomy, and the Aletti, Eisenkop and PCI models externally validated by Feng et al can only be performed at the time of laparotomy.

The majority of models (71%) aimed to predict suboptimal cytoreduction at the time of surgery, with only 12 models predicting a favourable outcome of complete cytoreduction. All models defined suboptimal cytoreduction as disease $\geq$1cm remaining at the end of the procedure. Of the 12 models predicting the favourable outcome at the time of surgery, all but two models defined complete cytoreduction as no remaining macroscopic disease. Gu et al and Stashwick et al defined a favourable outcome as residual disease <1cm.

Table 4.2 *Modalities and predictors used in included models*

Study	Modalities included	Number of predictors in model	Final predictors used	Surgical outcome predicted
Abdallah 2015	Genomic	58	gene expression data (58 genes included*)	suboptimal
Abdallah 2015	Genomic	58	gene expression data (58 genes included*)	suboptimal
Abdallah 2015	Genomic	58	gene expression data (58 genes included*)	suboptimal
Arab 2018	Biomarker, radiology	3	CA125>420, CT-ascites, liver metastasis	suboptimal
Borley 2014	Radiology	2	CT-ascites, mesentery upper abdominal disease	suboptimal
Chesnais 2017	Biomarker, radiology	3	CA125 >100 BMI >30 CT- parenchymal metastasis	suboptimal
Chesnais 2017	Biomarker, radiology	3	CA125 >100 BMI >30 CT- parenchymal metastasis	suboptimal
Chong 2019	Clinical, Radiology	2	ECOG status, FDG PET/CT metabolic uptake score	suboptimal
Enshaei 2015	Clinical, biomarker	5	age, stage, grade, histological type, CA 125	suboptimal
Fago-Olsen 2014	Clinical, biomarker	4	Transferrin, β2- macroglobulin, age, CA 125	complete
Feng 2018 (Suidan)	Clinical, biomarker, radiology	9	Age, CA 125, ASA, CT- aortic lymph nodes, small bowel thickening, small bowel mesenteric disease, mesenteric artery disease, peri-splenic disease, lesser sac disease	complete
Feng 2018 (Aletti)	Surgical-laparotomy	1	Surgical complexity score at laparotomy	complete
Feng 2018 (PCI)	Surgical-laparotomy	1	Tumour bulk at laparotomy	complete
Feng 2018 (Eisenkop)	Surgical-laparotomy	1	Tumour bulk at laparotomy	complete
Feng 2018 (Fagotti)	Surgical-laparoscopy	1	Laparoscopy score	complete
Feng 2020	Clinical, biomarker	3	Age, HE4, CA 125	suboptimal
Fujwara 2011 Model 1	Radiology	3	CT-pelvic lymph, cul-de-sac tumours, retroperitoneal tumours	suboptimal
Fujwara 2011 Model 2	Radiology	3	CT- DPT, bowel mesenteric tumours, bowel encasement	suboptimal
Gerestein 2011	Clinical, radiology	3	Platelets, CT- DPT, ascites	suboptimal
Gu 2020 (Suidan)	Clinical, radiology	9	Age, CA 125 ASA PET CT- aortic lymph nodes, small bowel thickening, small bowel mesenteric disease, mesenteric artery disease, peri splenic disease, lesser sac disease	suboptimal
Gu 2020	Clinical, radiology	6	Age, BMI, CT- bowel metastasis, spleen metastasis, diaphragmatic metastasis, lymph nodes	complete
Heitz 2020 (Reister)	Genomic	7	gene expression data (POSTN, CXCL14, FAP, NUAK1, PTCH1, TGFBR2, TNFAIP6)	suboptimal
Heitz 2020 (Liu)	Genomic	11	gene expression data (POSTN, FAP, TIMP3, COL11A1, EDNRA, CTSK, COL5A2, TNFAIP6, TMEM158, MMP11, CXCL14)	suboptimal
Heitz 2020 (Tucker)	Genomic	2	gene expression data (FABP4 ADH1B)	suboptimal
Heitz 2020	Genomic	126	gene expression data (126 genes included*)	suboptimal
Horowitz 2018	Clinical, radiology	5	disease score, stage, age, CA 125, CT- ascites	complete
Janco 2015	Clinical, radiology	4	Age, CT- no ascites, omental cake, DPT	complete
Janco 2015	Clinical, radiology	3	ECOG >2, CT- DPT, lymphadenopathy	suboptimal
Karlsen 2016	Clinical, biomarker	3	HE4, age, PS	complete
Kumar 2019 (Suidan)	Clinical, radiology	9	Age, CA 125 ASA PET CT- aortic lymph nodes, small bowel thickening, small bowel mesenteric disease, mesenteric artery disease, peri splenic disease, lesser sac disease	suboptimal
Llueca 2018 (PCI)	Radiology	1	Tumour bulk score at CT	suboptimal
Mackintosh 2014	Radiology	3	CT- para aortic nodes, liver surface disease	suboptimal
Mackintosh 2014	Radiology	3	CT- para aortic nodes, liver surface disease	suboptimal
Petrillo 2015	Surgical-laparoscopy	1	Score at laparoscopy	complete
Riester 2015	Biomarker	3	protein signature (POSTN, CXCL14, phosphorylated Smad2/3)	Suboptimal
Rutten 2016 (Ferrandina A)	Clinical, radiology	5	PS, CT- bowel mesentery disease, diaphragmatic metastasis, aortic lymph npdes, DPT	suboptimal
Rutten 2016 (Ferrandina B)	Clinical, radiology	5	PS, CT- bowel mesentery disease, diaphragmatic metastasis, liver disease, omental disease	suboptimal
Rutten 2016 (Gerestein)	Clinical, radiology	3	Platelet count, CT- DPT, ascites	suboptimal
Son 2017	Clinical, radiology	5	PS, CT- DPT, bowel mesentery, lymph nodes, pleural effusion	suboptimal
Stashwick 2011	Clinical, radiology	4	Albumin, CT- DPT, lymph nodes, splenic disease, bowel mesentery	complete
Suidan 2014	Clinical, radiology	9	Age, CA 125 ASA PET CT- aortic lymph nodes, small bowel thickening, small bowel mesenteric disease, mesenteric artery disease, peri splenic disease, lesser sac disease	suboptimal

*full list of genes included in gene panel provided in appendix

Predictors included in models can be grouped as patient and tumour characteristics, biochemical markers, radiological predictors, or surgical predictors.

4.2.3.1 Patient and disease characteristics

Age was the most common patient characteristic utilised in models, demonstrating significant association with surgical outcome by many studies. The majority of models kept this predictor as a continuous variable, however Gu et al and Feng et al dichotomised the predictor by defining cut off points of 60 and 69 years, respectively. ECOG PS was shown to be associated with both suboptimal and complete cytoreduction and was included in four models, with BMI also utilised by Chesnais et al who again dichotomised this continuous variable.

Table 4.3 *Patient characteristics individually associated with surgical outcome*

Patient characteristic	Study	p value	Associated debulking outcome
Age	Janco	0.003	Complete
Age	Karlsen	<0.002	Complete
Age	Fago-Olsen	0.005	Suboptimal
Age	Horowitz	0.01	Suboptimal
Age >60	Gu	0.016	Suboptimal
Age >69	Feng	0.042	Suboptimal
ECOG PS >2	Chong	0.052	Suboptimal
ECOG PS >2	Janco	0.004	Suboptimal
ECOG PS >2	Son	0.025	Suboptimal
ECOG PS ≤3	Karlsen	<0.001	Complete
BMI >30	Chesnais	<0.01	Suboptimal
Stage	Horowitz	0.009	Suboptimal

Tumour characteristics were utilised most commonly in models focussing on gene expression data, with Abdallah et al and Heitz et al developing gene panels, and Riester et al developing a three-protein panel based upon a gene panel. Two models used stage of disease as predictors and only one model developed by Enshaei et al included the tumour histology and grade in their model.

4.2.3.2 Biochemical biomarkers

Table 4.4 *Biochemical biomarkers individually associated with surgical outcome*

Biochemical marker	Study	p value	Associated debulking outcome
CA125	Horowitz	<0.001	Suboptimal
CA125	Fago-Olsen	0.014	Suboptimal
CA125	Gerestein	0.199	Suboptimal
CA125> 420	Arab	<0.001	Suboptimal
CA125> 100	Chesnais	<0.01	Suboptimal
CA125 > 800	Gu	0.033	Suboptimal
CA125 > 313	Feng	0.037	Suboptimal
HE4	Karlsen	<0.001	Complete
HE4> 777	Feng	0.007	Suboptimal
Transferrin	Fago-Olsen	0.0014	Suboptimal
β2–macroglobulin	Fago-Olsen	0.046	Suboptimal
platelet count	Gerestein	0.033	Suboptimal

CA 125 is the most commonly included biochemical biomarker in the models, with four of the models dichotomising this continuous variable, with cut off levels varying between 100 – 800. Two models included HE4 as a predictor, Karlsen et al keeping this as a continuous variable and Feng et al using 777 as a cut off value. Fago-Olsen et al found transferrin and β2-macroglobulin to be associated with surgical outcome and included these in their model. Only Gerestein et al used the common haematological marker of platelet count as a predictor in their model. All studies specify that biochemical markers are to be taken via serum samples pre-operatively as close to the time of the surgical procedure as is practical, however only Suidan et al and Karlsen et al set a definitive time scale, both specifying their samples were collected within 14 days of surgery.

4.2.3.3 Radiological predictors

Table 4.5 *Disease sites on CT and PET-CT individually associated with surgical outcome*

CT disease site	Study	p value	Associated debulking outcome
Ascites	Arab	0.01	Suboptimal
Ascites	Gerestein	0.0385	Suboptimal
Ascites	Horowitz	<0.001	Suboptimal
Diffuse peritoneal thickening	Janco	0.003	Suboptimal
Diffuse peritoneal thickening	Gerestein	0.0074	Suboptimal
Diffuse peritoneal thickening	Gu	0.046	Suboptimal
Diffuse peritoneal thickening	Son	0.01	Suboptimal
Diffuse peritoneal thickening	Fujwara	0.006	Suboptimal
Omental metastasis	Chesnais	<0.01	Suboptimal
Omental metastasis	Gu	0.008	Suboptimal
Omental metastasis	Janco	<0.001	Suboptimal
Diaphragmatic metastasis	Chesnais	<0.01	Suboptimal
Diaphragmatic metastasis	Gu	0.014	Suboptimal
Diaphragmatic metastasis	Son	0.004	Suboptimal
Pelvic bowel metastasis	Gu	0.007	Suboptimal
Abdominal bowel metastasis	Gu	0.034	Suboptimal
Bowel mesenteric metastasis	Son	0.007	Suboptimal
Bowel encasement tumours	Fujwara	0.002	Suboptimal
Liver surface disease	Mackintosh	<0.0001	Suboptimal
Parenchymal liver metastasis	Arab	0.041	Suboptimal
Parenchymal liver metastasis	Gu	0.005	Suboptimal
Parenchymal metastasis	Chesnais	<0.01	Suboptimal
Liver surface metastasis	Gu	<0.001	Suboptimal
Splenic metastasis	Gu	<0.001	Suboptimal
cul-de-sac metastasis	Fujwara	0.005	Suboptimal
Lymphadenopathy above inferior mesenteric artery	Gu	<0.001	Suboptimal
Lymphadenopathy	Janco	0.03	Suboptimal
Pelvic lymph nodes	Fujwara	0.007	Suboptimal
Supra renal lymphadenopathy	Son	0.008	Suboptimal
Para-aortic lymphadenopathy	Mackintosh	0.0065	Suboptimal
Pleural effusion	Son	0.02	Suboptimal

Radiology was the most commonly included modality across all studies, with 20/41 (49%) including the findings of pre-operative CT scans in their models and four studies including findings of PET-CT scans. All studies specified that scans were to be performed pre-operatively and were interpreted by radiologists trained in gynae-oncology. Scan-surgery time scales were often not reported, with only eight studies giving definitive timings, ranging from as long as 3 months (Rutten et al) to 21 days pre-operatively (Son et al). Table 4.5 demonstrates the different disease sites reported on CT found to have an association with surgical outcome, with diffuse peritoneal thickening being the most commonly included site in 12% of models.

4.2.3.4 Surgical predictors

Surgical predictors by way of triaging laparoscopy were described in by Feng et al and Petrillo et al. Both these models describe a scoring system (Predictive index value- PIV) at time of laparoscopy, calculated by the presence or absence of: omental cake, extensive peritoneal metastasis, diaphragmatic metastasis, bowel disease, stomach and/or spleen and/or lesser omentum metastasis and superficial liver metastasis. The scoring system is then used to triage patients towards either the PDS or NACT/IDS treatment pathways. Feng et al also describes three surgical scoring systems all performed at the time of staging laparotomy: Eisenkop score, peritoneal cancer index (PCI) and Aletti score. The Eisenkop score and the PCI both aim to quantify tumour bulk, and Aletti surgical complexity at the start of the procedure and at this point triage whether completion of surgery would result in good surgical outcome or not. All surgical scoring in all models was performed by specialist gynae-oncologists in tertiary level centres.

4.2.4 Model development and performance

4.2.4.1 Model development and reporting methods

Table 4.6 demonstrates the method of model development, their level of validation and methods alongside model performance. Of the 27 developed models identified by the review, all were developed via multivariable logistic regression modelling techniques, with the exception of one (Enshaei et al., 2015), who used the non-logistic artificial neural network (ANN) method. All but two (Borley et al., 2015; Heitz et al., 2020) of the 26 models selected the final included predictors for their models by applying univariable analysis to all candidate predictors prior to the multivariable analysis. Overfitting and model optimism were accounted for in all validated models, using a variety of techniques including bootstrapping, cross validation, splitting of data sets and the leave-one-out method.

Model discrimination was described in the majority of cases as area under the curve (AUC) of receiver-operating characteristic (ROC) curve. Three models (Chesnais et al., 2017; Fujwara et al., 2011; MacKintosh et al., 2014) did not report an AUC but instead reported the specificity

and sensitivity of their models. Model calibration was discussed and reported less commonly, however of the eight studies reporting its usage all demonstrated good concurrence using Hosmer-Lemeshow goodness-of-fit test (Abdallah, Chon, Zgheib, et al., 2015; Chesnais et al., 2017; Fago-Olsen et al., 2014; Feng et al., 2020; Y. Gu et al., 2020; J. M. T. Janco et al., 2015; Karlsen et al., 2016; Riester et al., 2014).

Table 4.6 *Model development methods and performance*

Model	Validation	Development method	AUC	95% CI		Sensitivity	Specificity
				Lower	Upper		
Arab 2018	Development	logistic regression	0.874	0.815	0.934		
Gu 2020	Development	logistic regression	0.788	0.72	0.856	89.10%	52.40%
Stashwick 2011	Development	logistic regression	0.9	0.87	0.93	94%	75%
Suidan 2014	Development	logistic regression	0.758				
Chong 2019	Development	logistic regression	0.775	0.64	0.88	82.40%	64.70%
Fujwara 2011 model 1	Development	logistic regression				96.5%%	50%
Fujwara 2011 model 2	Development	logistic regression				98.80%	50%
Abdallah 2015	Internal	logistic regression	0.73			73%	66%
Borley 2014	Internal	logistic regression	0.721	0.594	0.857	50%%	68%
Chesnais 2017	Internal	logistic regression	0.79	0.73	0.86	93.50%	91.4
Enshaei 2015	Internal	Artificial Neural network	0.75				
Fago-olsen 2014	Internal	logistic regression	0.64			70%	68%
Feng 2020	Internal	logistic regression	0.71	0.54	0.88	100%	44%
Gerestein 2011	Internal	logistic regression	0.67				
Horowitz 2018	Internal	logistic regression	0.73	0.79	0.88	80%	76%
Janco 2015	Internal	logistic regression	0.748			88.50%	45.80%
Janco 2015	Internal	logistic regression	0.685				
Karlsen 2016	Internal	logistic regression	0.839				
Riester 2014	Internal	logistic regression	0.89	0.84	0.93		
Son 2017	Internal	logistic regression	0.758	0.685	0.809	70%	73%
Heitz 2020	Internal	logistic regression	0.54				
Mackintosh 2014	Internal	logistic regression				91.4%	59.4%
Abdallah 2015	External		0.57				
Abdallah 2015	External		0.62				
Feng 2018 Suidan validation	External		0.548	0.436	0.659		
Gu 2020 Suidan validation	External		0.79	0.633	0.961		
Kumar 2019 Suidan validation	External		0.653	0.532	0.773		
Heitz 2020 Reister validation	External		0.5				
Heitz 2020 Liu validation	External		0.62				
Heitz 2020 Tucker validation	External		0.53				
Rutten 2016 Ferrandina A validation	External		0.59				
Rutten 2016 Ferrandina B validation	External		0.58				
Rutten 2016 Gerestein validation	External		0.66				
petrillo 2015	External		0.885				
Feng 2018 Aletti validation	External		0.524	0.413	0.634		
Feng 2018 PCI validation	External		0.797	0.713	0.882		
Feng 2018 Eisenkop validation	External		0.808	0.728	0.887		
Feng 2018 Fagotti validation	External		0.713	0.527	0.9		
Chesnais 2018	External						
Mackintosh 2014	External					41.70%	73.90%
Llueca 2018 PCI validation	External						

4.2.4.2 Model performances and validation

Model performance varied greatly between studies. Figure 4.2 demonstrates the varying AUC values between models, and the performance of models not reporting AUC values are also shown in table 4.6. The models at development stage reported AUC values between 0.73 and 0.90, models that had been internally validated demonstrated AUC values between 0.54 and 0.89, and models externally validated demonstrated AUC values between 0.50 and 0.88. Figure 4.2 demonstrates a general trend of decreasing AUCs with increasing levels of validation. Suidan et al demonstrated an AUC of 0.75 with a CT based model when originally developed and was externally validated twice, once successfully by Gu et al (AUC 0.79) and once unsuccessfully by Feng et al (AUC 0.54). Reister et al reported the most successful internally validated model with their three-protein signature, based upon a gene panel (AUC 0.89). The gene panel was validated unsuccessfully by Heitz et al (AUC 0.53), however the three-protein signature has not yet been externally validated. Although most models do show a decrease in performance with an increasing level of validation, the surgical based models demonstrate this less so. Petrillo et al demonstrate the highest-level performance with their externally validated laparoscopic scoring system (AUC 0.88). However, although Petrillo et al reported a good performance, when the same scoring system was externally validated by a separate surgical team, Feng et al, performance of the model dropped to an AUC of 0.71. The second and third best performing surgical models (Feng et al validation of Eisenkop and PCI scoring systems, AUC 0.79 and 0.80, respectively) both aimed to quantify disease bulk at the time of laparotomy, and therefore require major surgery for their usage.

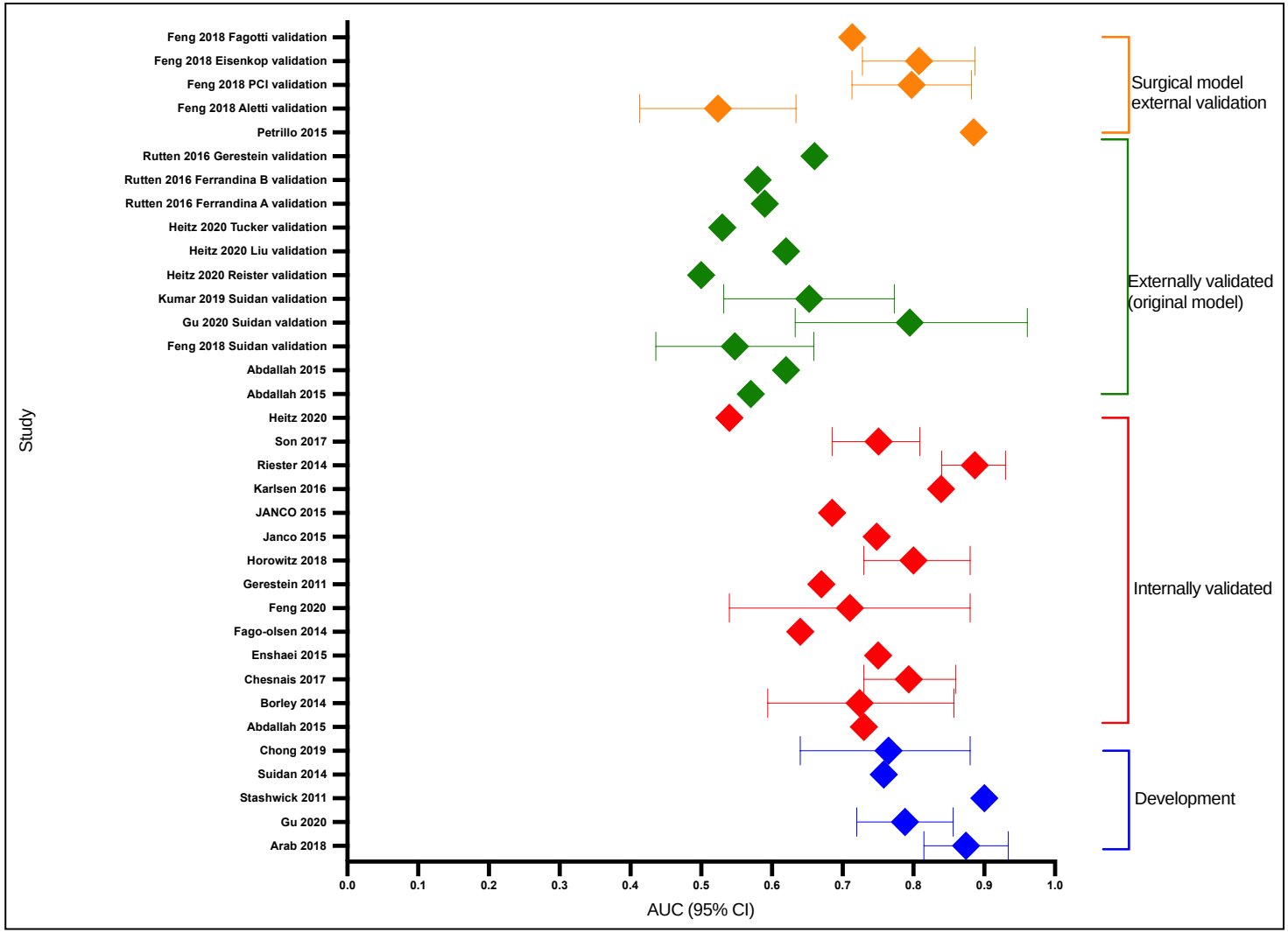


Figure 4.2 *Box plot illustrating performance of all models included in review*
AUC- area under curve, CI- confidence intervals. 95% confidence intervals shown when provided by studies. Figure excludes six models whose AUCs are not reported.

4.2.5 Risk of Bias assessment

The risk of bias (ROB) for all studies was assessed using the Prediction model risk Of Bias Assessment Tool (PROBAST). This tool is designed specifically for prediction models in health and is suitable to assess prognostic models. The PROBAST guides the systematic assessment of ROB by assessing all aspects of the model separately (participants, predictors, outcome, analysis) then combining to give an overall ROB. All areas are scored as either low ROB (L), high ROB (H) or unclear ROB (U). Figure 4.3 outlines the PROBAST assessment for each of the models included in this review.

STUDY	ROB				APPLICABILITY			OVERALL	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	ROB	Applicability
Abdallah 2015 (D)	L	L	L	H	L	L	L	H	L
Abdallah 2015 (EV)	L	L	L	U	L	L	L	U	L
Abdallah 2015 (EV)	L	L	L	U	L	L	L	U	L
Arab 2018 (D)	L	L	L	H	L	L	L	H	L
Borley 2014 (IV)	L	L	L	H	L	L	L	H	L
Chesnais 2017 (IV)	L	L	L	H	L	L	L	H	L
Chesnais 2017 (EV)	L	L	L	H	L	L	L	H	L
Chong 2019 (D)	L	L	L	H	L	L	L	H	L
Enshaei 2015 (IV)	L	L	L	U	L	L	L	U	L
Fago-Olsen 2014 (IV)	L	L	L	H	L	L	L	H	L
Feng 2018 (Suidan EV)	L	L	L	H	L	L	L	H	L
Feng 2018 (Aletti EV)	L	H	L	H	L	H	L	H	H
Feng 2018 (PCI EV)	L	H	L	H	L	H	L	H	H
Feng 2018 (Eisenkop EV)	L	H	L	H	L	H	L	H	H
Feng 2018 (Fagotti EV)	L	L	L	H	L	L	L	H	L
Feng 2020 (IV)	L	L	H	H	L	L	H	H	H
Fujwara 2011 (D)	L	L	L	H	L	L	L	H	L
Fujwara 2011 (D)	L	L	L	H	L	L	L	H	L
Gerestein 2011 (IV)	L	L	L	H	L	L	L	H	L
Gu 2020 (D)	L	L	L	H	L	L	L	H	L
Gu 2020 (Suidan EV)	L	L	L	H	L	L	L	H	L
Heitz 2020 (Reister EV)	L	L	L	L	L	L	L	L	L
Heitz 2020 (Liu EV)	L	L	L	L	L	L	L	L	L
Heitz 2020 (Tucker EV)	L	L	L	L	L	L	L	L	L
Heitz 2020 (IV)	L	L	L	L	L	L	L	L	L
Horowitz 2018 (IV)	L	L	L	H	L	L	L	H	L
Janco 2015 (IV)	L	L	L	L	L	L	L	L	L
Janco 2015 (IV)	L	L	L	H	L	L	L	H	L
Karlsen 2016 (IV)	H	L	L	L	H	L	L	H	H
Kumar 2019 (Suidan EV)	L	L	L	H	L	L	L	H	L
Llueca 2018 (PCI EV)	L	L	L	H	L	L	L	H	L
Mackintosh 2014 (IV)	L	L	L	H	L	L	L	H	L
Mackintosh 2014 (EV)	L	L	L	H	L	L	L	H	L
Petrillo 2015 (EV)	L	L	L	L	L	L	L	L	L
Riester 2015 (IV)	L	L	L	L	L	L	L	L	L
Rutten 2016 (Ferrandina A EV)	L	L	L	L	L	L	L	L	L
Rutten 2016 (Ferrandina B EV)	L	L	L	L	L	L	L	L	L
Rutten 2016 (Gerestein EV)	L	L	L	L	L	L	L	L	L
Son 2017 (IV)	H	L	L	H	H	L	L	H	H
Stashwick 2011 (D)	L	L	L	H	L	L	L	H	L
Suidan 2014 (D)	L	L	L	H	L	L	L	H	L

Figure 4.3 *Risk of Bias (ROB) for all included models using PROBAST assessment*

Each domain is assessed for ROB; Low ROB (L), High ROB (H), Unclear ROB (U) as well as applicability in practice and given an overall score. Each model is shown along with the level of validation of the model; Development (D), Internal validation (IV), External validation (EV).

4.2.5.1 Domain one: Participants

All models included in the review had appropriate sources of data for their patient selection. The majority used data from cohort studies, with two models using patient data from RCTs (Abdallah, Chon, et al., 2015; Fago-Olsen et al., 2014; Heitz et al., 2020) and one model from data registries (Abdallah, Chon, et al., 2015). The majority of models had appropriate inclusion and exclusion criteria for their patient inclusion and therefore scored low risk of bias for this category. Karlsen et al scored high risk however, as they excluded all patients with an ECOG-PS ≥4, and Son et al also scored high risk as they excluded patients with an ECOG-PS ≥2 and patients over the age of 80 years. These exclusions potentially eliminate the sickest patients from their models which could potentially create bias and affect the applicability of the study as in practice these patients are not always excluded as surgical candidates.

4.2.5.2 Domain two: Predictors

All models both defined and assessed the predictors in a similar way for all participants and made predictor assessments without knowledge of the outcome. For the majority of models, the predictors would be available at the time the models are designed to be used, i.e. pre-operatively. As already discussed, some predictors used require additional testing not routinely performed under current UK guidance, such as additional tumour biopsy or serum blood sample testing. However, image guided biopsy or laparoscopic assessment is occasionally undertaken in current clinical practice. Three of the surgical prediction models undertaken by Feng et al: the external validation of the PCI, Aletti and Eisenkop scores, require laparotomy in order for the prediction to be made. This would suggest additional morbidity and mortality and therefore the three models were scored as high ROB and high ROB for applicability in clinical practice.

4.2.5.3 Domain three: Outcome

Outcome was well defined in all models included. As discussed, the majority of the models aimed to predict suboptimal surgical outcome, with 12/41 predicting complete debulking. All models gave clear definitions for surgical outcome except one (Feng et al., 2020), which whilst

did clearly state the aim of the model to be prediction suboptimal debulking, did not define the surgical outcome clearly. For this reason, this model has a high ROB in this category.

4.2.5.4 Domain four: Analysis

Statistical analysis is a critical part of prognostic prediction model development and validation. Model development and validation studies can include many steps where flawed methods can distort results. As part of the PROBAST assessment models must show the following: robust methodology in sample size; appropriate handling of categorical and continuous predictors and missing data; appropriate selection of predictors; use of relevant performance measures; and account adequately for overfitting and optimism. Ten models demonstrated robust methodology in all categories and were therefore deemed low ROB in this area (Heitz et al., 2020; J. M. T. Janco et al., 2015; Karlsen et al., 2016; Petrillo et al., 2015; Riester et al., 2014; Marianne Jetske Rutten et al., 2015). The remaining 31 models were all deemed high risk for ROB.

With respect to sample size, model development ideally requires >20 events per predictor (EPP), with >10 events adequate if 20 is not achievable. Eight of the development models did not provide information to assess the number of EPV (Abdallah, Chon, et al., 2015; Chesnais et al., 2017; Chong et al., 2019; Fago-Olsen et al., 2014; Heitz et al., 2020; MacKintosh et al., 2014; Riester et al., 2014; Stashwick et al., 2011), and three showed <10 EPV for the included predictors (Fujwara et al., 2011; Son et al., 2017). Of the remaining development models, five demonstrated >20 EPV for all predictors (Feng et al., 2020; Cornelis G. Gerestein et al., 2011; Y. Gu et al., 2020; Karlsen et al., 2016), with the remainder of models having between 10 – 20 EPV.

The handling of categorical and continuous predictors was assessed. A large number of studies demonstrated dichotomisation of continuous predictors such as age, BMI, CA 125 and HE4. Dichotomisation of continuous predictors requires choosing an arbitrary cut off point above which values are classified as high and below which values are classified as low. Although this is often carried out to aid clinical interpretation, dichotomisation can lead to the loss of information and therefore the introduction of bias (Moons et al., 2019). The most

appropriate handling of missing data in prediction models is the use of multiple imputation, leading to a reduced level of bias in results (Moons, Kengne, Woodward, et al., 2012). Of the assessed models, 34 studies did not address missing data. When not addressed, the most common explanation was that patients with missing data were excluded from analysis, as many statistical packages automatically exclude patients with missing data fields (Moons et al., 2019). Six models specified that patients with missing data were excluded from analysis (Chesnais et al., 2017; Y. Gu et al., 2020; Rutten et al., 2016), and only one model specified the use of multiple imputation to account for missing data (Cornelis G. Gerestein et al., 2011). When a model is developed using a small data set, or if the EPP is not adequate, the model has high ROB due to overfitting of the data (Moons et al., 2019). To reduce this risk, all models require internal validation. Of the development models, seven (Arab et al., 2018; Chong et al., 2019; Fujwara et al., 2011; Y. Gu et al., 2020; Stashwick et al., 2011; Suidan et al., 2014) were not internally validated and therefore present with high ROB.

Most models sourced data using cohort studies, therefore datasets were likely to include desirable features to be used as candidate predictors. To allow for ease in clinical practice, many researchers aim to reduce the number of predictors required during model development, to produce a simplified model. Researchers often use univariable analysis of predictors as a triage step in model building, by only including predictors that reach a statistically significant univariable association. This approach can lead to incorrect predictor selection, as predictors are chosen on the basis of their statistical significance as a single predictor, rather than in context with other predictors (Moons et al., 2019). The method of selection of predictors was well described by all included development models. As discussed, all but one model (Enshaei et al., 2015) used multivariable logistic regression for model development. Of these models only two (Borley et al., 2015; Heitz et al., 2020) avoided the use of univariable analysis as a triage step.

When assessing the performance of a prognostic prediction model, both model calibration and discrimination must be assessed. The most common methods used for calibration is the Hosmer-Lemeshow test and for discrimination the AUC of ROC curve (Moons, Kengne, Woodward, et al., 2012). All included models provided a method of model discrimination,

with AUC most commonly employed (table 4.6). However, only nine models provided information on method of model calibration.

Taking all described domains into consideration, only 10 of the included models were deemed to have an overall low ROB when assessed with PROBAST.

4.3 Discussion

The ability to determine pre-operatively which patients will undergo complete debulking at the time of primary surgery would allow for the correct treatment pathway for each patient and in turn reduce morbidity and mortality as well as increasing survival. For a prediction model to be successful in clinical practice it must be clinically required, easy to use with the reliance upon readily available data and be successfully validated to allow it to be applicable across different populations. Despite many published models in this field, none are currently used in routine clinical practice.

This systematic review identified 26 publications describing varying levels of development and validation of 27 prognostic prediction models, and the external validation of 14 pre-existing models. The publications included differing predictors in their models including patient characteristics such as age, BMI and ECOG-PS, biochemical markers including tumour markers, CT scan findings, and laparoscopy.

The majority of models displayed a good ability to predict surgical outcome at the development and internal validation stage, however failed to replicate this predictive ability when externally validated on a separate patient cohort. The failure to externally validate models may due to differences between patient populations, differing surgical practice and surgical effort or using different methods to develop prediction models leading to overfitting to the described population. The majority of models included relied heavily upon CT findings, as CT images are the largest influence in clinical decision making in MDTs in current practice (Scott et al., 2020). However, the sensitivity and specificity of CT scans for identifying intra-abdominal metastasis range from 25 - 93% and 57 - 96%, respectively, casting some doubt

upon the current heavy reliance of this modality (Altman et al., 2012; Bailly et al., 2009; Coakley et al., 2002; Gemer et al., 2009).

Although a variety of modalities were included in the models, none addressed surgical heterogeneity. More radical surgery, including more extensive upper abdominal surgery, is associated with increased debulking rates, and in turn increased overall survival (Aletti, Dowdy, et al., 2006; Chi et al., 2009). It is well established that surgical heterogeneity exists, not only on an individual level within units, but also between units, and even countries (Aletti, Gostout, et al., 2006; J. M. Janco et al., 2015; Jones et al., 2018). The tendency of the surgeon towards performing radical procedures has been found to be associated with optimal cytoreduction (Aletti, Gostout, et al., 2006). Range of debulking rates between surgeons can be marked (42-67%) depending upon their surgical tendencies (Aletti, Dowdy, et al., 2006). It has been suggested that in excluding this factor, all prediction models will be unsuccessful, as debulking rates rely so heavily on surgical practice (Aletti, Dowdy, et al., 2006).

Although many models used a variety of modalities as predictors, all but three used univariable analysis to streamline the number of predictors before proceeding to the multivariable analysis and model development. This method is not recommended, as it can lead to incorrect predictor selection, as included predictors are selected based on their individual association with debulking, which excludes any significance that might occur when combined with other predictors. It is instead recommended that methods based on existing knowledge of previously established predictors in combination with statistical methods be used. It is not recommended that the number of predictors be reduced but instead to allow all clinically credible predictors to be retained in a model regardless of statistical significance (Harrell et al., 1996; Moons, Kengne, Woodward, et al., 2012; Sun et al., 1996).

The most successfully externally validated models relied upon laparoscopic triaging of patients (Feng et al., 2018; Petrillo et al., 2015). This method has received criticism in the past, as laparoscopy is not currently routinely performed as part of the diagnostic pathway. There were concerns that this procedure may introduce a time delay, increase morbidity due to complications and port site metastasis (PSMs) as well as have a monetary implication. On

systematic review of laparoscopy as a triage tool, it was found that the procedure does not increase healthcare costs, adversely affect complication rates or negatively impact on quality of life (van de Vrie et al., 2019). In addition, a large meta-analysis concluded that although rates of PSMs did increase following diagnostic laparoscopy, patient outcome was not negatively affected. The advantages to using laparoscopy as a triage tool could be twofold; firstly, it would allow for the completion of the prediction score, and secondly would allow for the pre-operative attainment of tissue via biopsy. By routinely achieving tissue pre-operatively, predictors requiring tissue biopsy such as protein panels and genomic markers could also be included.

The majority of models displayed high risk of bias, mainly due to the methods of model development and lack of successful validation. For this reason, none of the currently published models would be appropriate to be used in clinical practice in their current form. Some models do show promise when internally validated, and it would be beneficial for these models to be validated further. The use of all available predictors in a model, including laparoscopy, as well as incorporating the heterogeneity of surgical practice, may well improve model performance and should therefore be explored further.

5 External validation of a three-protein surgical prediction signature.

5.1 Introduction

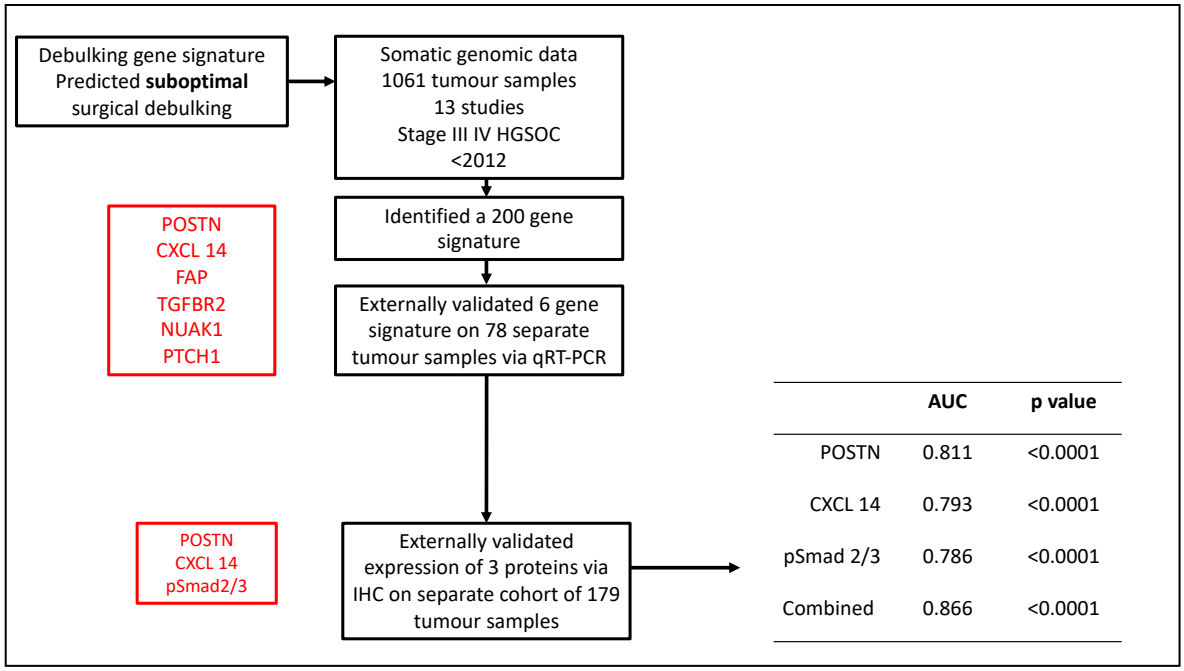
The systematic review in chapter 4 demonstrates the many published models that aim to predict poor surgical outcome at the time of primary surgery. As shown, none validate sufficiently enough to be of use in clinical practice. Often, external validation is not performed due to a lack of suitable databases large enough to perform a significant validation.

External validation to this point has mainly been confined to radiological and surgical models, with few of the newer biological models being validated. One such published model was described by Riester et al, in 2014. They developed a model to predict suboptimal debulking, firstly using genomic screening to create a gene panel, and then translating this to a three-protein signature (POSTN, CXCL 14 and pSmad2/3). The model utilised the presence of the three proteins in stage III and IV high grade epithelial ovarian cancer tumours to predict suboptimal surgical debulking status with a sensitivity of 92.8% and an AUC of 0.89 (Riester et al., 2014). Despite showing promise, this protein signature has yet to be externally validated on an independent cohort.

The model utilises IHC, a method that is commonly used in histopathology labs in the UK as part of the diagnostic pathway in ovarian cancer (Shah et al., 2012). IHC is readily available, fast and relatively cheap (Raab, 2000) when compared to alternative techniques such as genomic analysis. As our current treatment pathway includes a tissue biopsy for patients having NACT and IDS, if this model were to successfully validate, it could become routine for all patients to undergo a tissue biopsy pre-first line treatment.

The development of the protein signature is outlined in figure 5.1. The team collated data from eight different genomic databases and applied univariable logistic regression to identify 200 genes demonstrating association with suboptimal surgical debulking. The genes were analysed using Pathway Studio 7.1 software (Ariadne Genomics, Rockville, MD) to identify relationships between the genes, and resulted in the isolation of a panel of six genes: POSTN, CXCL 14, FAP, TGFBR2, NUAK1 and PTCH1.

To assess the six gene signature on a functional, rather than genomic level, three proteins were selected to act as surrogate biomarkers to represent the aforementioned six gene panel. Three proteins, POSTN, CXCL 14 and pSmad2/3 were selected, whose expression in tumour cells would represent the hyperactivation of the three pathways. This protein panel was internally validated on a separate cohort of 179 tumour samples with an AUC of 0.866.



	AUC	p value
POSTN	0.811	<0.0001
CXCL 14	0.793	<0.0001
pSmad 2/3	0.786	<0.0001
Combined	0.866	<0.0001

Figure 5.1 *Development and internal validation of the three-protein model (Riester et al., 2014). AUC- area under curve.*

The three proteins identified in the protein signature (POSTN, CXCL 14 and pSmad2/3) all have precedent in the literature for having a role in the encouragement of migration, as well as increasing vascularity of tumours in high grade epithelial ovarian cancer tumours. POSTN, (periostin, OSF-2) is a protein expressed in tumour stroma (Kujawa et al., 2020) , encoded by the POSTN gene in humans. POSTN encodes an extracellular matrix protein that functions in tissue development and regeneration, including wound healing. It functions as a ligand for alpha-V/beta-3 and alpha-V/beta-5 integrins to support adhesion and migration of epithelial cells (Gillan et al., 2002). In ovarian cancer, it is believed tumours associated macrophages contribute to tumour progression, and POSTN has been reported to be an important factor in macrophage recruitment in the tumour microenvironment through involvment in the interactions between macrophages and ovarian cancer cells (Tang et al.,

2018). POSTN can be overexpressed in cancer stroma in epithelial ovarian cancer (EOC) patients, with immunohistochemistry analysis showing that the overexpression of stromal POSTN was a powerful independent poor prognostic predictor for EOC patients associated with platinum resistance. POSTN also regulates ovarian cancer cell adhesion and motility (Choi et al., 2011; Gillan et al., 2002). High expression of POSTN in tumour stroma has been associated with a worse prognosis (Karlan et al., 2014).

Chemokine (C-X-C motif) ligand 14 (CXCL 14), is an antimicrobial gene belonging to the CXC chemokine family, and encodes for CXCL 14, a protein involved in immunoregulatory and inflammatory processes (Li et al., 2020). In ovarian cancer, CXCL 14 leads to multi-effects in tumorigenesis and development (Li et al., 2020), and is preferentially expressed in ovarian cancer (Bedognetti et al., 2013). CXCL 14 regulates several different biological processes in the body, including inflammatory immune responses, angiogenesis in cancer, host-specific tumour-specific immunity activation and autocrine tumour growth regulation (Lu et al., 2016). *In vitro* and *in vivo* experiments have both confirmed that the overexpression of CXCL 14 promotes ovarian cancer cell proliferation. The upregulation of CXCL 14 is associated with poor survival outcomes and promotes ovarian cancer cell proliferation. CXCL 14 expression is disproportionately increased in patients with metastatic disease (Li et al., 2020).

The SMAD gene codes for the protein pSmad 2/3, a member of a family of proteins acting as signal transducers and transcriptional modulators that mediate multiple signalling pathways. The protein also mediates the signal of the transforming growth factor (TGF) beta, and therefore regulates several cellular processes, such as cell proliferation, apoptosis, and differentiation. The disruption of TGF-beta has been strongly linked with ovarian cancer, leading to increased levels of pSmad 2/3 being expressed in cells (Alsina-Sanchís et al., 2017). TGF-beta signalling plays a role in ovarian cancer physiology as well as acting as a tumour promotor controlling proliferation (Alsina-Sanchís et al., 2017).

All three proteins promote the proliferation of ovarian cancer and have been associated with poor outcomes. Although these poor outcomes have been mainly attributed to chemotherapy resistance, there may be a surgical element to this. If a tumour is more disseminated, with extensive vasculature, it may well be more difficult to remove surgically.

The ICON5 trial; a trial conducted by the International Collaborative Ovarian Neoplasm (ICON) group, aimed to determine if the incorporation of an additional cytotoxic agent improved overall survival and progression free survival for women with advanced stage high grade epithelial ovarian carcinoma and primary peritoneal carcinoma who received carboplatin and paclitaxel (Bookman et al., 2009). The patients recruited for ICON5 were identified as an appropriate test bed for the external validation of the three-protein panel, as it was multicentred, included relevant data, and was a negative trial, so chemotherapy was not a confounder for the results.

5.2 Hypothesis and aim

5.2.1 Hypothesis

The three-protein signature originally developed by Riester et al. will successfully predict suboptimal cytoreduction in the ICON5 cohort of patients.

5.2.2 Aim

Using the ICON5 cohort of patients' tumour samples, validate the three-protein panel developed by Riester et al. via immunohistochemistry.

5.3 Results

5.3.1 Patient demographic and tumour samples comparison between original and validation cohorts

5.3.1.1 Original cohort

The original prediction model was developed in a cohort of 179 patients. Patients originated from a single centre and had all undergone PDS between 1993 and 2009. All patients were diagnosed with stage III and IV HGSOC. Suboptimal was defined as residual tumour ≥1cm and optimal as residual tumour <1cm.

5.3.1.2 Validation cohort

In total, 357 paraffin fixed high grade epithelial cancer patient samples from the ICON5 trial were identified. A minimum of three sections per tumour sample were required for the validation, and surgical outcome information was required. Of the 357 slides, 265 patient samples were identified as fitting the criteria for inclusion, figure 5.2.

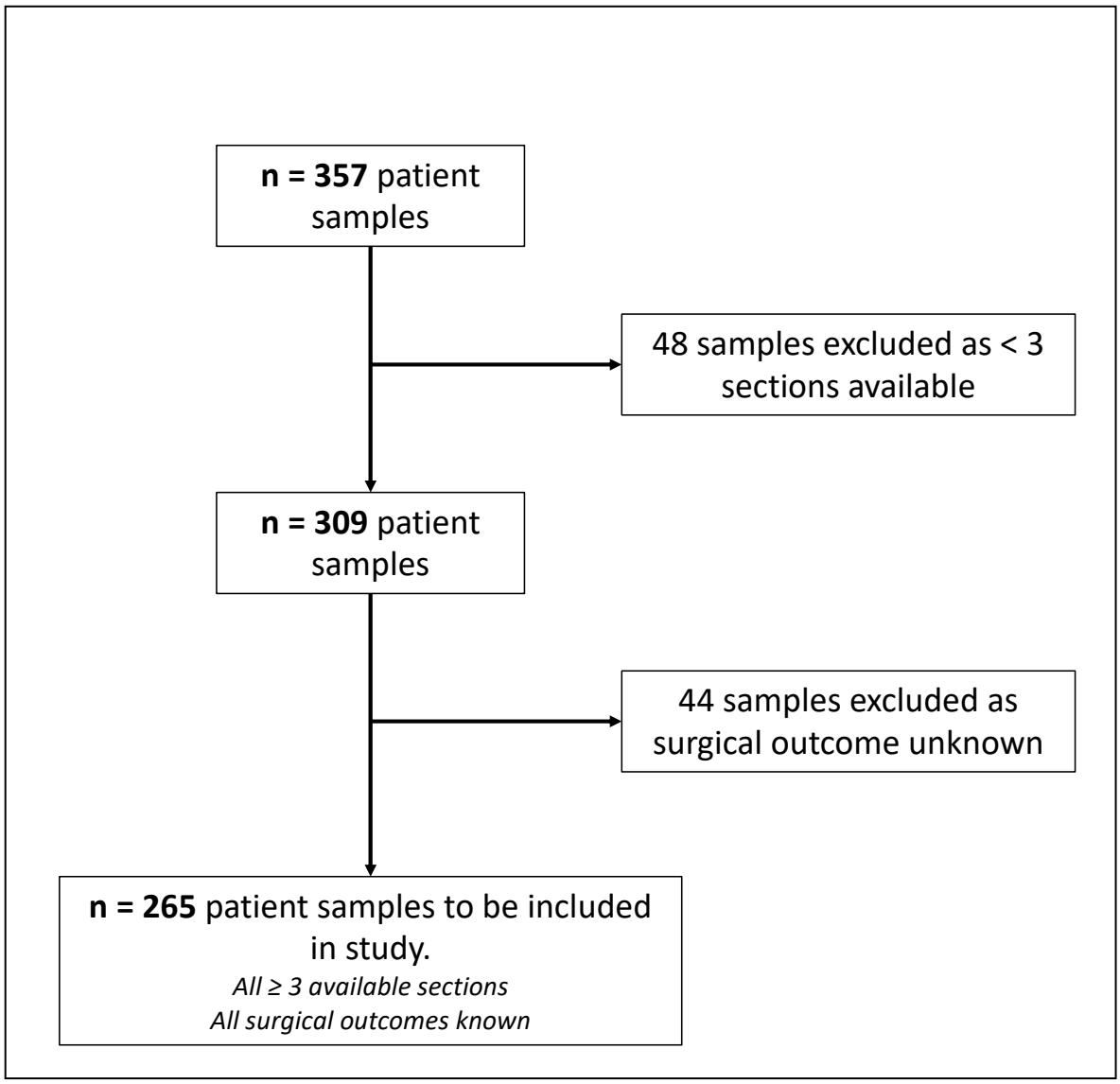


Figure 5.2 *CONSORT diagram showing final selection of samples for analysis*

Of the 265 patient samples included, all were histologically proven high-grade epithelial cancer and originating from stage III and IV tumours. In total, 85% of patients underwent primary debulking surgery, and the overall suboptimal rate was 46%, a higher percentage than reported for the whole trial cohort (30%). Suboptimal was defined as residual tumour ≥1cm and optimal as residual tumour <1cm.

Table 5.1 *Differences in patient and tumour characteristics between cohorts*

Demographic	Original dataset n= 178	Validation dataset n= 238	p value
Patient age	unknown	unknown	
Stage at diagnosis n (%)		unknown	
III	142 80%		
IV	36 20%		
Primary debulking surgery n (%)	100%	85%	
Suboptimal debulking rates n (%)	43 (24%)	112 (47%)	p <0.0001
Dates surgeries performed	1993 - 2009	2001 - 2004	
Age of samples at time of IHC	4 – 20 years	15 – 18 years	
Single or multicentre	single	multi	

As both models were created on cohorts developed for other purposes, the known demographic data was limited. Patient age, BMI, deprivation status, or the co-morbidities was not known about either cohort. The original geographic location of the first dataset was unknown, although this was a single centre study. Conversely, the validation data set included patient tumours from multiple different centres internationally.

Both datasets were made up entirely of high grade stage III and IV epithelial ovarian cancer tumours. The original dataset was split between stage III and stage IV respectively, however the breakdown of stages was unknown for the validation dataset. All tumours from the original dataset were taken at the time of primary surgery, however in the validation cohort 85% of tumours were taken at the time of primary surgery, and 15% at the time of interval debulking surgery.

The original dataset contained tumour samples that were between 4 – 20 years old at the time of IHC, and the validation dataset contained tumours between 15 – 18 years old at the time of IHC being performed.

5.3.2 Optimisation of antibody concentrations for validation

Optimisation of the antibody concentrations required for both for hand staining and then for automated staining on the BOND-III was performed. The concentrations used in the original study including for the negative control of Rabbit IgG are shown below in table 5.2. Details of brand and working concentrations used are described in methods section 3.3.1.

Table 5.2 *Antibody concentrations in original study*

Protein staining	Antibody concentrations
POSTN	1:800
CXCL 14	1:400
pSmad 2/3	1:200
Rabbit IgG	1:500

Firstly, optimisation by hand staining was performed by AH. A range of concentrations both more and less concentrated than the concentrations in the original study was performed. Three separate slides were stained for each concentration to ensure consistency of results, as shown in figure 5.3. Once stained, slides were reviewed by AH with GW to ensure adequate staining had been achieved. Adequate staining was achieved with the same antibody concentrations as the original study, as highlighted below in figure 5.3.

Further optimisation via automated staining on BOND III was performed. A range of concentrations both more and less concentrated than the original paper was used. Three sections were stained per antibody concentration to ensure consistency of staining. The concentrations used in the original paper produced adequate staining for the POSTN, CXCL 14 and negative control of Rabbit IgG. However, the concentration of 1:200 produced under-stained sections for the pSmad 2/3. For this reason, a further optimisation for the pSmad 2/3 in isolation was performed at the concentrations shown in figure 3. Review of confirmed adequate staining for pSmad 2/3 at the concentration 1:50.

Optimisation of antibody concentrations

1. Manual initial optimisation (x3 each)

POSTN	CXCL 14	pSmad 2/3	Rabbit IgG
1:400	1:200	1:100	1:250
1:800	1:400	1:200	1:500
1:1600	1:800	1:400	1:1000

Slides reviewed by consultant histopathologist specialising in gynae-oncology

2. BOND III initial optimisation (x3 each)

POSTN	CXCL 14	pSmad 2/3	Rabbit IgG
1:400	1:200	1:100	1:250
1:800	1:400	1:200	1:500
1:1600	1:800	1:400	1:1000

Slides reviewed by consultant histopathologist specialising in gynae-oncology

3. BOND III second optimisation (x 3 each)

pSmad 2/3
1:50
1:100

→ Slides reviewed by consultant histopathologist specialising in gynae-oncology →

4. Final antibody concentrations to be used

POSTN	CXCL 14	pSmad 2/3	Rabbit IgG
1:800	1:400	1:50	1:500

Figure 5.3 *Optimisation of antibody pathway- hand and automated staining*

Step one- demonstrates concentrations both used in original study and achieving adequate staining. Step two- bold text demonstrates adequate staining with automated staining. Red text highlights inadequate staining with both 1:200 concentration and more concentrated 1:100. Step three demonstrates further optimisation- green text representing adequate staining. Step four- final chosen antibody concentrations

Before scoring the full cohort, examples of percentage positive slides and intensity of staining were agreed, figure 5.4.

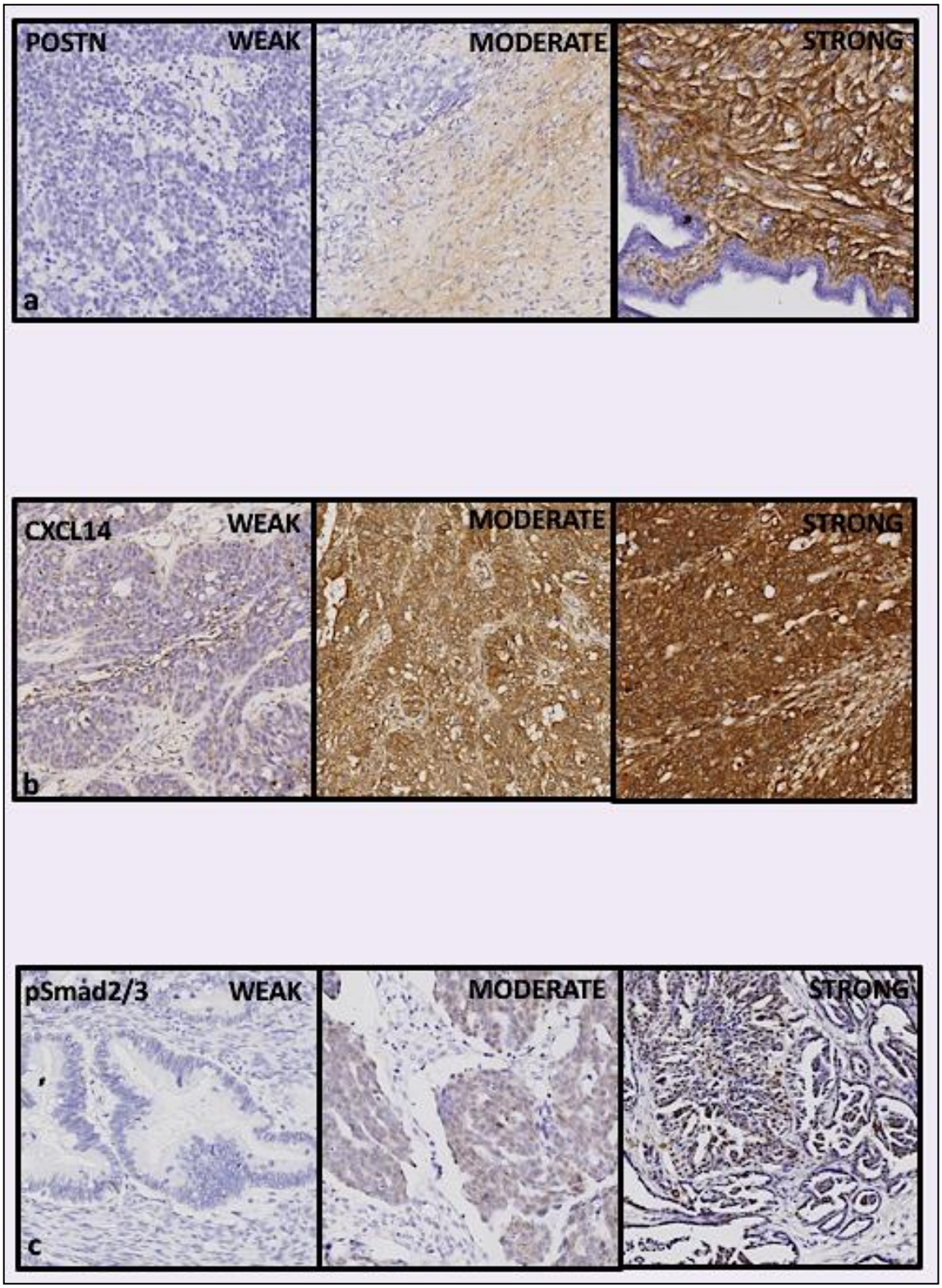


Figure 5.4 *Range of staining intensities*

Viewed using QuPath at 20 times magnification. Figure 5.4a demonstrating cytoplasmic staining of POSTN protein, figure 5.4b and 5.4c demonstrating nucleic staining of CXCL 14 and pSmad 2/3 respectively.

A sample of 50 slides were scored independently to ensure adequate agreement between scorers before the entire cohort was assessed. Correlation between the two scorers was very strong (Spearman r 0.8549 95% CI 0.7516 – 0.9173 p<0.001), figure 5.5.

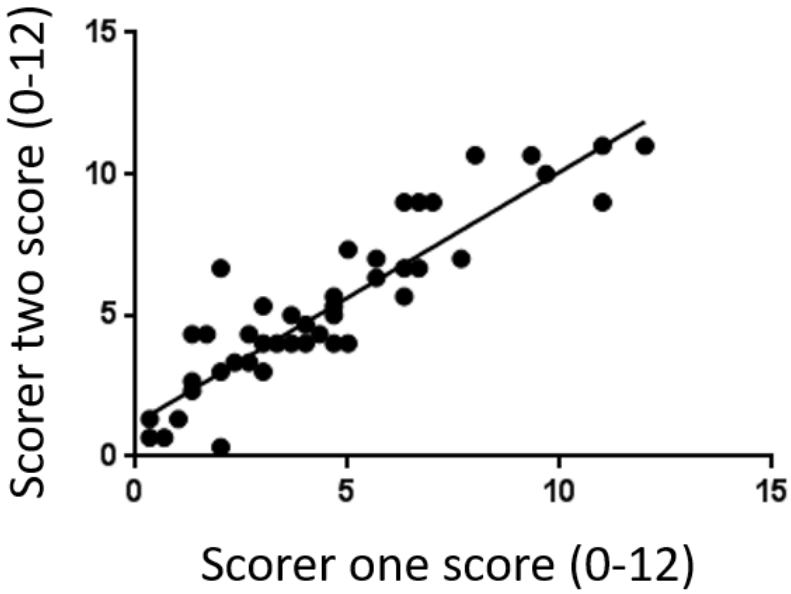


Figure 5.5 *Scatter graph demonstrating very strong correlation between two independent scorers*

p = <0.001

As very strong correlation between scores was demonstrated, the full completion of scoring of all 265 stained samples for each of the three proteins was next performed by both independent scorers.

5.3.3 External validation

5.3.3.1 Final sample size

Final scoring was completed for all 265 stained samples for each of the three samples, resulting in 795 slides scored per scorer. The tissue on a small number of slides had folded during the IHC process deeming them un-scorable, and a small number of slides had been scanned inadequately resulting in blurred images. The final number of tumour samples included in analysis was 238, figure 5.6.

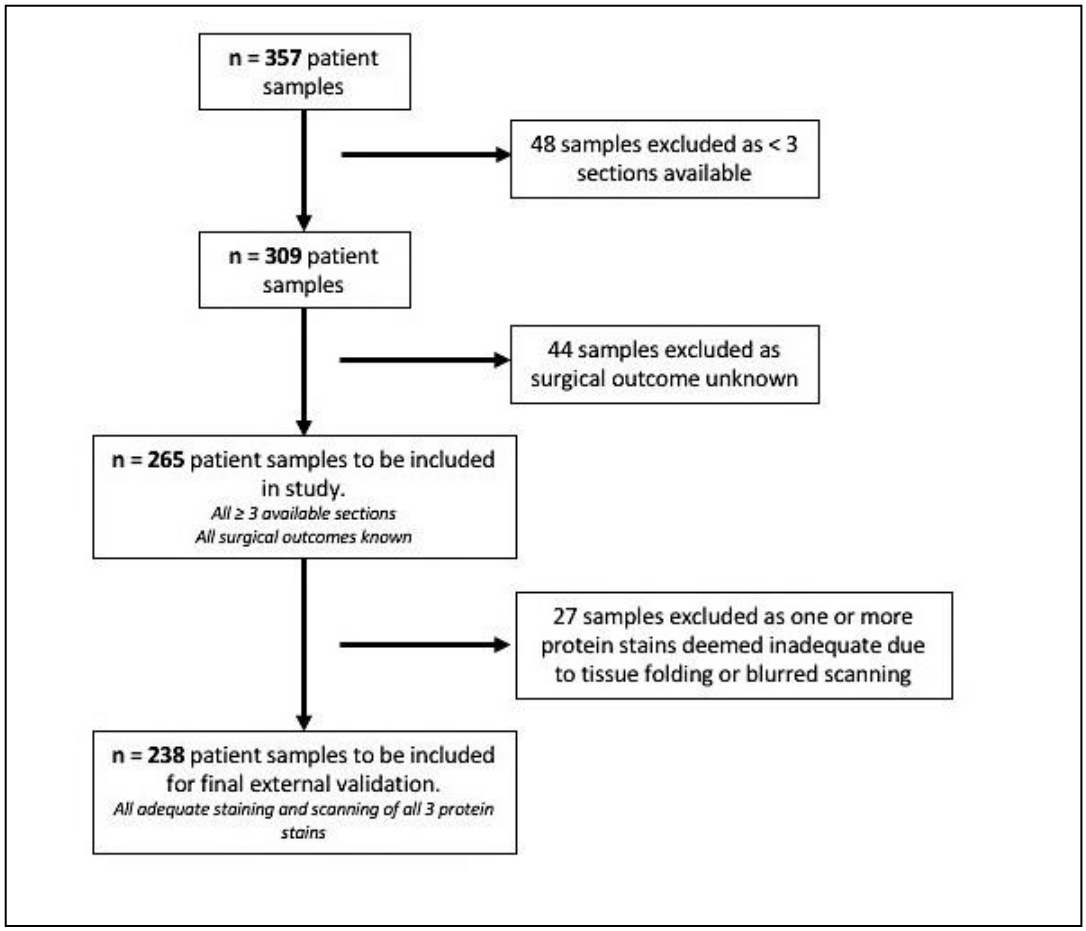
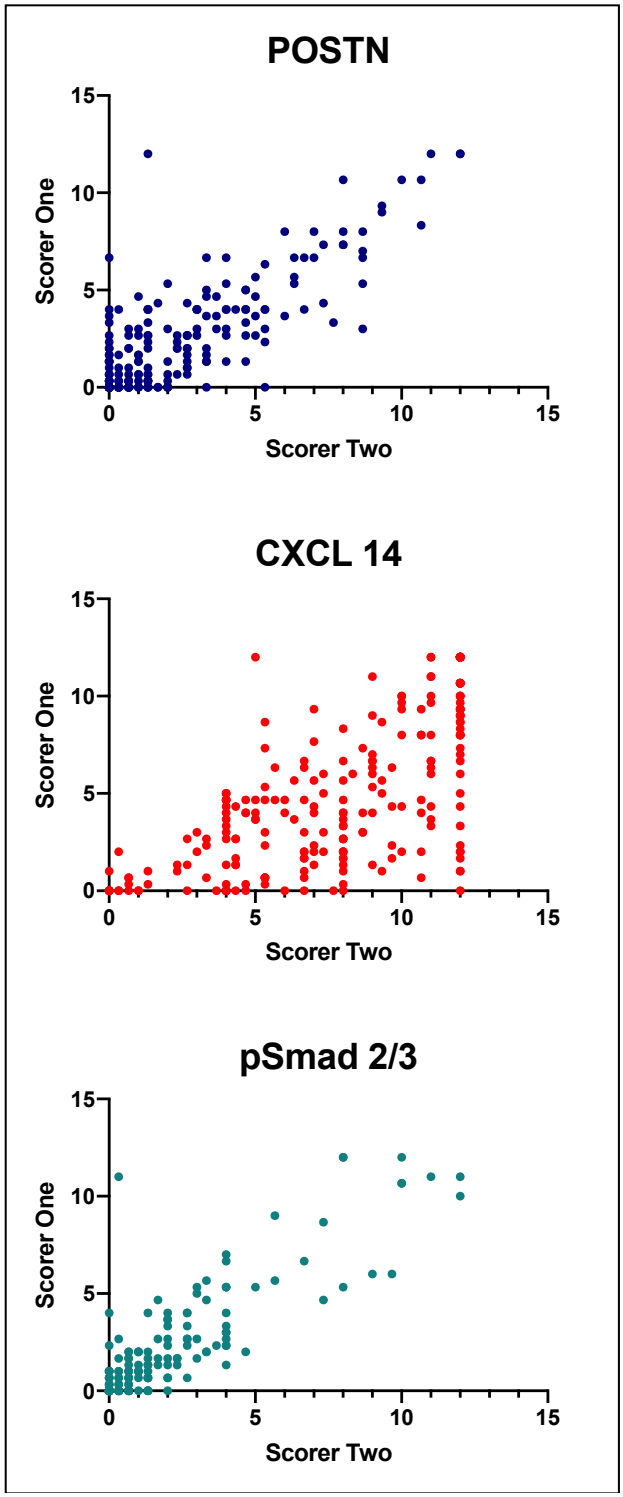


Figure 5.6 *CONSORT diagram demonstrating final number of samples in external validation*

Of the 238 tissue samples included for the final external validation, 46% had a surgery resulting in suboptimal debulking. This resulted in a number of events n=109 for the validation cohort.

5.3.3.2 Inter-scorer variability

Each scorer independently scored the 238 slides for each of the three proteins (795 slides in total). The slides were scored in three areas and a final score was calculated by averaging the scores for the three sections, described in section 3.3.2. To ensure agreement between scorers, inter scorer variability was assessed. Good correlation of scores was achieved for all three proteins, figure 5.7.



	R_2 value (95% CI)	p value
POSTN	0.749 (0.68 – 0.80)	<0.001
CXCL 14	0.700 (0.62 – 0.76)	<0.001
pSmad 2/3	0.836 (0.79 – 0.87)	<0.0001

Figure 5.7 *Scatter graphs showing inter-scorer variability*

Corresponding R^2 and p values seen in table

Correlation between both scorers in pSmad 2/3 and POSTN were very strong, however correlation between scorers for CXCL 14 was slightly weaker. Figure 5.7 demonstrates the

wider variation in scores for CXCL 14 when compared to POSTN and pSmad 2/3, which could explain the slightly lower correlation for this protein.

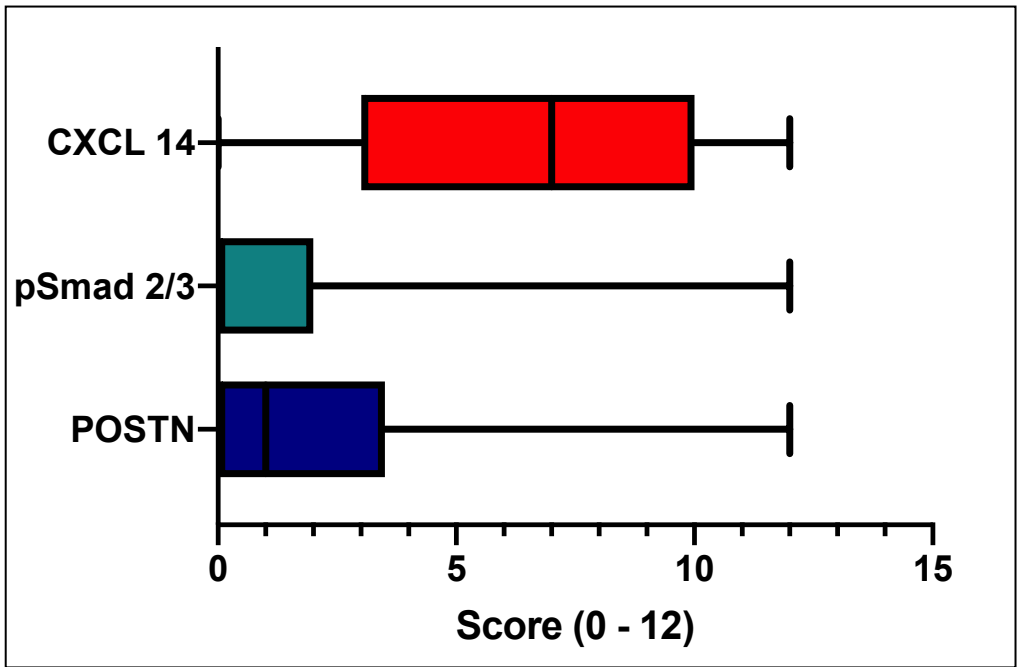


Figure 5.8 *Box and whisker chart showing breakdown of combined scores*

476 slides for each protein. The scores for CXCL 14 demonstrate a much wider variation than the other two proteins.

The three individual scores were combined to give an overall score between 0 – 36 for each tissue sample. Figure 5.9 demonstrates very similar distribution of total scores between the two independent scorers. Spearman's rank correlation coefficient demonstrated a strong positive correlation between the two independent scorers with an R_2 value of 0.8025 (95% CI 0.7506 – 0.8445) and $p < 0.0001$.

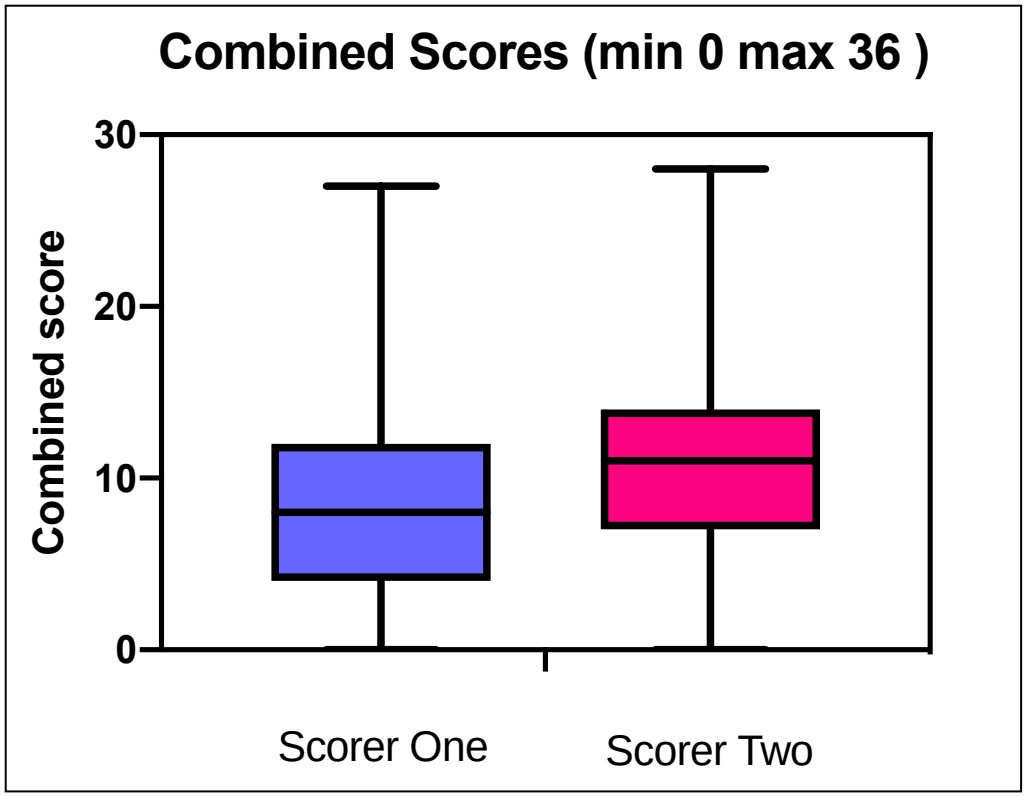


Figure 5.9 *Box and whisker chart showing ranges of scores between scorers*

Scorer two scores were used to externally validate the original three-protein signature. These scores demonstrate a wider range and therefore will represent our cohort more comprehensively than either scorer one or an average of the two scorers combined together.

All slide scores from the validation cohort along with clinical data including surgical outcome were consolidated to allow the external validation of the three-protein signature

The original three-protein validation model described the ability to predict suboptimal surgical outcome. This predictive ability was present for each three proteins individually, in addition when the scores for all three proteins were combined.

5.3.3.3 Comparison of methods

Methods between the two studies were replicated as closely as possible, however some variations did occur. The source of some antibodies varied, however for the anti-POSTN and the anti-pSmad 2/3, both the working antibody concentrations and the dilution of antibodies used in the IHC methods remained consistent across the two studies. As the working

concentrations of the CXCL 14 in the original study was unknown, this may have differed in the validation study, and may also be reflected in the difference in the optimisation of antibody concentrations in the validation study (1:400 original vs 1:50 validation).

The original study conducted the IHC entirely by hand, and slides were scored through visualisation through a light microscope. Conversely the validation study used a combination of hand staining and automated staining techniques for IHC, and all slides were scanned and scored manually on digital images.

The scoring system for each study was replicated exactly, as was the main statistical analysis.

5.3.3.4 Comparison of scores

Both models were created based on slides of high grade epithelial ovarian cancer tumour samples. All the slides were stained under very similar IHC conditions as above and scored using the same scoring system. Despite this, when the scores for each individual protein were compared between the two studies, they varied significantly in distributions between the two cohorts, figure 5.10. The scores for both POSTN and pSmad 2/3 showed significantly greater variation in standard deviation in the original cohort. Conversely, CXCL 14 showed a greater variation in standard deviation in the validation cohort.

Table 5.3 demonstrates the difference between the scoring of slides in the two cohorts. There are significant differences between both cohorts, when comparing them as a whole, and when broken down by surgical cytoreductive status.

Table 5.3 *Comparison of mean scores for each protein*

Cohorts		Original dataset Median score (IQ range)	Validation dataset Median score (IQ range)	p value
Complete cohort	CXCL 14	5 (4 - 7)	8 (5 - 12)	<0.0001*
	PSMAD 2/3	5 (3 - 7)	0 (0 - 2)	<0.0001*
	POSTN	4 (1 - 7)	1 (0 - 3)	<0.0001*
	Combined score	15 (9 - 19)	11 (7 - 14)	<0.0001*
Suboptimal cytoreduction	CXCL 14	7 (6 - 9)	7 (4 – 10.75)	0.548
	PSMAD 2/3	7 (6 - 8)	0 (0 - 2)	<0.0001*
	POSTN	8 (6 - 10)	1 (0 - 4)	<0.0001*
	Combined score	23 (17 - 26)	11 (7 - 13)	<0.0001*
Optimal cytoreduction	CXCL 14	5 (3 - 7)	9 (7 - 12)	<0.0001*
	PSMAD 2/3	5 (3 - 6)	0 (0 - 2)	<0.0001*
	POSTN	3 (1 - 5)	1 (0 - 3)	<0.0001*
	Combined score	12 (8 - 16)	12 (8 - 16)	0.557

p values calculated by Mann-Whitney U test, p <0.05 considered significant

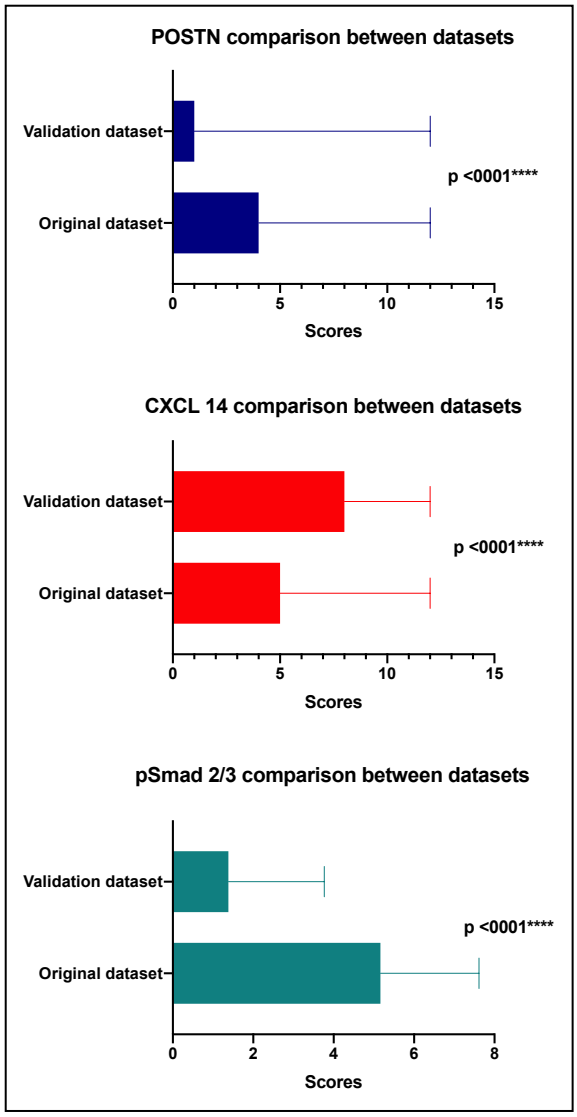
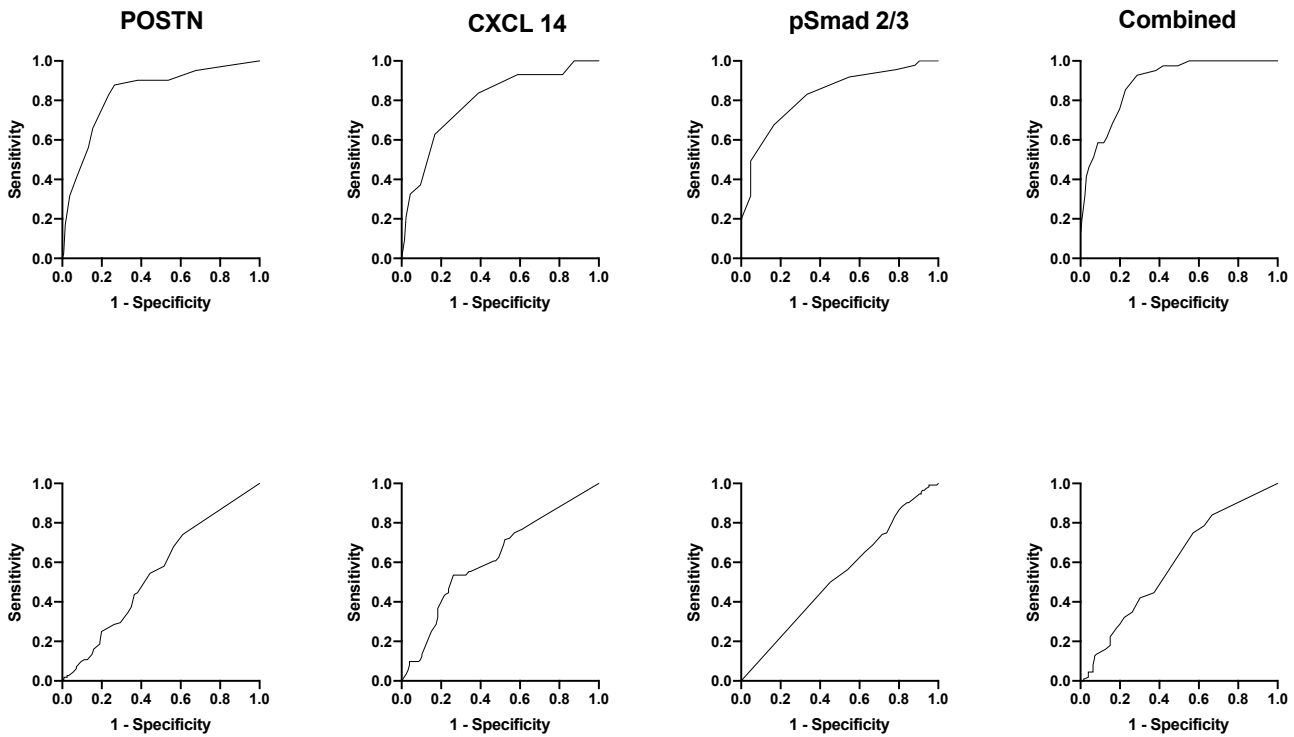


Figure 5.10 Box and whisker chart showing variation of scores between datasets

Plots showing mean and standard deviation of scores. Significantly different distributions between the two cohorts for all three proteins. P values calculated using Mann-Whitney U test.



POSTN	AUC	p value
Original	0.811	<0.0001****
Validation	0.551	0.174

CXCL 14	*AUC*	p value
Original	0.793	<0.0001****
Validation	0.621	0.0012**

pSmad 2/3	AUC	p value
Original	0.786	<0.0001****
Validation	0.529	0.4371

Combined	AUC	p value
Original	0.866	<0.0001****
Validation	0.593	0.0131**

Figure 5.11 *ROC curves demonstrating AUC of original and validation cohorts for three proteins individually and combined*

5.3.3.5 POSTN validation

In the original model, POSTN was predictive for suboptimal surgical outcome on a dataset of 178 high grade ovarian cancer tumour samples with an AUC of 0.811 and p value of <0.0001. When validated on our cohort of 238 high grade ovarian cancer tumour samples the predictive affinity was poor, with an AUC of only 0.551 and a p value of 0.174, figure 5.11.

5.3.3.6 pSmad 2/3 validation

In the original model, pSmad 2/3 was predictive for suboptimal surgical outcome on a dataset of 178 high grade ovarian cancer tumour samples with an AUC of 0.786 and p value of <0.0001. When validated on our cohort of 238 high grade ovarian cancer tumour samples the predictive affinity was again poor, with an AUC of only 0.529 and a p value of 0.3471, figure 5.11.

5.3.3.7 CXCL 14 validation

In the original model, CXCL 14 was predictive for suboptimal surgical outcome on a dataset of 178 high grade ovarian cancer tumour samples with an AUC of 0.793 and p value of <0.0001. When validated on our cohort of 238 high grade ovarian cancer tumour samples the predictive affinity was again poor, with an AUC of only 0.621 and a p value of 0.0012, figure 5.11.

5.3.3.8 Combined scores

In the original model, when the scores of all three proteins were added together to create a combined score, this score was predictive for suboptimal surgical outcome on a dataset of 178 high grade ovarian cancer tumour samples with an AUC of 0.866 and p value of <0.0001. When validated on our cohort of 238 high grade ovarian cancer tumour samples the predictive affinity was again poor, with an AUC of only 0.593 and a p value of 0.0131, figure 5.11.

5.3.3.9 Further exploratory analysis

As the four models above (three individual proteins and one combined score) all utilise univariable analysis, further analysis was performed using data from the original paper to explore whether the three protein scores could each be used as a separate predictor in a multivariable model. When a logistic regression multivariable mode was applied to the original dataset (n = 179), an AUC of 0.889 was achieved. However, when this was externally validated using the ICON5 dataset (n = 238), AUC dropped to 0.429.

The original prediction model was built on PDS patients only. The validation cohort was mainly PDS patients (85% n = 202), however, there were a small cohort of tumours taken at the time of interval debulking surgery (15%, n = 36). When the 36 patients who underwent IDS were excluded from analysis, the predictive affinity of each of four models remained largely unchanged (POSTN AUC 0.56, pSmad 2/3 AUC 0.54, CXCL 14 AUC 0.63, combined AUC 0.58).

5.4 Discussion

Despite showing promise and achieving successful internal validation, the successful predictive performance of the original model was not replicated when applied to an external cohort of patients.

There are no set parameters for prediction models in medicine required to be reached to declare the model successful and suitable for clinical practice. Instead, success can be measured by achieving an accuracy that is superior to existing methods designed to accomplish the same task (Myers et al., 2020). Although there are no currently successfully externally validated prediction models able to predict suboptimal surgical outcome published in the literature, the performance of this model when validated was poor. The accuracy levels achieved (AUC $\leq$ 0.621) for all models were not significant, therefore this model would not be acceptable for use in clinical practice based upon this validation.

The selection of the three proteins used in the original model was a complex process and involved somewhat opaque transitions between genomic analysis and functional level analysis. Although the presence of the three proteins in an ovarian cancer tumour do have

some scientific rationale as to the removability of the disease, the combination of the three proteins used has no precedence in the literature.

The original model was validated via IHC on a cohort of 179 patient samples, with a reported suboptimal debulking rate of 24%. Therefore, the number of events (number of patients undergoing suboptimal debulking) in the cohort was 43. The literature recommends at least 100 events for the validation of a clinical prediction model (Moons, Kengne, Grobbee, et al., 2012), and therefore the original model validation was underpowered. This could have resulted in overfitting of the model to the data and could account for the high reported accuracy in the original model. Conversely, the external validation was performed on a cohort of 238 patient samples with a much higher reported suboptimal debulking rate of 46%. Therefore, the number of events stands at 109, which allows greater confidence for the reliability of the failed external validation.

There are many previously published surgical prediction models that have also failed to successfully validate when applied to external cohorts, often attributed to the differences between the cohorts, and this study is no exception. Although all tumours in both cohorts were epithelial ovarian cancers of the same grade, and all late stages, there remain differences between the two groups. The most notable differences are the timing and the locations of the surgery performed. The original study contained tumours with surgery dates between 1993 – 2009, whereas the tumours in the validation cohort were removed in surgeries between 2001 – 2004. Although there is some time overlap between the groups, it is possible that surgical practice has changed over time. Furthermore, the original study cohort were operated upon within a single centre, whereas the validation cohort were made up of patients from multiple different centres internationally. Variation in surgical practice within centres is well established, and surgeon heterogeneity between centres is vast (Aletti, Gostout, et al., 2006; J. M. T. Janco et al., 2015; Jones et al., 2018). This variation in practice could contribute to the failed external validation of this model. If a tumour was deemed operable by the surgeons operating on the original cohort, this may well not be the case for a surgeon in one of the validation centres, deeming the model unreproducible anywhere other than the centre in which it was developed. The initial cohort contained 100% patients undergoing PDS, whereas the validation cohort included 15% of patients who had undergone

IDS. The patients undergoing IDS would no longer have chemotherapy naïve tumours, and this could affect results. There is an argument for excluding IDS patients from this analysis completely.

External validation was performed with care to ensure the techniques used to create the original model were replicated as closely as possible in the validation set. The conditions in which the IHC were undertaken were as similar as possible, although the original model utilised hand staining, and the validation used automated staining for the main part. Despite this difference, hand staining was also successfully performed for the optimisation of antibodies in the validation model and all other materials and methods were kept consistent, however, the concentrations of the antibody anti-pSmad 2/3 did differ between the two studies. However, the working concentrations of anti-pSmad 2/3 used in the original study were unknown. Therefore, if the original study was undertaken using a more concentrated batch, this could account for the discrepancy, as a higher concentration would be required for the validation study to achieve adequate staining.

There was very strong positive correlation between the two scorers in the validation cohort for both the POSTN and pSmad 2/3 staining, which gives confidence in the consistency of the scoring. Although there was still good positive correlation between scorers for the CXCL 14 staining, there was slightly more discrepancy than with the other two proteins. This was despite a learning period between the two scorers with a consultant histopathologist before scoring began. Due to the statistically proven strong correlation between the two scorers, the decision was made not to re-score the CXCL 14 slides. Re-scoring could introduce scoring bias, as the slides had been previously viewed, and the outcomes of both scorers known, and therefore a re-score would no longer be blinded. Differences were also identified in the scoring patterns between the two cohorts for each of the three proteins. Given that the two cohorts were made up of tumours removed in varying time and place, as well as varying storage ties, this is not surprising.

Both studies used historical slides that had been stored between 4 – 20 years before IHC was undertaken. The validation cohort were stored at room temperature in a pre-cut paraffin fixed state. The method of storing used in the original study was not known. There are very

few studies exploring the relationship between the time fixed slides are stored and the accuracy of IHC results. Some studies have suggested that longer storage time may be detrimental to antigenicity in tumour samples, resulting in false negative findings (Economou et al., 2014). Conversely, other studies have contradicted this theory, with Forse et al reporting adequate staining of breast cancer tissue via IHC following 12 years of storage (Forse et al., 2019), although these slides were stored at -80°C and not at room temperature. There is literature agreement that if slides are to be stored over prolonged periods, they must be paraffin fixed, as they were in the validation study. Both studies also included successful negative controls, suggesting that a positive result was indeed a true positive. Staining was also reviewed by an experienced consultant histopathologist, who confirmed that despite their age, the slides has stained adequately.

This external validation was conducted with adequate power, with the methods and materials followed as closely as possible. Despite this, the three-protein prediction model failed to accurately predict suboptimal surgical outcome in an external cohort. Without successful external validation, this model is not suitable for progression forward and consideration for use in clinical practice. Although future work could include the repeated validation on a separate cohort, the failed validation presented here would still stand, and therefore doubts for the efficacy of the model would remain, and further biomarkers are therefore needed.

6 Cancer genomics as predictors of surgical outcome

6.1 Introduction

The genomic landscape of ovarian cancer has been refined over recent years with data from the Cancer Genome Atlas (TCGA) and other consortia outlining classifications based on DNA damage repair status (Mukhopadhyay et al., 2012; TCGA, 2011) mutation profiling (TCGA, 2011) gene expression (Tothill et al., 2008), and copy number changes (Cerami et al., 2012; Macintyre et al., 2018).

To date, only the HR repair status of tumour has been identified to be a prognostic marker for progression free survival (PFS) and overall survival (OS) (Gee et al., 2018). A patient carrying a germline or somatic mutation in the HR pathway rendering the tumour HR deficient (HRD) demonstrates increased PFS and OS when compared to patients with tumours with functioning (competent) HR pathway (HRC). This survival advantage is attributed to an increased sensitivity to both platinum-based chemotherapy and PARP inhibitors (Macintyre et al., 2018; TCGA, 2011; M. Tumiati et al., 2018).

In 2011, the TCGA published a comprehensive overview of their findings of the genomics of ovarian cancer. They describe the full exome sequencing findings for 316 HGSOC patients (TCGA, 2011). HR status of both patient and tumour was recognised as playing an important role in both OS and response to medical treatment. HR defects were present in over half of all HGSOC cases. The most common gene defect involved in the HR pathway were BRCA1 and BRCA2 (20%). After analysis of over 72 genes thought to be involved in the HR pathway, 16 other genes were identified to be relevant. The study defined a unique HR gene panel, comprising of the genes identified to utilise HR for DNA repair in HGSOC; BRCA1, BRCA2, C11of30, PTEN, RAD51C, ATM, ATR, PALB2, FANCA, FANCC, FANCCI, FANCL, FANCD2, FANCE, FANCG, FANCM.

The correlation between HR status and surgical outcome has not been explored to date. If a patient's HR status was found to correlate with surgical outcome, this could become a valuable part of prediction tools to guide surgical decision making and increase the rate of

good surgical outcomes. It is estimated that approximately 90% of all HR mutations in ovarian cancer are germline, with the majority also being carried through into the tumour as somatic mutations, and approximately 10% being isolated somatic mutations (Capoluongo et al., 2017; Pennington et al., 2014).

Next generation sequencing (NGS) genomic testing and functional HR assays are becoming more commonplace in clinical practice, as well as becoming increasingly cheaper and faster to process. The testing for somatic germline testing via a small tumour sample at the time of diagnosis is becoming part of routine practice (Capoluongo et al., 2017; Konstantinopoulos et al., 2020).

6.2 Hypothesis

Patients who harbour a germline or somatic mutation rendering them HRD will be more likely to have favourable surgical outcomes than those who are HRC.

6.3 Aims

Determine correlation between tumour HR status and surgical outcome in MOCHR patient cohort utilising a functional HR assay.

Determine correlation between HR status and surgical outcome in the TCGA patient cohort utilising the HR gene panel identified by the TCGA.

6.4 Results

6.4.1 Patient cohorts

6.4.1.1 Manchester ovarian cancer homologous recombination cohort (MOCHR)

The MOCHR database currently comprises 101 HGSOC patient samples. Surgical debulking status is known for 46 of these samples. Of the 46 patients, eight were defined as being homologous recombination heterogenous (HRH) and were excluded from the analysis as further outlined in methods section 3.4. This resulted in a final n=38 MOCHR patients, all HGSOC.

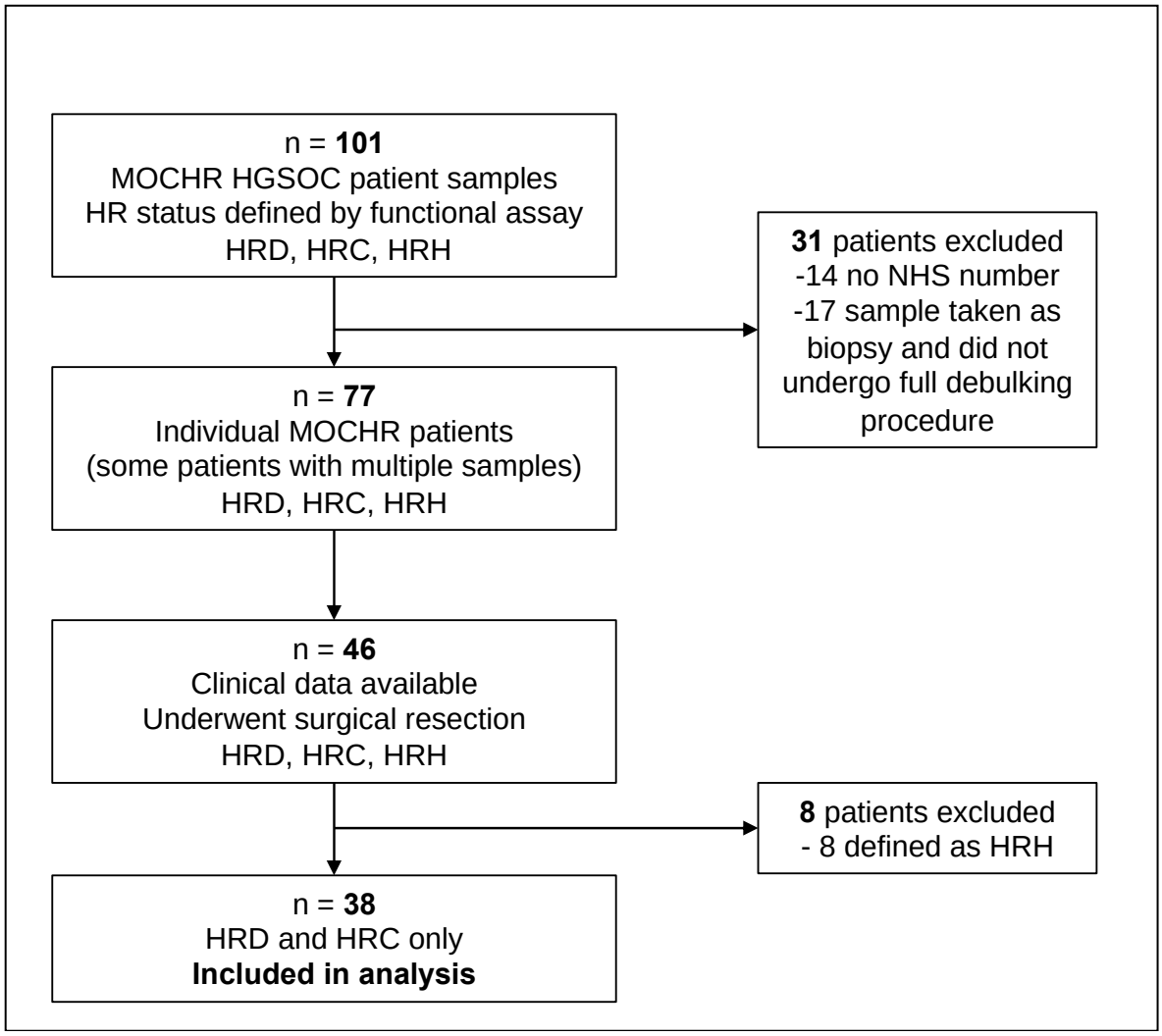


Figure 6.1 *CONSORT diagram illustrating patient selection for inclusion*

All included patients were HGSOC, had undergone surgical resection with debulking status available, and were defined as HRC or HRD by way of the HR functional assay.

Patient tumour samples included in this cohort all originated from a single tertiary UK cancer centre, collected between 2013 – 2018. The formal histology reports for all patients were reviewed to ensure disease was of HGSOC subtype, and tissue samples were characterised to ensure they were representative of the patient's disease. HR status was determined via a functional HR assay, and described as HRD or HRC, see section 3.4.

A full breakdown of demographics for both cohorts can be found in table 6.1. Patient mean age was 72 years with a range between 42 – 91 years. Samples originated from a variety of anatomical sites including omentum (41%), ascites (25%), ovarian tumour (24%) and peritoneal disease (10%). The majority of tumours were from stage III disease (74%), with the remaining classified as stage IV disease (26%). Half of the patients (n=19) underwent PDS, with the other half undergoing IDS following neo-adjuvant chemotherapy. At the time of surgery, 22/38 (58%) patients achieved complete cytoreduction (no macroscopic disease remaining), 9/38 (24%) achieved optimal cytoreduction (<1cm visible disease remaining), and 7/38 (18%) achieved suboptimal cytoreduction (≥1cm disease remaining). Median survival in the MOCHR cohort was 42 months (SEM 1.1), with survival data correct as of July 2020. Of the 38 tumours, 21 (55%) were categorised as homologous recombination repair deficient (HRD) with the remainder homologous recombination repair competent (HRC).

6.4.1.1.1 Homologous recombination functional assay

All 38 included tumours underwent HR characterisation by way of the functional assay. All functional assays were performed by MP, section 3.4. Sufficient damage to promote DNA repair was confirmed if ≥100% increase in H2AX foci was observed between control and UV treated cells. Samples were categorised as HRC if cells displayed ≥100% increase in Rad51 foci over the control. Samples were categorised as HRD if cells displayed <100% increase in rad51 foci, figure 6.2

A breakdown of all 38 patient samples showing average number of H2AX and Rad foci as an average over 100 cells before and after UV treatment with percentage increase can be seen in table 6.1. All samples demonstrated >100% increase in H2AX foci indicating sufficient DNA damage. Of the competent tumours, percentage increase in Rad51 foci ranged between 614.8%

in MOCHR sample number 157 and 103.6% in MOCHR sample number 49. In the deficient tumours Rad51, percentage increase ranged between -91.0% in MOCHR sample number 126 and 83.4% in MOCHR sample number 61.

Table 6.1 *Number of H2AX and Rad51 foci in MOCHR sample cells pre and post UV exposure*

HR status	MOCHR number	H2AX foci number Control	H2AX foci number Post UV exposure	Percentage increase in H2AX foci number	Rad51 Foci number control	Rad51 Foci number post UV exposure	Percentage increase in H2AX foci number
HRC	19	2.2	22.3	903.9%	2.1	4.2	104.3%
	32	0.3	11.6	3360.7%	0.2	1.5	586.9%
	49	1.1	2.4	123.5%	0.3	0.7	103.6%
	69	0.5	2.3	330.4%	0.6	1.9	231.0%
	71	1.2	10.2	739.3%	1.6	4.0	146.0%
	74	0.8	7.7	862.5%	2.4	5.3	124.0%
	78	1.1	6.7	537.3%	0.2	1.3	474.1%
	120	0.4	1.1	164.5%	0.6	1.9	223.0%
	127	1.7	7.5	328.2%	1.6	4.6	188.0%
	138	0.5	6.0	1211.9%	2.9	8.3	182.9%
	140	0.9	7.3	679.0%	2.6	5.4	109.0%
	142	0.8	8.6	1043.7%	2.5	6.5	161.5%
	153	3.5	10.1	184.9%	5.1	11.4	122.4%
	156	0.6	3.6	480.9%	0.5	2.0	308.0%
	157	0.8	10.7	1276.3%	0.2	1.6	614.8%
	160	1.3	5.6	330.8%	1.7	6.3	271.5%
	172	1.5	17.4	1082.8%	10.0	23.2	132.7%
HRD	16	2.5	6.7	163.7%	1.8	1.4	-23.6%
	34	0.4	0.9	118.7%	1.6	1.0	-41.4%
	61	0.6	3.7	479.3%	8.7	15.9	83.4%
	62	3.1	11.9	283.6%	6.6	9.1	38.1%
	67	1.2	7.0	499.9%	5.1	6.8	35.2%
	75	0.7	3.4	368.6%	3.4	3.5	2.0%
	77	1.9	4.0	108.7%	3.6	1.5	-57.4%
	107	1.7	11.5	555.3%	2.2	3.3	48.9%
	108	6.1	15.1	147.7%	5.7	7.1	25.1%
	124	1.1	5.6	417.1%	2.5	2.1	-17.1%
	125	1.5	11.1	623.7%	2.2	1.9	-13.7%
	126	3.3	13.8	316.0%	1.1	0.1	-91.0%
	136	0.9	7.1	723.9%	0.6	0.3	-45.4%
	137	3.3	19.5	488.4%	24.3	19.2	-20.8%
	146	0.5	1.7	257.8%	2.5	1.4	-43.9%
	155	0.1	0.2	296.0%	9.6	11.0	14.9%
	161	1.1	5.9	437.8%	1.4	1.1	-21.2%
	169	1.2	5.9	404.3%	7.4	8.1	9.4%
	173	0.7	9.7	1260.5%	6.0	9.3	54.7%
	178	3.4	8.1	141.8%	6.6	4.6	-30.8%
	179	2.3	8.2	253.5%	8.6	9.2	7.7%

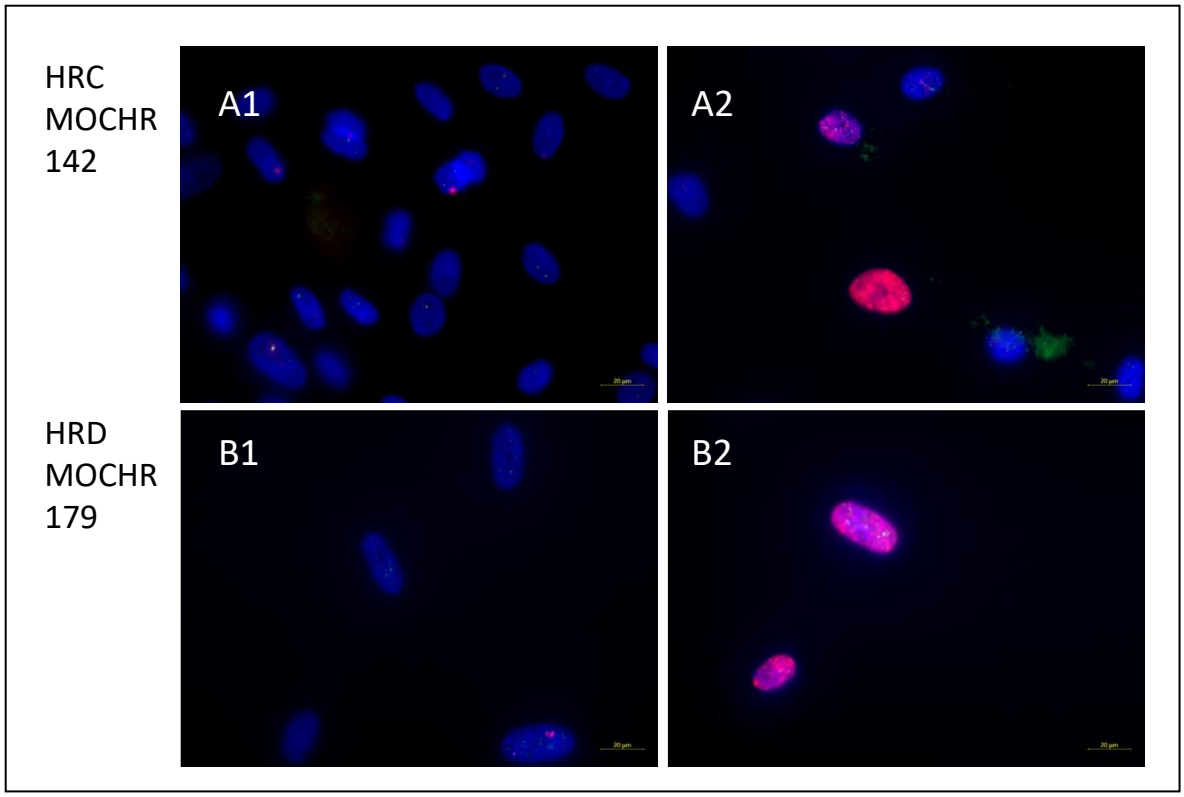


Figure 6.2 *- Zeiss Axio Observer microscope images of MOCHR samples 142 and 179*

Figure A showing untreated (A1) and treated (A2) MOCHR 142 cells. The increase of H2AX foci (red) and Rad51 foci (green) demonstrates a competent HR pathway.
Figure B showing untreated (B1) and treated (B2) MOCHR 179 cells. The smaller increase (<100%) of H2AX foci (red) and Rad51 foci (green) demonstrates a deficient HR pathway.

6.4.1.2 The Cancer Genome Atlas cohort (TCGA)

Following the mining of the cBioportal online database and the corresponding data in the TCGA database, as per section 3, 258 patients were identified for inclusion in the study. All patient tumours were HGSOC, and surgical outcome and survival data was available for all included patients.

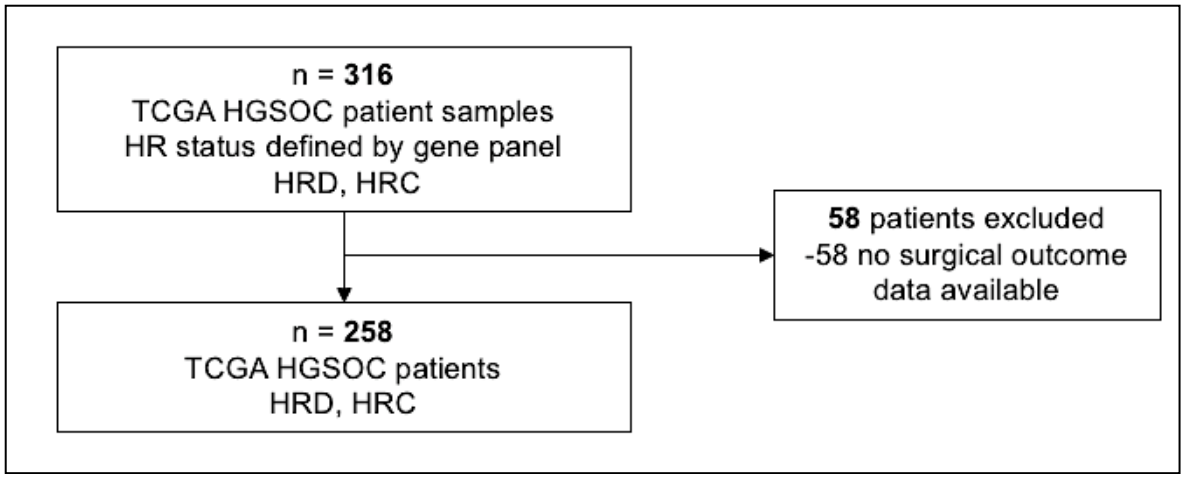


Figure 6.3 *CONSORT diagram demonstrating patient selection for inclusion in the TCGA cohort*

All included patients were HGSOC, had undergone surgical resection with debulking status available, and were defined as HRC or HRD by way of the gene panel.

Patient tumour samples included in this cohort were collected as part of the cancer genome atlas project between 2005 – 2009, from multiple different centres. A stringent quality control process was applied in order to ensure all tumours were HGSOC in nature, as per section 3. HR status was determined, and a tumour classified as HRD if any mutation was detected (germline or somatic) in any of the 14 genes included in the HR gene panel, see section 3. If no mutation was detected in the selected genes, a tumour was classified as HRC.

Patient mean age was 60 years with a range between 28 – 87 years. All specimens were primary HGSOC tumours, however the exact anatomical site of origin for all samples was unknown. The majority of tumours were from stage III disease (79%), with the remaining stage IV disease (21%). All samples were collected at the time of primary debulking surgery, before the patients received any chemotherapy. At the time of surgery 56/258 (22%) patients achieved complete cytoreduction, 137/258 (53%) achieved optimal cytoreduction and 65/258 (25%) achieved suboptimal cytoreduction. Median survival in the TCGA cohort was 41.4 months (SEM 1.0), with survival data correct as of August 2010.

Of the 258 tumours, 114 (44%) were categorised as HRD with the remainder categorised as HRC.

The pattern of gene mutations involved in the HR pathway for each of the 114 patient samples, as well as a breakdown of their surgical outcome can be seen in figure 6.4. Of the HRD

tumours, the genetic alterations included missense mutations, truncating mutations, germline mutations, amplifications and deep deletions. The most commonly mutated genes were BRCA 1 (12%) and BRCA 2 (12%).

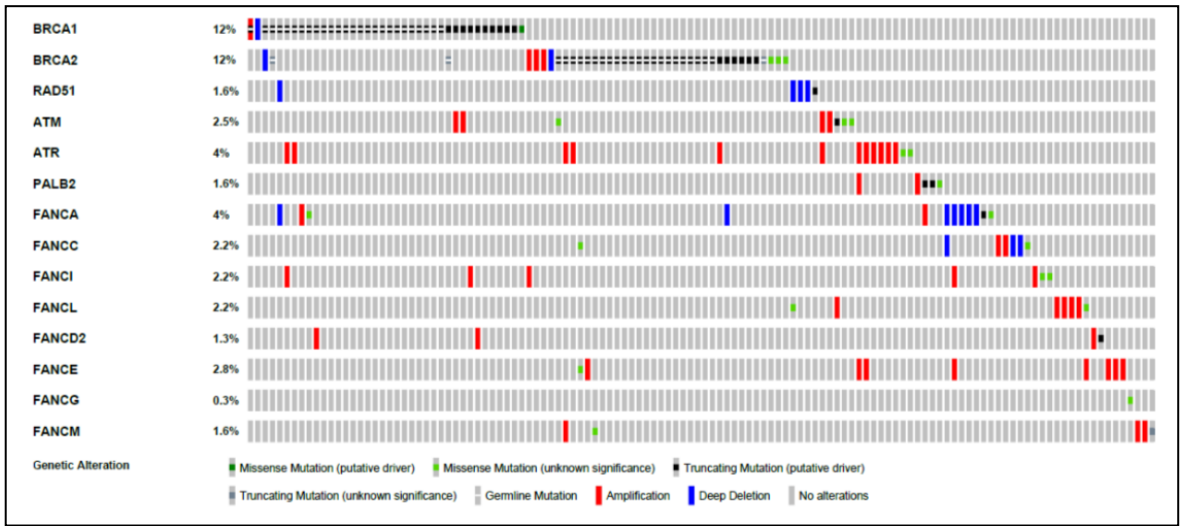


Figure 6.4 *Genetic alterations in the 114 HRD tumours in the TCGA cohort*

Figure created using software available on cBioportal (Cerami et al., 2012)

6.4.1.3 Cohort comparisons

Both databases were analysed separately to determine any association between tumour HR status and surgical resection rates. However, when directly compared, some notable similarities and differences can be seen between the two cohorts, table 6.2.

Table 6.2 *Demographics comparison of MOCHR and TCGA cohorts*

	MOCHR (%)	TCGA (%)	p value (RR 95% CI)
n	38	258	-
Age mean (SD)	72.6 (12.1)	60.2 (11.4)	*<0.0001
Range	45 - 91	28 - 87	
Histology HGSOC (%)	38 (100)	258 (100)	-
Sample source (%) Ovary	9 (24)	-	-
Ascites	10 (25)	-	-
Omentum	15 (41)	-	-
Peritoneum	4 (10)	-	-
FIGO (%) Stage III	28 (74)	204 (79)	**0.526 (0.772 0.408-1.511)
Stage IV	10 (26)	54 (21)	
Surgery (%) PDS	19 (50)	258 (100)	**<0.0001 (0.068 0.044-0.140)
IDS	19 (50)	0 (0%)	
Surgical outcome (%) complete	22 (58)	56 (22)	***<0.0001
optimal	9 (24)	137 (53)	
suboptimal	7 (18)	65 (25)	
HR status (%) HRD	21 (55)	114 (44)	**0.2245 (1.473 0.817-2.657)
HRC	17 (45)	144 (56)	
Median survival months (SE)	42 (1.1)	41.4 (1.0)	****0.793 (0.932 0.552-1.574)

*n- number, * Mann-Whitney test, **fishers exact test, ***Chi-sq test SD standard deviation, ****log rank test. SE standard error, RR risk ratio, CI confidence intervals*

Both cohorts contained tumours of FIGO stage III and IV disease only, with a HGSOC histological subtype. They have similar distributions of stages with the majority of tumours being from stage III disease, as well as having a similar split between HRD and HRC tumours.

The median survivals between the two cohorts are similar 42 vs 41.4 months for HRD vs HRC, respectively, with no statistical difference between survival curves found (p=0.793 log rank test HR 0.9362 95% CI 0.5596 – 1.1787), as seen in figure 6.5.

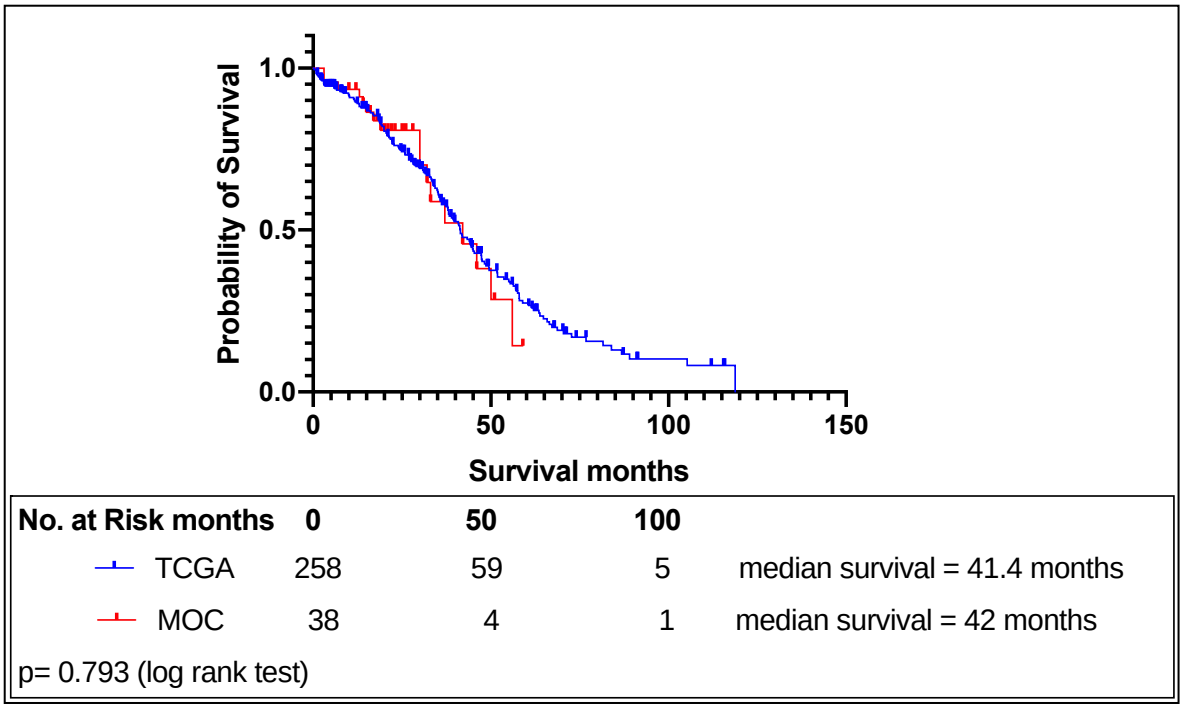


Figure 6.5 *Kaplan-Meier survival curve comparing MOCHR and TCGA cohorts*

The TGCA cohort were on average older (p<0.0001), and represented all primary surgery tumours, making all tumours chemotherapy naïve. Conversely, the MOCHR cohort were evenly split between PDS and IDS and therefore only half of the cohort was chemo naïve, with the other half having been exposed to three cycles of platinum-based chemotherapy before resection.

A greater number of patients in the MOCHR compared to the TCGA cohort underwent a complete cytoreductive procedure, with more of the TCGA cohort achieving an optimal cytoreduction. Both groups defined complete cytoreduction as no macroscopic disease remaining, optimal as <1cm disease and suboptimal as disease remaining ≥1cm. Differences between surgical resection rates in the two cohorts can be seen in figure 6.6.

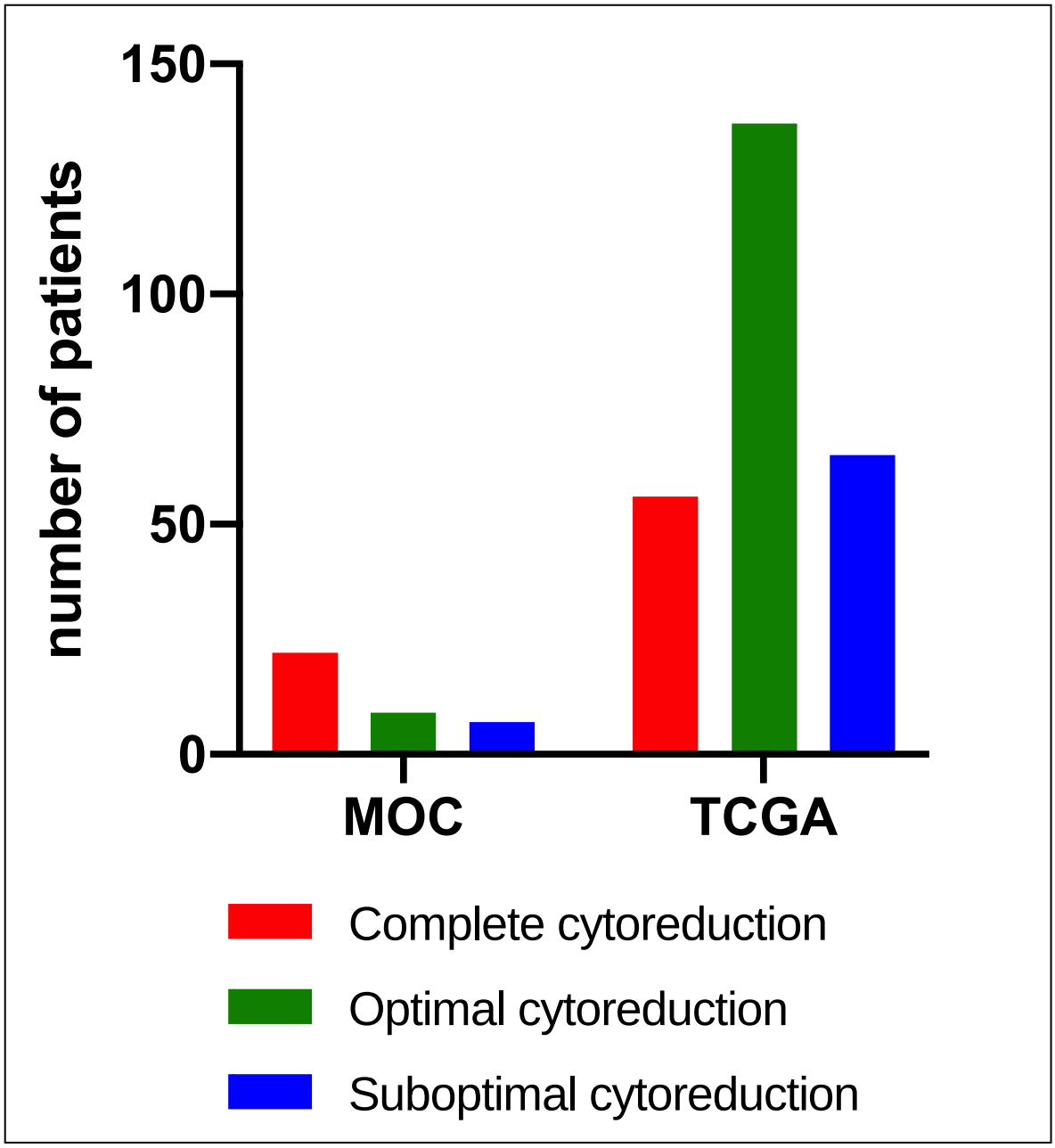


Figure 6.6 *Bar chart showing differences between resection rates*

Chi sq. analysis p <0.0001.

HRD status has previously translated to increased patient survival when compared to HRC tumours, a difference that is reflected in the TCGA cohort (p=0.0047 log rank test HR 1.577 95% CI 1.148 - 2.165), where the HRD tumours had a median survival of 51.7 vs 37.9 months when compared to HRC tumours. The tumours in the MOCHR cohort however did not display this same survival advantage for the HRD tumours, figure 6.7.

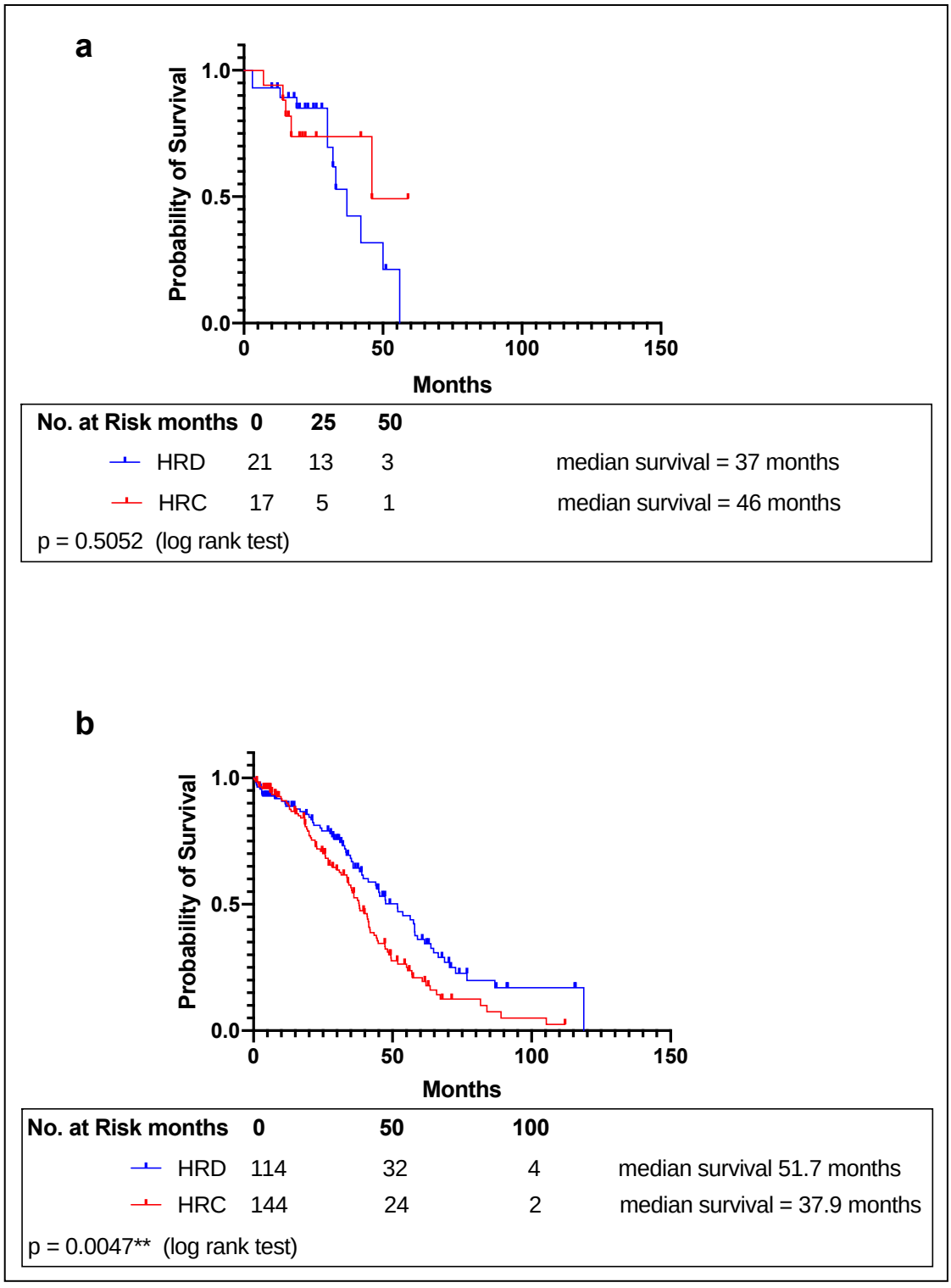


Figure 6.7 *Kaplan-Meier survival curves showing difference in survival by HR status*

Figure 6.7a. *Illustrates slightly worse non-significant survival in the HRC group compared to the HRD group in the MOCHR cohort.*
Figure 6.7b. *Illustrates significantly increased survival in the HRD group when compared to the HRC group in the TCGA cohort.*

6.4.2 Determining a binary outcome

In both cohorts, surgical outcome was recorded as complete, optimal or suboptimal cytoreduction in line with international guidance.

Overall survival is well documented to be inversely correlated with residual disease at the end of surgical debulking. In the MOCHR cohort complete, optimal and suboptimal outcomes showed an inversely proportional correlation with survival (median survival 50, 33 and 15 months, respectively, p=0.0014), see figure 6.8a. The same correlation was seen also in the TCGA dataset with median survival for complete, optimal and suboptimal increasing as residual tumour volume decreased (58, 41, 32 months, p=0.0002), see figure 6.9a.

Figure **6.8** and figure 6.9 demonstrate that in both cohorts, complete cytoreduction significantly increased survival over optimal and suboptimal debulking. However, there was no significant survival difference between optimal and suboptimal debulking.

These findings suggest that in both cohorts, patients gained a survival advantage if no visible disease remained (complete cytoreduction). However, if any visible disease remained, even <1cm (optimal and suboptimal cytoreduction), there was limited further survival advantage. Therefore, the three surgical outcomes, complete, optimal and suboptimal were combined to create a binary outcome of good or bad, as shown in table 6.3.

Table 6.3 *Binary cytoreduction outcomes*

Cytoreduction outcome	Binary outcome
Complete (no macroscopic disease)	Good outcome
Optimal (macroscopic disease <1cm)	Bad outcome
Suboptimal (macroscopic disease ≥ 1cm)	Bad outcome

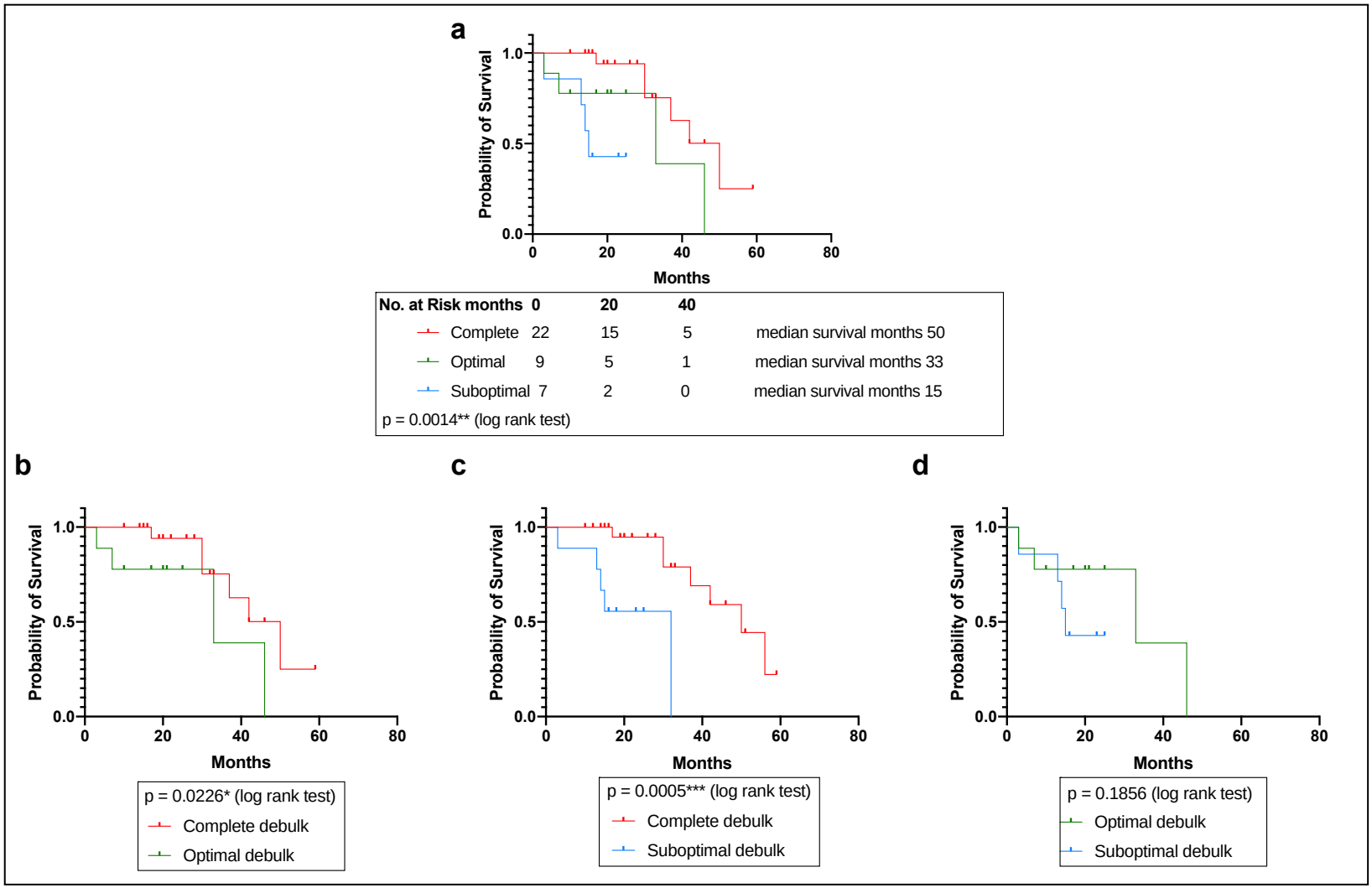


Figure 6.8 *Kaplan-Meier curves demonstrating survival between surgical outcomes in MOCHR cohort*

n=38 (6.7a), complete vs optimal (6.7b), complete vs suboptimal (6.7c), optimal vs suboptimal (6.7d).

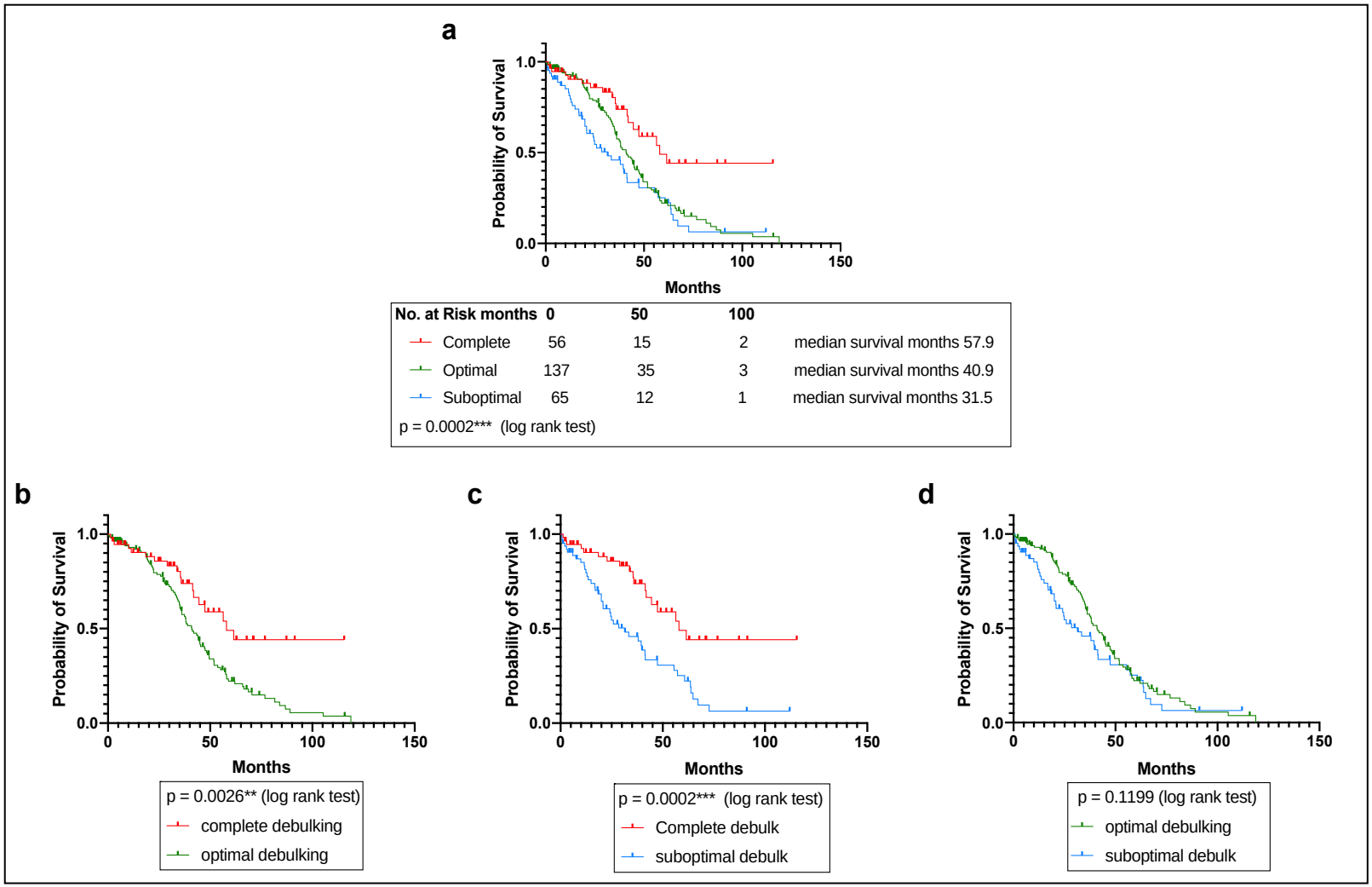


Figure 6.9 *Kaplan-Meier survival curves showing survival difference between surgical outcomes in the TCGA cohort*

n=258 (6.8a), complete vs optimal (6.8b), complete vs suboptimal (6.8c), optimal vs suboptimal (6.8).

6.4.3 Correlation between HR status and surgical outcome

6.4.3.1 MOCHR cohort

Of the 38 patients included in the MOCHR cohort, 17 were classified as HRC and 21 as HRD by the HR functional assay. The hypothesis suggests that the 21 patients in the HRD group would be more likely to achieve a favourable (complete) surgical outcome, and that the 17 patients in the HRC group would be less likely to achieve a favourable (therefore optimal/suboptimal) surgical outcome.

In the MOCHR cohort, the most common surgical outcome regardless of HR status was good surgical outcome (22 vs 16 patients). Of these 22 patients, 50% were HRD and 50% HRC, as demonstrated in figure 6.10b.

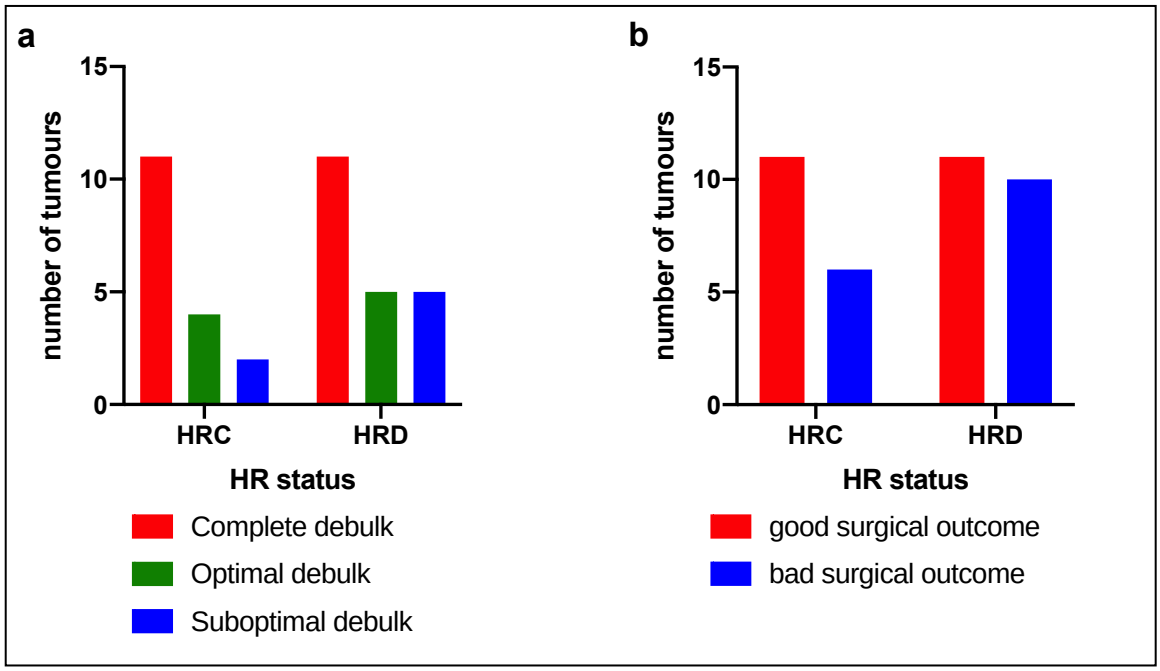


Figure 6.10 *Bar charts showing distribution of binary surgical outcome by HR status*

when surgical outcome is defined as complete, optimal, suboptimal in 9a, and as a binary outcome of good and bad outcome in 9b.

The contingency table (table 6.4) highlights the even split between HR status in the good surgical outcome group. There were a higher number of patients in the bad surgical outcome group who were HRD (10/16, 63%) compared to HRC.

Table 6.4 *Contingency table of HR status vs surgical outcome in MOCHR cohort*

n= 38	Good surgical outcome	Bad surgical outcome
HRC	11	6
HRD	11	10

When the correlation between the data was assessed via Fisher's exact test, table 6.5, there was no statistical correlation between HR status and surgical outcome (p=0.5205). PPV and NPV were low (0.6471 and 0.4762, respectively), suggesting HR has no predictive value for surgical outcome in this cohort.

Table 6.5 *Correlation results between HR status and surgical outcome MOCHR cohort*

p value (OR 95% CI)	PPV (95% CI)	NPV (95% CI)
0.5205 (1.667 0.4516 – 5.433)	0.6471 (0.4130 – 0.8269)	0.4762 (0.2834 – 0.6763)

Key- OR odds ratio, CI- confidence intervals, PPV- positive predictive value, NPV- negative predictive value

6.4.3.2 TCGA cohort

Of the 258 patients included in the TCGA cohort, 144 were classified as HRC and 114 as HRD via the HR gene panel. The hypothesis suggests that the 114 patients in the HRD group be more likely to achieve a favourable (complete) surgical outcome, and that the 144 patients in the HRC group would be less likely to achieve a favourable (therefore optimal/suboptimal) surgical outcome.

In the TCGA cohort, the most common surgical outcome regardless of HR status was bad surgical outcome (202 vs 56 patients). Of these 202 patients, 59% were HRD and 41% HRC, as demonstrated in figure 6.11b.

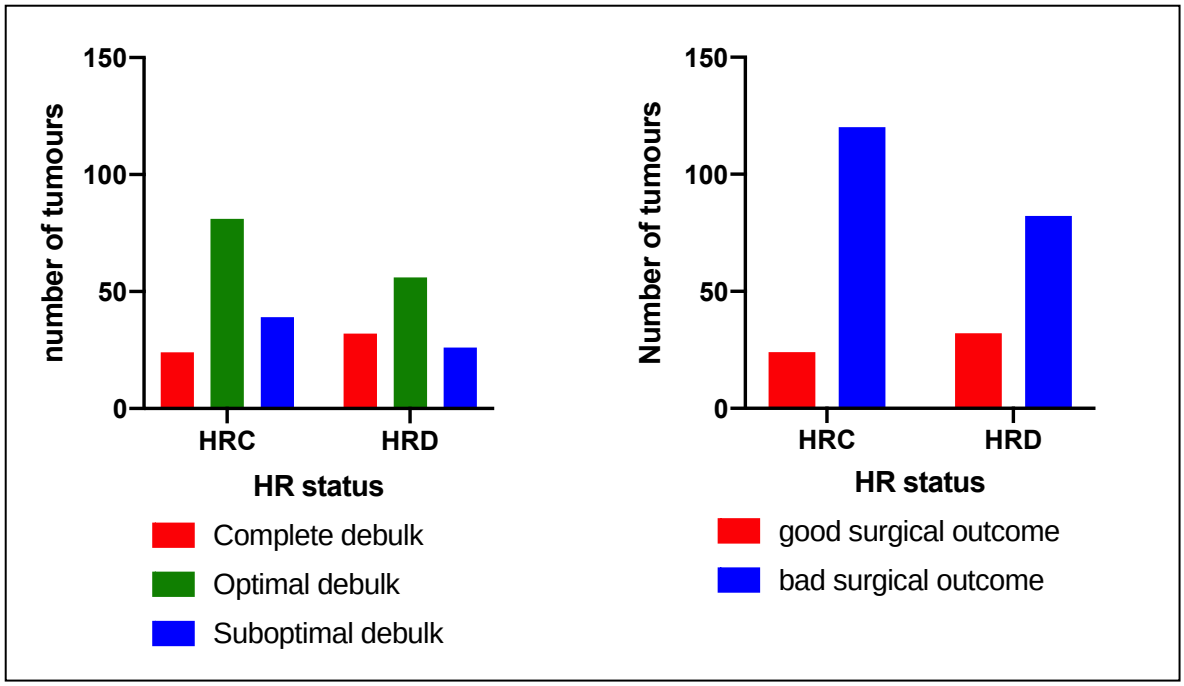


Figure 6.11 *Bar chart showing distribution of surgical outcome by HR status*

Surgical outcome is defined as complete, optimal, suboptimal in 10a, and as a binary outcome of good and bad outcome in 10b.

The contingency table (table 6.6) highlights there were a higher number of HRD patients who achieved a good surgical outcome compared to HRC patients. Conversely, more patients with HRC achieved a bad surgical outcome compared to HRD patients.

Table 6.6 *Contingency table of HR status vs surgical outcome for TCGA cohort*

n= 258	Good surgical outcome	Bad surgical outcome
HRC	24	120
HRD	32	82

Fisher's exact test, table 6.7, demonstrated there was a statistical correlation between HR status and surgical outcome (p=0.0332). PPV was low (0.1667), suggesting HRD has limited predictive value for determining good surgical outcome. However, the NPV was relatively high (0.7193), which suggests that an HRC status could be more predictive of a bad surgical outcome.

Table 6.7 *Correlation between HR status and surgical outcome in TCGA cohort*

p value (OR 95% CI)	PPV (95% CI)	NPV (95% CI)
0.0332 (1.91 1.071 – 3.492)	0.1667 (0.1146 – 0.2360)	0.7193 (0.6307 – 0.7936)

OR odds ratio, CI- confidence intervals, PPV- positive predictive value, NPV- negative predictive value

6.5 Discussion

The ability to accurately predict surgical debulking outcome for HGSOC based upon a patient's germline or somatic HR status established at the time of diagnosis would be of significant clinical benefit. This predictive ability could be utilised as a stand-alone variable, or as part of a larger, multivariable model. It would facilitate triaging of patients to the best first line treatment on an individual basis. The HR status would be best established via somatic testing, to ensure inclusion of all mutations. A tissue biopsy could be performed either radiologically or during a diagnostic laparoscopy. With the recent introduction of PARP inhibitors to the routine first-line treatment algorithm in HGSOC (Banerjee et al., 2020), a patient's HR status is now extremely clinically relevant at the time of diagnosis, and therefore a part of initial diagnostic investigations.

Analysis of the TCGA cohort indicated a statistically significant correlation between HR status and surgical outcome. This analysis defined HRD as a defect in one of the 14 genes included in the TCGA HR gene panel using next generation sequencing (NGS) analysis. NGS holds the advantage of screening for defects within a wide variety of genes, however when a specific gene panel is applied, the interpretation is limited to the genes of the panel (Frey & Pothuri, 2017). NGS does not evaluate HRD due to other aetiologies such as epigenetic modifications (Frey & Pothuri, 2017). The 14 genes included in the gene panel used in this study is comparable to panels described in the literature (Matondo et al., 2017; van Wijk et al., 2020; Vanderstichele et al., 2017). Each panel identifies a slightly different array of genes utilising HR as their method of repair. This analysis should be repeated on all available HR gene panels to ensure consistency of results with respect to HR status. Ideally this analysis should be performed on a separate patient cohort.

The patients and tumour samples included in the TCGA analysis originated from between 2005 – 2009. The large size of the study (n=258) gives confidence in the statistical significance

of the results. Patient selection included multiple different geographical sites and therefore allowed for selection from a wide variety of patient cohorts. The historic nature of the samples could present difficulties replicating the results on more recent patient cohorts, however when primary surgery is considered, management has been relatively consistent across this time period with the exception of the advocating for more radical surgical management (Lheureux et al., 2019).

Analysis of the MOCHR cohort did not show statistically significant correlation between HR status and surgical cytoreduction. Only a small number of patient samples were available for analysis, and therefore the negative outcome of this study may be attributed to a lack of power, rather than a true clinical finding. The MOCHR cohort underwent functional analysis of their tumour. This method offers the advantage of determining HR status regardless of the underlying genetic or epigenetic mechanism, and can therefore be preferable over NGS (Frey & Pothuri, 2017; Manuela Tumiati et al., 2018). This cohort also contained patients who had undergone IDS as well as PDS. The inclusion of the IDS cohort could skew results as these tumours would no longer be chemotherapy naïve, which could alter their ability to be surgically removed. Future studies should exclude IDS and focus purely on PDS patients.

The TCGA cohort were comprised solely of patients who underwent primary surgery and therefore were all chemotherapy naive. The MOCHR cohort were split evenly between PDS and NACT/IDS, exposing 50% of the cohort to three cycles of chemotherapy prior to testing. Exposure to chemotherapy can alter the HR status of tumour cells in HGSOC (Damia & Broggini, 2019) which could account for the lack of correlation in this cohort. To exclude the IDS/NACT patients from the analysis would require reduction of the already small cohort to a number where useful statistical analysis would not be possible. Further analysis with a larger cohort would be required.

Analysis from the TCGA cohort is promising, and further validation of these findings on an external large cohort is essential. The 100K genome project hold germline and somatic genomic data for 316 HGSOC patients collected from multiple centres across England, and analysis is underway. Further work is currently in progress to collect relevant clinical data

fields to complement the available genomic data. Once complete this dataset would be ideally placed for the external validation of the TCGA panel.

7 A prognostic model to predict suboptimal surgical outcome

7.1 Introduction

The importance of reducing rates of suboptimal cytoreduction in advanced HGSOC by improving treatment pathway selection is paramount. A suboptimal surgical outcome translates to reduced overall survival, increased morbidity and a delay in commencing chemotherapy treatment (Bristow et al., 2002; Chi et al., 2009; Fagotti et al., 2006; M. J. Rutten et al., 2015). In cases where suboptimal debulking is thought to be likely at the time of MDT, NACT and consideration of IDS is appropriate (Kehoe et al., 2015; van Meurs et al., 2013; Vergote et al., 2010; Wright et al., 2016). The accurate and reproducible prediction of surgical outcome allows for better treatment pathway selection. However, as described in chapter four many published pre-existing models have failed when validation is attempted. With the exception of laparoscopic models, those utilising single modalities have to this point been unsuccessful, as demonstrated in chapters five and six.

There are currently no universally accepted indications for NACT in clinical practice. Decisions are made based mainly upon radiological imaging and clinician opinion, with additional available information not part of the decision process (Scott et al., 2020). The rates of suboptimal debulking vary greatly but have been reported between 9% - 76% (Horowitz et al., 2018; J. M. Janco et al., 2015), highlighting the urgent clinical need for a tool to guide decision making in this field.

There are many patient and tumour factors that have been associated with surgical outcome, including: age, BMI, ECOG-PS, Hb, platelet count, albumin levels, WCC, CA 125, HE4, tumour protein and gene panels, tumour stage and histology, as well as disease distribution at the time of CT scan and laparoscopy (Abdallah, Chon, et al., 2015; Chesnais et al., 2017; de Jong, Eijkemans, Fong, et al., 2007; Enshaei et al., 2015; Fagotti et al., 2006; C. G. Gerestein et al., 2011; Horowitz et al., 2018; Riester et al., 2014; Wang et al., 2015).

Multiple prognostic prediction models have attempted to combine the above features in order to accurately identify which patients will have both good and bad outcomes at the time of primary surgery. However, to date, none have been validated for use in clinical practice. When assessed via the PROBAST tool, the majority of models show high risk of bias, see figure 4.3. Although many researchers consider a wide variety of predictors to feature in their models, most significantly reduce the number included by using univariable analysis as a triaging tool before performing multivariable analysis and creating the final model. Although it is important to have a tool that is clinically easy to use, performing univariable analysis may limit model performance (Moons, Kengne, Woodward, et al., 2012).

Prediction models in medicine can be developed using a variety of both traditional statistics and machine learning. For prognosis and prediction, regression models are the most commonly used and include linear, logistic and Cox regression. A systematic review and meta-analysis comparing the use of machine learning with traditional statistics in prediction models concluded that one technique was not superior above the other (Christodoulou et al., 2019).

The description of residual disease following PDS has varied slightly over time. Current UK practice is to define surgical success as one of three outcomes; complete: no macroscopic disease remaining; optimal: disease remaining <1cm; suboptimal: disease remaining ≥1cm (Winter et al., 2008). It is now widely accepted that an outcome of complete debulking, where no macroscopic disease remains confers the highest survival benefit (Chi et al., 2006; Eisenkop et al., 2003; Hamilton et al., 2011). For this reason, the majority of prognostic models define good outcome as no macroscopic disease (complete debulking), and bad outcome as macroscopic disease of any size (optimal and suboptimal debulking).

Survival rates and good surgical outcomes have been shown to significantly improve with the introduction of radical procedures such as upper abdominal surgery and peritoneal stripping. (Aletti, Dowdy, et al., 2006; Aletti et al., 2009; Chi et al., 2009; Dowdy et al., 2008; Eisenhauer et al., 2006; Jones et al., 2018). However, rates of maximum surgical efforts vary between units, and even between individual surgeons, which is reflected in the range of debulking rates reported (Jones et al., 2018). Despite this clear association between surgical heterogeneity and debulking rates, very few researchers have considered ‘performing

surgeon' as a predictor in their models (Jung et al., 2013). It is unlikely that a model would be transferrable between patient populations, unless this factor is addressed.

A significant proportion of patients with advanced HGSOC undergo surgery which results in residual disease at the time of PDS. Although multiple models exist to predict outcome, attempts to translate the research into clinical practice have failed at the validation stage. For this reason, there is a valid clinical need remaining in this area.

7.2 Hypothesis

The combining of multiple predictors associated with surgical outcome into a single prognostic model will allow the accurate prediction of surgical outcome at the time of primary debulking surgery in patients with advanced HGSOC.

7.3 Aims

To develop an internally validated multi-predictor prognostic model to predict residual disease at the time of primary surgery in stage III and IV HGSOC patients using logistic regression.

To develop an internally validated multi-predictor prognostic model to predict residual disease at the time of primary surgery in stage III and IV HGSOC patients using machine learning techniques.

7.4 Results

7.4.1 Cohort characteristics

Between 2013 and 2018, 100 patients within the Manchester Database met the inclusion criteria and therefore were included in the model development (figure 7.1).

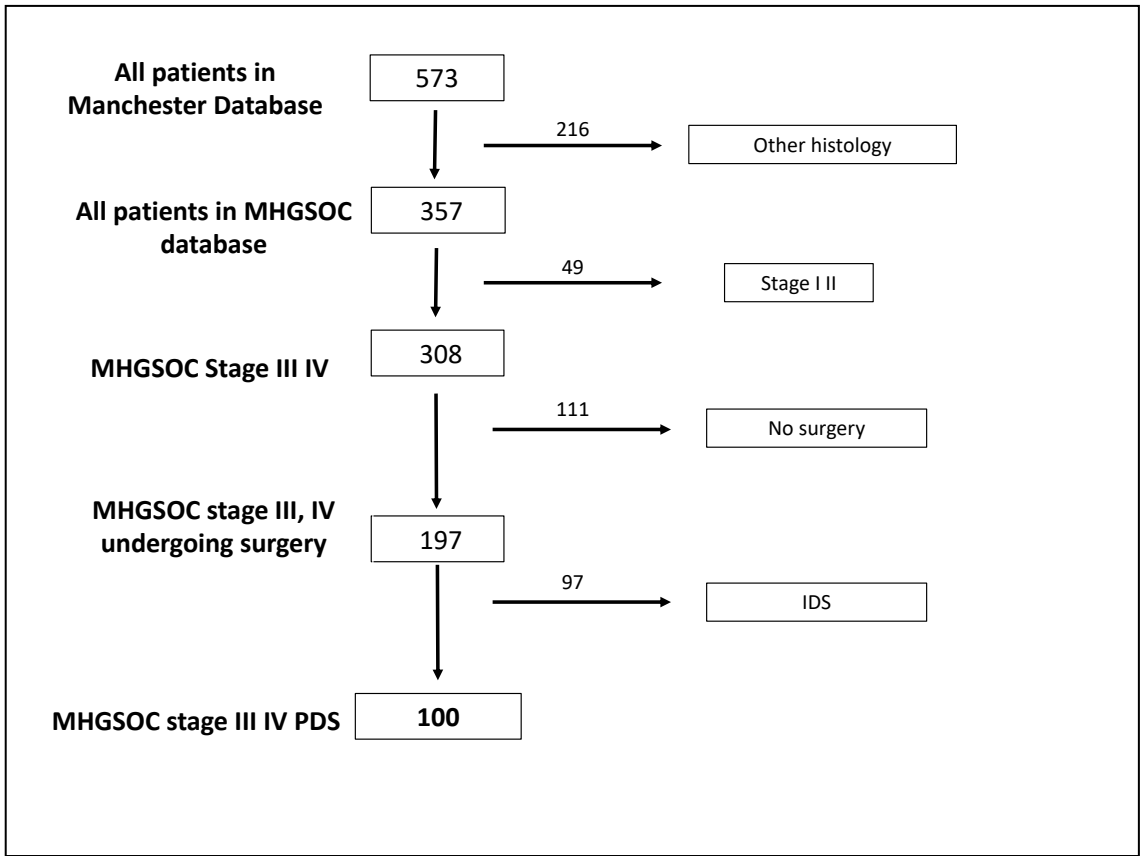


Figure 7.1 *Flow diagram demonstrating identification of patients included in model*

MHGSOC- Manchester high grade serous ovarian cancer database IDS- interval debulking surgery, PDS- primary debulking surgery.

7.4.1.1 Patient characteristics

Good surgical outcome (complete cytoreduction) was achieved in 59% (59/100) of the patients with stage III and IV HGSOC. The clinical characteristics of the patients and disease are shown in tables 7.1 and 7.2.

Table 7.1 *Comparative patient and disease characteristics for whole cohort*

Demographic	Whole cohort n= 100	Good surgical outcome n= 59	Bad surgical outcome n= 41	p value
Age (years)				
Mean	64.5	63.6	65.9	0.315
SD	11.3	11.6	10.7	
BMI				
Mean	25.9	26.8	24.8	0.113
SD	4.8	5.6	4.0	
Stage				
III	82 (82%)	54 (92%)	28 (68%)	0.003*
IV	18 (18%)	5 (8%)	13 (32%)	
ACE-27 (0-5)				
n (%)				
Low (0-1)	79 (79%)	44 (75%)	31 (76%)	0.999
High (2-5)	21 (21%)	15 (25%)	10 (24%)	
IMD centile (1-10) n (%)				
Low (1-5)				
High (6-10)	65 (65%)	41 (69%)	23 (56%)	0.2060
	35 (35%)	18 (31%)	18 (44%)	
PS (0-4)				
n (%)				
Fit (0,1)	77 (77%)	51 (86%)	28 (68%)	0.044*
Unfit (2,3,4)	23 (23%)	8 (18%)	13 (32%)	

**Indicates p<0.05, normally distributed data presented as mean and SD and non-normally distributed data as median and range.*

Table 7.2 *Comparison serum blood values for whole cohort*

Blood value		Whole cohort n = 100	Good surgical outcome n= 59	Bad surgical outcome n= 41	p value
Haemoglobin		(g/L)			
	Median	127	132	122	0.002*
	IQ Range	117 - 136	123 – 138	112 - 131	
Platelet		(10^9/L)			
	Median	337	312	381	0.002*
	IQ Range	272 - 418	264 - 397	283 - 500	
Neutrophil		(10^9/L)			
	Median	5.42	5.29	5.45	0.521
	IQ Range	4.21 – 7.00	4.08 – 7.04	4.55 – 7.06	
Lymphocyte		(10^9/L)			
	Median	1.65	1.93	1.30	<0.001*
	IQ Range	1.17 – 2.18	1.36 – 4.18	0.96 – 1.74	
Albumin		(g/L)			
	Median	37	37	36	0.014*
	IQ Range	35 - 39	35- 40	28 - 38	
CA 125		kU/L			
	Median	445	319	721	
	IQ Range	193 - 1262	94 - 798	316 - 2443	0.001*

**indicates p<0.05, IQ- interquartile, median and IQ range presented for non-normally distributed data*

The mean age of the patients was 64.5 ± 11.5 years, and BMI 25.9 ± 4.8m^2 with no significant difference between surgical outcome groups. There was no statistical difference between ACE-27 or IMD between the two groups, however patients in the bad surgical outcome group were less fit with more unfit patients in the good outcome group (13/41 (32%) vs 8/59 (18%) p=0.044).

There were several significant differences in the haematological markers between the two groups, with the bad surgical outcome group having lower haemoglobin levels (p=0.002), lymphocyte count (p=<0.001) and albumin (p=0.014), and a higher platelet count (p=0.002). There was no difference between neutrophil levels between the two groups. CA 125 was significantly higher in the bad surgical outcome group (median 721kU/L vs 319kU/L p=0.001).

7.4.1.2 Disease characteristics

All patients included in the study were diagnosed via histological analysis with high grade serous ovarian cancer of FIGO stage III and IV. The majority of patients (82/100, 82%) were FIGO stage III, with proportionately more stage IV patients in the bad than the good surgical outcome group (32/41, 32% vs 5/59, 8%, p=0.003).

Presumed disease distribution, defined from specialist radiologist reports at the time of pre-operative CT following MDT review is shown in table 7.3 and is similar between the two groups. The patients in the good surgical outcome group were more likely to have ascites reported on scan (85% vs 51%, p=<0.0001), and patients in the bad surgical group had slightly more "un-removable" disease noted on scan, although this was not statistically significant.

Table 7.3 *Comparison of reported disease distribution for whole cohort*

Positive mention on CT		Whole cohort n=100	Good surgical outcome n=59	Bad surgical outcome n=41	p value
Removable disease					
	n (%)				
		71 (71%)	42 (71%)	29 (71%)	0.820
Ascites					
	n (%)				
		71 (71%)	50 (85%)	21 (51%)	<0.0001*
Bowel disease					
	n (%)				
		18 (18%)	9 (15%)	9 (22%)	0.601
Nodal disease					
	n (%)				
		20 (20%)	11 (19%)	9 (22%)	0.802
Unremovable disease					
	n (%)				
		20 (20%)	9 (15%)	11 (27%)	0.210
Chest disease					
	n (%)				
		22 (22%)	10 (17)	12 (29%)	0.223

7.4.1.3 Surgical debulking rates and surgeon heterogeneity

Over the period of 2013 – 2018, 100 patients underwent PDS for stage III and IV HGSOC, with five specialist gynaecological oncology surgeons operating over this time period. All surgeons attended the same MDT, and on the whole operated independently of each other. Debulking status was defined by the lead surgeon at the end of the procedure and recorded on operative notes as per national guidance. Surgical outcome was recorded as either complete, optimal or suboptimal.

Survival of this cohort, stratified by surgical outcome, is shown in figure 7.3. Patients achieving complete debulking survived significantly longer than those with disease remaining (median survival of 69.4 months vs 38.7 months vs 23.1 months for complete debulk vs optimal vs

suboptimal, p=<0.0001). Patients with complete debulking rates survived significantly longer than those with optimal and suboptimal debulking (p=0.0187 vs p<0.0001), however there was no significant survival difference between optimal and suboptimal outcomes (p=0.0510). This could be attributed to small numbers in the cohorts. The three surgical outcomes were therefore separated into two distinct groups: good outcome (complete debulking) and bad outcome (optimal and suboptimal debulk).

The number of surgeries performed by surgeon and their surgical debulking rates demonstrated heterogeneity, figure 7.2, table 7.4. Good outcomes ranged between 71% (surgeon one) to 50% (surgeon two and three).

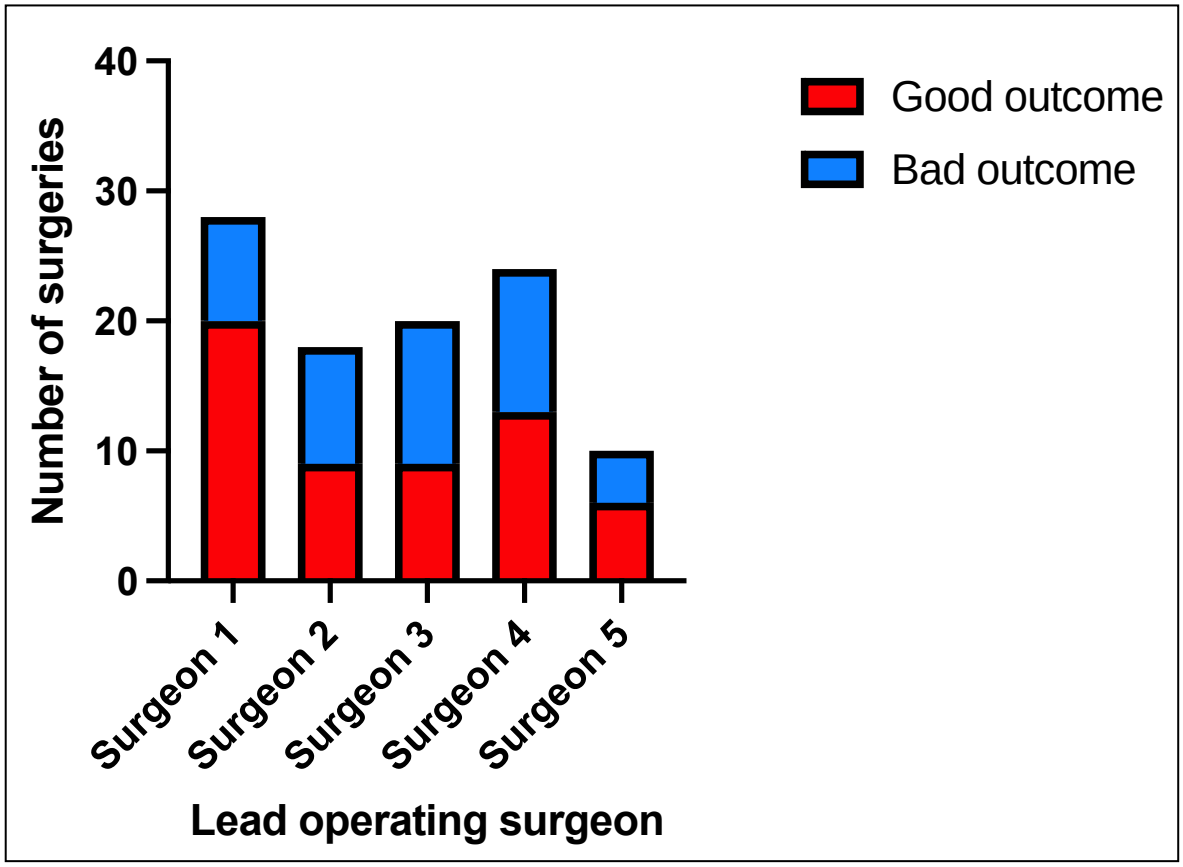


Figure 7.2 *Bar graph illustrating heterogeneity between surgeons*

Includes data for 5 operating surgeons over 100 surgeries.

Table 7.4 *Varying surgical outcome rates between surgeons*

Surgeon	Total surgeries n=100	Good outcome n= 59 (% per surgeon)	Bad outcome n= 41 (% per surgeon)
1	28	20 (71%)	8 (29%)
2	18	9 (50%)	9 (50%)
3	20	10 (50%)	10 (50%)
4	24	13 (54%)	11 (46%)
5	10	7 (70%)	3 (30%)

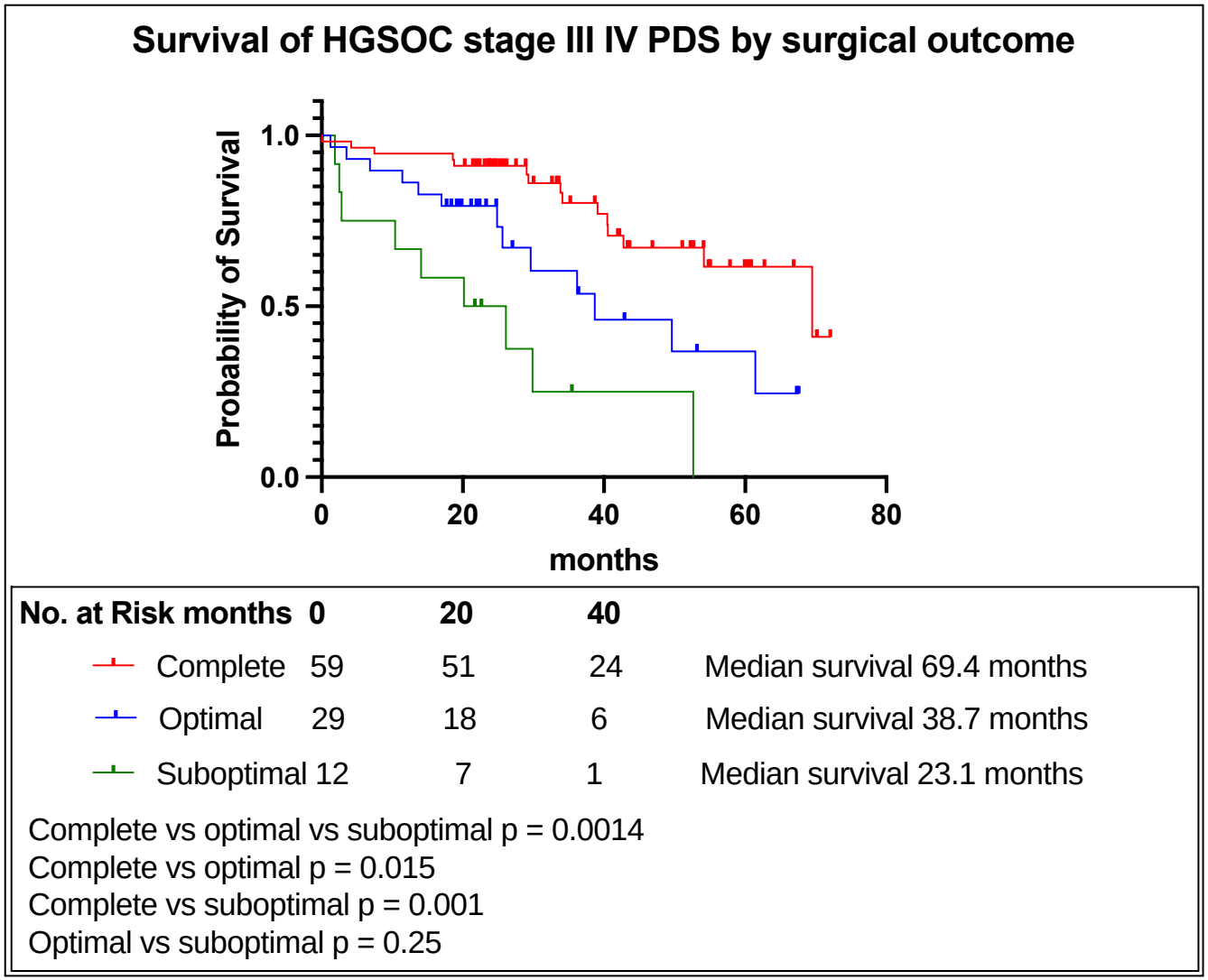


Figure 7.3 *Kaplan-Meier survival curve comparing median survival between surgical outcomes*

Log rank tests used for median survival comparison

7.4.2 Model building

7.4.2.1 Included predictors

The multivariable logistic regression model and the random forest model (machine learning) were built by applying the same 19 predictors, which included patient characteristics (age, BMI, PS, IMD, ACE-27), patient haematological markers (Hb, plt, lymphocyte and neutrophil count, albumin, CA 125), tumour characteristics (FIGO stage, CT disease distribution) and operating surgeon, table 7.5. Both models used the binary outcome of good or bad surgical outcome and aimed to predict bad surgical outcome. All continuous variables remained continuous and categorical variables were grouped as per table 7.1.

Within the categorical variables, event per predictor (EPP) number ranged between 3 events for surgeon 5, and 32 events for no nodal disease and no bowel disease. Despite this range, 76% of EPP were above the proposed minimum standard of EPP=10.

Of all selected predictors, higher levels of CA 125 ($p<0.0001$) and platelet count ($p=0.0036$) and lower levels of haemoglobin ($p<0.0001$), leucocyte count ($p=0.0003$) and albumin ($p=0.0144$) were associated with bad surgical outcome when univariable logistic regression was applied. There was no statistically significant association shown between any of the other included predictors.

Table 7.5 *Breakdown of al predictors included*

Predictor	p value (95% CI)	Variable outcomes	Event per predictor (bad surgical outcome)
Age	0.476	NA	
ECOG- PS	0.138	Fit Unfit	28 13
BMI	0.085	NA	
ACE-27 score	0.699	Low High	31 10
Indices of multiple deprivation score (IMD)	0.353	Low High	23 18
FIGO stage	0.670	III IV	28 13
CA 125 kU/L	<0.0001*	NA	
Haemoglobin (Hb) (g/L)	<0.0001*	NA	
Platelet count (plt) (10^9/L)	0.0036*	NA	
Lymphocyte count (10^9/L)	<0.0001*	NA	
Neutrophil count (10^9/L)	0.524	NA	
Albumin (g/L)	0.006*	NA	
un-removable disease	0.435	Yes No	11 30
removable disease	0.957	Yes No	29 12
Nodal disease	0.668	Yes No	9 32
Bowel disease	0.460	Yes No	9 32
Ascites	0.291	Yes No	21 20
Surgeon	0.367	1 2 3 4 5	8 9 10 11 3

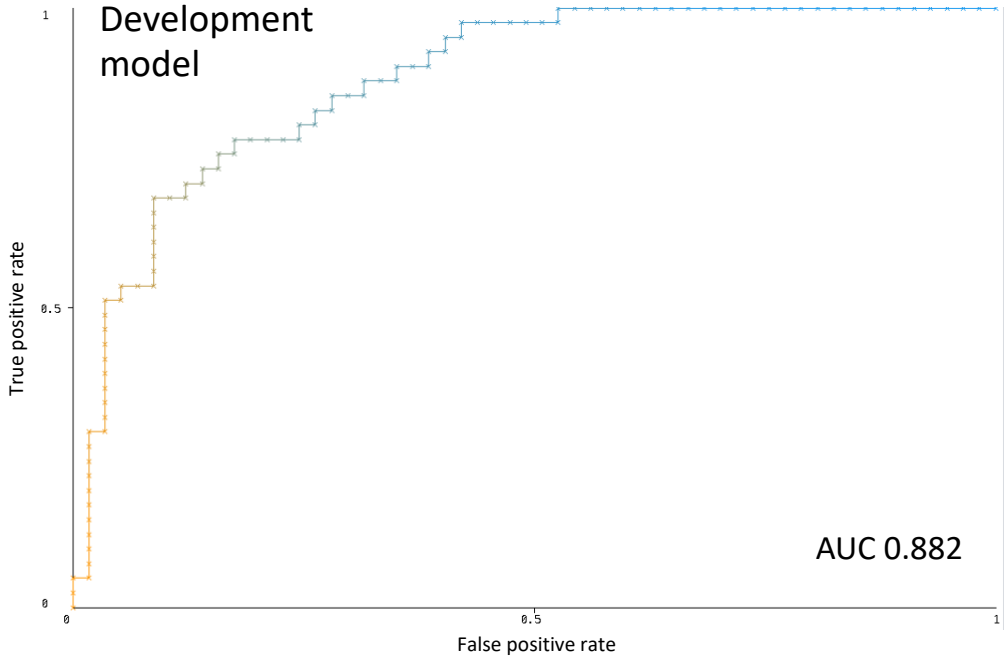
**indicates p<0.05*

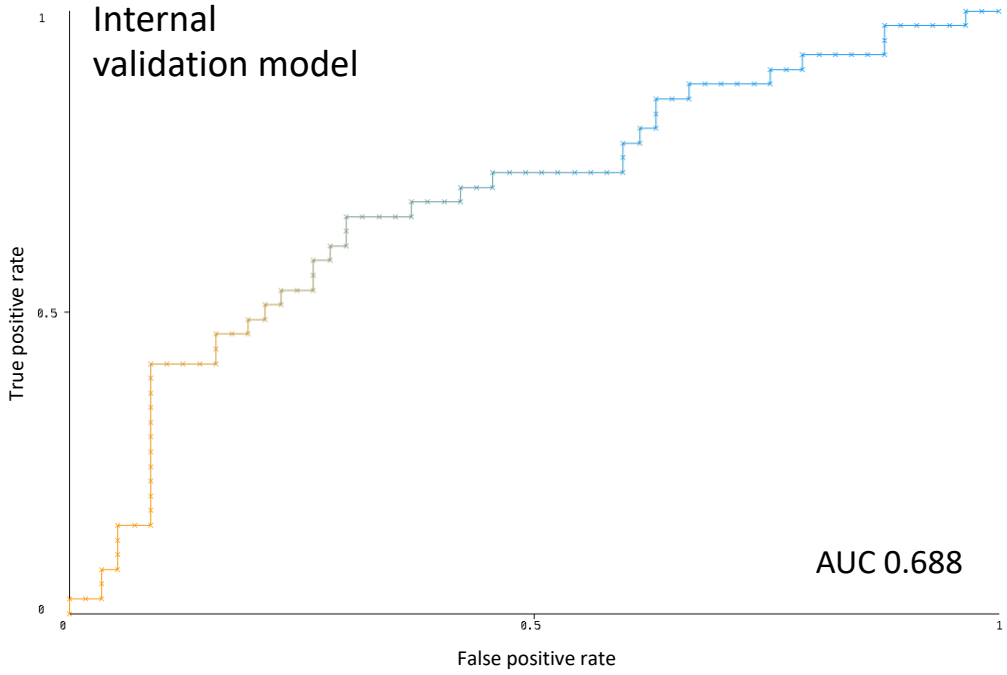
7.4.2.2 Multivariable logistic regression model for predicting bad surgical outcome

A logistic model was constructed to predict bad surgical outcome, using the data shown in table 7.5 for the 100 patients included in our cohort, incorporating 19 predictors and a binary surgical outcome of good or bad.

During initial development, the model demonstrated good calibration with a non-significant Hosmer-Lemeshow goodness of fit test of p=0.8920, indicating the model fits the data well. The model discrimination performance demonstrated an area under the curve of receiver operating characteristic curve (AUC ROC curve) to be 0.882 (95% CI 0.818 – 0.947). Sensitivity and specificity of the model were 79.4% and 79.6%, respectively. NPV and PPV of the model were 85.5% and 72%. The model demonstrated an overall accuracy of 79.5%.

When internal validation was applied to the model by way of leave-one-out cross validation to account for overfitting, model performance reduced, with an AUC of 0.688 (95% CI 0.611 – 0.712). Sensitivity and specificity also reduced to 61.5% and 70.1%, respectively, as did NPV and PPV (72.7% and 58.5%). Post internal validation, the overall accuracy of the model was 66.6%, figure 7.4.





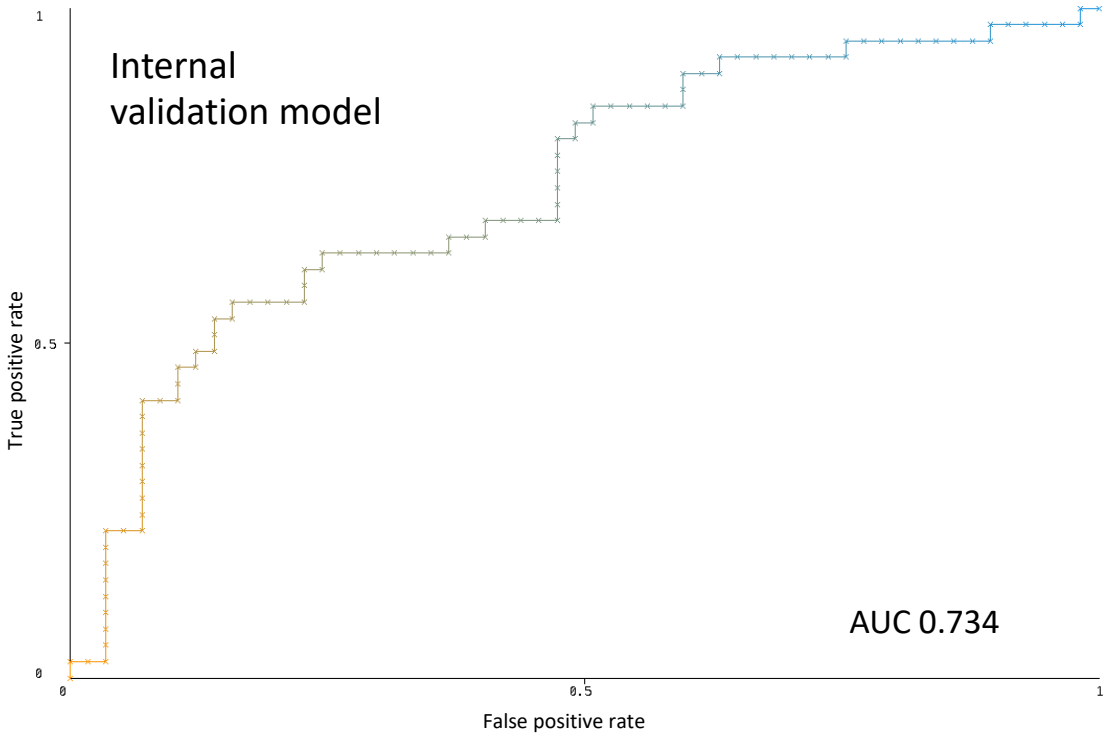
Model (95% CI)	AUC	Sensitivity	Specificity	NPV	PPV	Accuracy
Development	**0.882** *(0.818 – 0.947)*	**79.4%** *(63.5% - 90.7%)*	**79.6%** *(67.1% - 89%)*	**85.4%** *(75.7% - 91.7%)*	**72.0%** *(60.3% - 81.4%)*	**79.5%** *(70.2% - 87%)*
Internal validation	**0.688** *(0.611 – 0.712)*	**61.5%** *(44.6% - 76.6%)*	**70.1%** *(56.6% - 81.5%)*	**72.7%** *(63.4% - 80.4%)*	**58.5%** *(46.9% - 69.3%)*	**66.6%** *(56.3% - 75.9%)*

Figure 7.4 *ROC curve demonstrating calibration and discrimination performance measures of logistic regression prognostic model*

7.4.2.3 Random Forest Tree model for predicting bad surgical outcome

A Random Forest was constructed to predict bad surgical outcome for the 100 patients included in our cohort, incorporating 19 predictors and a binary surgical outcome of good or bad.

Following initial development and internal validation by way of leave-one-out cross validation to account for overfitting, the model discriminative performance demonstrated an AUC of 0.734, sensitivity and specificity of 75.8% and 71.6%, respectively, NPV and PPV of 87.2% and 53.6%, respectively, and an overall accuracy of 72.9%, figure 7.5.



Model (95% CI)	AUC	Sensitivity	Specificity	NPV	PPV	Accuracy
Internal validation	**0.734**	**75.8%** *(56.4% - 89.7%)*	**71.6%** *(59.3% - 81.9%)*	**87.2%** *(77.9% - 93.0%)*	**53.6%** *(42.9% - 64.0%)*	**72.9%** *(62.8% - 81.4%)*

Figure 7.5 *ROC curve demonstrating calibration and discrimination performance measures of Random Forest prognostic model*

7.5 Discussion

7.5.1 PROBAST assessment

As discussed in the systematic review, 4.2.5 a high risk of risk of bias exists in the majority of these published models in the current literature. Therefore, both presented multi-predictor prognostic models were assessed with the PROBAST tool.

7.5.1.1 Participants

Participants for both models were selected from a pre-existing database. Data were collected retrospectively, which is less desirable than prospectively collected data, however data collection was performed with the use of a pre-defined data dictionary and data collection guide in order to maintain consistency. There were no excluded patients, and the cohort of patients included were reflective of the target population for this tool.

7.5.1.2 Predictors

All selected predictors were defined and assessed in a consistent manner. As all patients were from a single centre, all haematological markers were assessed in the same lab, with the exclusion of CA 125. All CT scan assessments were performed by the same radiologists. All predictor assessments were performed without knowledge of outcome data, and all included predictors would be freely available to clinicians at the time of the model's intended use. All predictors included had clinical rationale or had been shown to be associated with surgical outcome in previous studies. Univariate analysis was avoided during predictor selection.

7.5.1.3 Outcome

Bad surgical outcome was clearly rationalised and defined and was applied in a similar way for all participants. However, at the time the outcome was determined by the surgeons, they may have been completely blind to all included predictors.

7.5.1.4 Analysis

The sample size for the model development was limited at 100 patients. Despite this, the majority of predictors had appropriate EPP numbers. However, a small number of predictors had an EPP <10, which increases the risk of bias due to overfitting. Data were handled appropriately with no dichotomization of continuous predictors. Several categorical predictors were grouped in order to increase EPP numbers. There was no exclusion of data, with all enrolled participants included in the analysis. All missing data was handled appropriately by way of multiple imputation. Model performance was analysed both by calibration and discrimination in both models, and overfitting was accounted for by internal validation using leave-one-out cross validation.

Overall both models were at high risk of risk of bias, mainly due to the small sample size and number of predictors with EPP <10, figure 7.6. However, this could be overcome by externally validating both models with a sample size of >100.

STUDY	ROB				APPLICABILITY			OVERALL	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	ROB	Applicability
Manchester logistic regression	L	L	L	H	L	L	L	H	L
Manchester Random Forest	L	L	L	H	L	L	L	H	L

ROB- risk of bias, L- Low, H- high

Figure 7.6 *PROBAST assessment for both prognostic models*

Two prediction models were developed. The first, developed via machine learning methods, performed well post internal validation (AUC 0.73, accuracy 73%). The second, developed using more traditional statistical methods, had a poorer performance post internal validation (AUC 0.68, accuracy 66%). Other than superseding current practice, there is no agreed level at which a model must perform to be of clinical use, however it is widely agreed that external validation is vital. The machine learning model outperformed the logistic regression model. However, it is possible that due to the small sample size, despite internal validation, the Random forest model may be overfitted to the data. For this reason, external validation of both models is a vital next step.

Patients included in the prediction model had already been assessed via the MDT decision making process, and therefore were patients selected for PDS. Consequently, this cohort of patients were thought to have a high chance of achieving complete cytoreduction, otherwise would have been put forward for NACT/IDS. This cohort is therefore a selected sub-cohort of the HGSOC population.

If these tools were to be externally validated, and used in clinical practice, they would need to be applied to the same population on which it was developed. Therefore, they should be applied after the original MDT decision has been made, as a separate triage step. For this reason, the tools would only act to reduce the number of patients operated on. However, if their implementation led to a reduced number of bad surgical outcomes, a reduction in morbidity and improvement in OS could occur.

Both developed models incorporated all data that would be available to clinicians at the time of discussion at MDT in the UK, in order to be as individualised as possible. However, the most successful currently published models rely on diagnostic laparoscopy. Neither of the two developed models incorporate surgical predictors. It is possible that the combination of our developed models with a laparoscopic scoring system might further improve model performance. Conversely to currently published models, our models attempt to address surgeon heterogeneity, with the aim of improving the chances of successful external validation.

Although our random forest model performed well following internal validation, it is not currently suitable for use in clinical practice. Firstly, external validation, ideally in multiple centres across multiple populations would need to be implemented successfully. Secondly, a prospective study to investigate the model as a pre-evaluation before diagnostic laparoscopy should also be considered. Notwithstanding these limitations, the concept of a multifactorial prediction tool is promising.

8 Discussion

8.1 Predicting surgical outcomes in ovarian cancer

Ovarian cancer is responsible for more deaths in women of the developed world than any other gynaecological cancers (Colombo et al., 2019). It is well established that surgical outcome is the most important surrogate marker of survival in advanced HGSOC (Griffiths et al., 1979; Marianne Jetske Rutten et al., 2015). It is of the utmost importance that the correct treatment pathway is selected for each patient in order to achieve the best surgical outcomes. Not only does this reduce morbidity and mortality, but also allows patients to commence medical treatment in a timely manner.

The treatment pathway decision is currently being made at the time of MDT, using mainly CT images, a modality known to have sensitivity and specificity for identifying intra-abdominal metastasis as low as 25% and 57%, respectively (Altman et al., 2012; Scott et al., 2020). With the current decision making process in clinical practice, rates of poor outcome at surgery stand between 9% and 76%, implying there are a large number of patients for whom PDS may not have been the correct pathway. These patients may have gained better survival benefit from NACT/IDS, or even no surgical management at all.

There are currently no clear guidelines, or validated prediction models to aid decision making in clinical use, and therefore decision trends vary nationally.

At present it is not routine for patients with suspected HGSOC to undergo a tissue biopsy prior to surgery. With the improved survival of patients with *BRCA* mutations and HRD tumours following treatment with PARP inhibitors proven, tissue samples pre-operatively may become more commonplace for better treatment planning (Banerjee et al., 2020; Moore et al., 2018). A tissue sample would also allow for analysis of gene panels, immunohistochemistry and further analysis of tissue biology, something that is often missed in the patients with the poorest outcome, as they die before undergoing surgery (Hawarden et al., 2021).

This thesis aimed to summarise current modalities used to predict surgical outcome at the time of PDS in HGSOC, as well as investigate possible future modalities and methods that may progress the field.

8.2 Summary of results

Chapter 4 outlined a systematic review identifying multiple different prognostic models with the goal of accurately predicting surgical outcome at the time of PDS. The models incorporated a wide range of predictors including clinical, biochemical, genomic, radiological and surgical markers. Many models demonstrated promise at the development stage, however when the PROBAST tool was applied, the majority of models displayed a high risk of bias. High risk of bias was most commonly attributed to model development, and the univariable logistic regression triage step in the predictor selection process. Only models utilising surgical markers at the time of laparoscopy were successfully externally validated. Absence of validation in the remaining models could be attributed to differing patient populations, as well as a lack of inclusion of surgical heterogeneity as a predictor.

Chapter 5 described the external validation of a promising three-protein signature that aimed to predict surgical outcome. The validation was performed using a large patient cohort from an RCT. The use of a protein signature to determine surgical outcome would be very appealing. Although a tissue biopsy would be required, once achieved, IHC could have been performed quickly, cheaply and readily. Unfortunately, the external validation failed, which may have been as a result of overfitting in the original model, alongside use of widely differing patient populations. The samples used for validation were also historic, and although stored appropriately, this could have affected the results.

Chapter 6 investigated the association between genotype (by gene panel) and phenotype (by HR assay) HR status, and surgical outcome. The TCGA is currently the largest readily available database where both debulking status and HR status are known. By determining HR status using an established gene panel and NGS, association with surgical outcome was shown in the TCGA cohort. However, this finding was not replicated in the MOCHR cohort. Absence of

correlation in the MOCHR cohort could be attributed to a smaller sample size, a combination of PDS and IDS samples being used, or the use of functional assay to determine HR status.

Chapter 7 discussed the development of two novel prognostic models to predict surgical outcome at the time of PDS. Two models were developed, one via logistic regression and one via machine learning techniques, and internally validated. The models were developed without the univariate analysis triage step and incorporated all available data available at the time of MDT discussion, including operating surgeon. The machine learning model out-performed the logistic regression model, which may have been due to overfitting of the logistic regression model. Validation using a larger cohort could help clarify the reasons for poorer model performance. Both models were developed on a sub cohort of all MDT patients, as they had been selected for PDS. Therefore, further validation and subsequent use of the model would need to be implemented following initial MDT discussion.

8.3 Significance of results presented

The systematic review presented in chapter 4 is the first to compile all PDS prognostic models and compare their risk of bias and validation status. The results and analysis will allow future developers to learn from predictors and methods used. The review also highlighted the models with the most promise, in order for them to be targeted for external validation, and forward movement towards use in clinical practice.

The three-protein panel was a model that had been identified as showing promise. Chapter 5 demonstrated a definitive failure of the model to externally validate when applied to a large cohort taken from multiple centres. This negative result is important, as it allows elimination of these predictors as future candidates.

Chapter 6 described the first reported assessment of HR status and surgical outcome. The result using the TCGA cohort shows definite promise. With the establishment of PARP inhibitors into routine clinical pathway, the *BRCA* and HR status of a patient will have enhanced relevance, which should allow for further cohorts to be available for validation.

The novel prediction models developed demonstrated the improved performance in a machine learning model over traditional statistics. The developed model avoided the elimination of any predictors prior to model building, and also factored in the operating surgeon. This allows for greater granularity, and therefore the model could be more likely to successfully validate on external cohorts. If successfully validated, the machine learning model could be used alongside clinician decision-making, as well as laparoscopic assessment.

8.4 Future work

Patients with an HRD status have shown improved surgical outcomes in the TCGA cohort. For this reason, further validation would be a good next step. The 100K genome project hold germline and somatic genomic data for 316 HGSOC patients collected from multiple centres across England, and analysis is underway. Work is currently ongoing to collect relevant clinical data fields to complement the available genomic data. Once complete, this dataset would be ideally placed for the external validation of the TCGA panel, as well as others available.

The machine learning prediction model described performed well when internally validated. External validation on a large multi-centre cohort is now required to progress its development towards clinical practice. The IMPRESS grant has been awarded by Ovarian Cancer Action as part of their IMPROVE programme which secures funding to take this project forward.

At present, all work described in this thesis focusses solely on predicting surgical outcome at the time of PDS. However, it would be of interest to broaden the prediction parameters to include overall survival, morbidity (of surgery and medical treatment) and quality of life. It is possible that future prediction tools will use predicted surgical outcome as one of a number of predictors, rather than as the outcome (figure 8.1). This would require a multi-step tool, likely incorporating laparoscopy and tissue biopsy. This tool would allow discussion with patients, to ascertain which factors hold most importance to them, with the treatment pathway being decided upon by all involved parties.

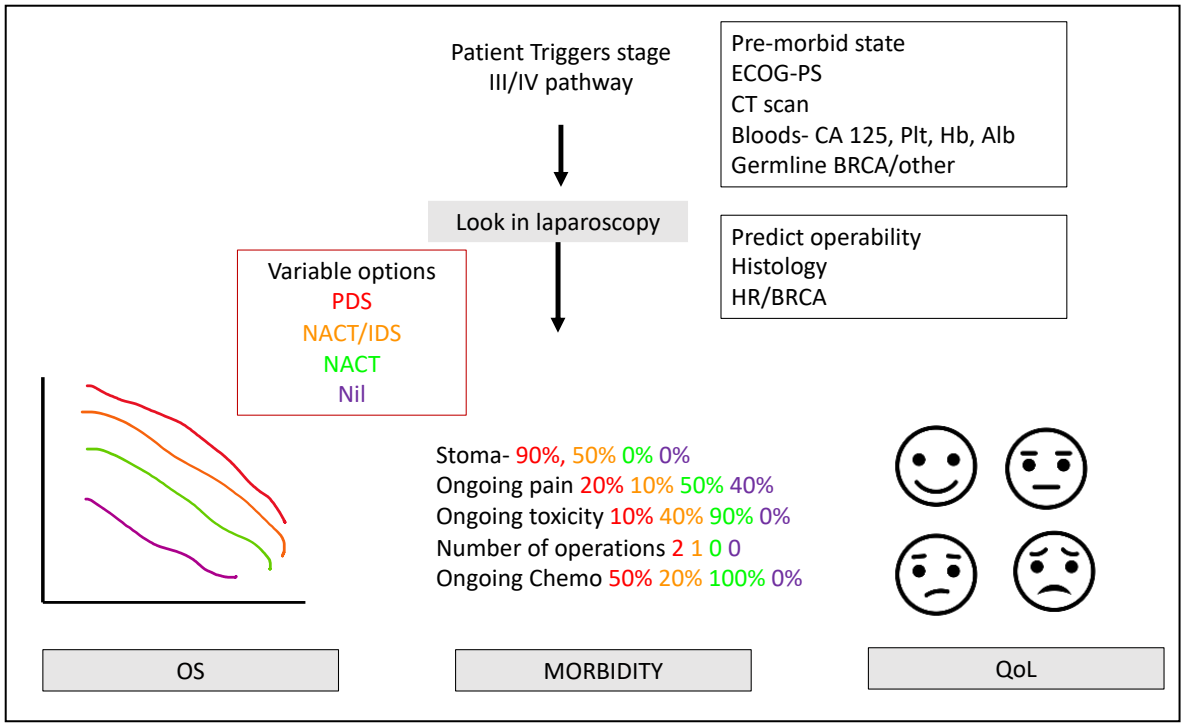


Figure 8.1 *Future aspirations of patient outcome prediction model*

QOL- Quality of life

8.5 Final conclusion

There are currently a significant proportion of patients undergoing major surgery and suffering suboptimal outcomes. These patients may be being subjected to increased morbidity, mortality and delay in commencing medical treatment. At present there is no established, or adequately validated tool to help make this decision.

Many modalities have shown promise, however the systematic review highlights the issue that models with single modalities often fail to validate. Models incorporating all available predictors including the operating surgeon require further validation. Future efforts should focus on the validation of promising models on large multi-centre cohorts as well as incorporating overall survival, morbidity and quality of life into models. Combining all of the above may allow for reduced suboptimal outcomes, improved mortality, morbidity and quality of life in patients with HGSOC.

9 References

- (NICE), N. I. f. H. a. c. E. (August 2019). *Olaparib for maintenance treatment of BRCA mutation-positive advanced ovarian, fallopian tube, or peritoneal cancer after response to first-line platinum based chemotherapy. Guideline TA 598*. NICE
- Abdallah, R., Chon, H. S., Bou Zgheib, N., Marchion, D. C., Wenham, R. M., Lancaster, J. M., & Gonzalez-Bosquet, J. (2015a). Prediction of Optimal Cytoreductive Surgery of Serous Ovarian Cancer With Gene Expression Data. *International Journal of Gynecological Cancer*, *25*(6), 1000-1009. https://doi.org/https://dx.doi.org/10.1097/IGC.0000000000000449
- Abdel-Azeez, H. A., Labib, H. A., Sharaf, S. M., & Refai, A. N. (2010). HE4 and mesothelin: novel biomarkers of ovarian carcinoma in patients with pelvic masses. *Asian Pac J Cancer Prev*, *11*(1), 111-116.
- ACOG Practice Bulletin. Management of adnexal masses. (2007). *Obstet Gynecol*, *110*(1), 201-214. https://doi.org/10.1097/01.aog.0000263913.92942.40
- Alberts, D. S., Green, S., Hannigan, E. V., O'Toole, R., Stock-Novack, D., Anderson, P., . . . Jolles, C. J. (1992). Improved therapeutic index of carboplatin plus cyclophosphamide versus cisplatin plus cyclophosphamide: final report by the Southwest Oncology Group of a phase III randomized trial in stages III and IV ovarian cancer. *J Clin Oncol*, *10*(5), 706-717. https://doi.org/10.1200/jco.1992.10.5.706
- Aletti, G. D., Dowdy, S. C., Gostout, B. S., Jones, M. B., Stanhope, C. R., Wilson, T. O., . . . Cliby, W. A. (2006). Aggressive surgical effort and improved survival in advanced-stage ovarian cancer. *Obstet Gynecol*, *107*(1), 77-85. https://doi.org/10.1097/01.AOG.0000192407.04428.bb
- Aletti, G. D., Dowdy, S. C., Gostout, B. S., Jones, M. B., Stanhope, R. C., Wilson, T. O., . . . Cliby, W. A. (2009). Quality improvement in the surgical approach to advanced ovarian cancer: the Mayo Clinic experience. *J Am Coll Surg*, *208*(4), 614-620. https://doi.org/10.1016/j.jamcollsurg.2009.01.006
- Aletti, G. D., Gostout, B. S., Podratz, K. C., & Cliby, W. A. (2006). Ovarian cancer surgical resectability: relative impact of disease, patient status, and surgeon. *Gynecol Oncol*, *100*(1), 33-37. https://doi.org/10.1016/j.ygyno.2005.07.123
- Alsina-Sanchís, E., Figueras, A., Lahiguera, A., Gil-Martín, M., Pardo, B., Piulats, J. M., . . . Viñals, F. (2017). TGFβ Controls Ovarian Cancer Cell Proliferation. *Int J Mol Sci*, *18*(8). https://doi.org/10.3390/ijms18081658
- Altman, A. D., Nelson, G., Chu, P., Nation, J., & Ghatage, P. (2012). Optimal Debulking Targets in Women With Advanced Stage Ovarian Cancer: A Retrospective Study of Immediate Versus Interval Debulking Surgery. *Journal of Obstetrics and Gynaecology Canada*, *34*(6), 558-566. https://doi.org/10.1016/S1701-2163(16)35272-0

- *american cancer society: evolution of cancer treatments:surgery.* Retrieved 19/02/2018 from
- Arab, M., Jamdar, F., Sadat Hosseini, M., Ghodssi- Ghasemabadi, R., Farzaneh, F., & Ashrafganjoei, T. (2018). Model for Prediction of Optimal Debulking of Epithelial Ovarian Cancer. *Asian Pacific journal of cancer prevention : APJCP*, *19*(5), 1319-1324.
- Armstrong, D. K., Bundy, B., Wenzel, L., Huang, H. Q., Baergen, R., Lele, S., . . . Burger, R. A. (2006). Intraperitoneal cisplatin and paclitaxel in ovarian cancer. *N Engl J Med*, *354*(1), 34-43. https://doi.org/10.1056/NEJMoa052985
- Ashrafganjoei, T., Mohamadianamiri, M., Farzaneh, F., Hosseini, M. S., & Arab, M. (2016). Investigating Preoperative Hematologic Markers for Prediction of Ovarian Cancer Surgical Outcome. *Asian Pac J Cancer Prev*, *17*(3), 1445-1448. https://doi.org/10.7314/apjcp.2016.17.3.1445
- Bailly, C., Bailly-Glatre, A., Alfidja, A., Vincent, C., Dauplat, J., & Pomel, C. (2009). [Peritoneal carcinosis in ovarian cancer: conventional imaging (CT-scan and MRI)]. *Bull Cancer*, *96*(12), 1155-1162. https://doi.org/10.1684/bdc.2009.0981 (Imagerie conventionnelle << péritonéale >> des cancers de l'ovaire (scanner, IRM).)
- Banerjee, S., Gonzalez-Martin, A., Harter, P., Lorusso, D., Moore, K. N., Oaknin, A., & Ray-Coquard, I. (2020). First-line PARP inhibitors in ovarian cancer: summary of an ESMO Open - Cancer Horizons round-table discussion. *ESMO Open*, *5*(6). https://doi.org/10.1136/esmoopen-2020-001110
- Bankhead, C. R., Kehoe, S. T., & Austoker, J. (2005). Symptoms associated with diagnosis of ovarian cancer: a systematic review. *Bjog*, *112*(7), 857-865. https://doi.org/10.1111/j.1471-0528.2005.00572.x
- Barclay, M., Gildea, C., Poole, J., Hirschowitz, L., Menon, U., & Nordin, A. (2016). Factors Affecting Short-term Mortality in Women With Ovarian, Tubal, or Primary Peritoneal Cancer: Population-Based Cohort Analysis of English National Cancer Registration Data. *Int J Gynecol Cancer*, *26*(1), 56-65. https://doi.org/10.1097/igc.0000000000000562
- Barnett, R. (2016). Ovarian cancer. *Lancet*, *387*(10025), 1265.
- Bast, R. C., Jr., Klug, T. L., St John, E., Jenison, E., Niloff, J. M., Lazarus, H., . . . Knapp, R. C. (1983). A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer. *N Engl J Med*, *309*(15), 883-887. https://doi.org/10.1056/nejm198310133091503
- Baxter, L., Ayres, C., Cohen, P., Kader Ali Mohan, R., & Leung, Y. (2019). EP797 20 years of ultra radical surgery for ovarian cancer patients in perth western australia, what have we learnt? *International Journal of Gynecologic Cancer*, *29*(Suppl 4), A438. https://doi.org/10.1136/ijgc-2019-ESGO.847
- Bedognetti, D., Spivey, T. L., Zhao, Y., Uccellini, L., Tomei, S., Dudley, M. E., . . . Marincola, F. M. (2013). CXCR3/CCR5 pathways in metastatic melanoma

patients treated with adoptive therapy and interleukin-2. *British Journal of Cancer*, *109*(9), 2412-2423. https://doi.org/10.1038/bjc.2013.557

- Benedet, J. L., Bender, H., Jones, H., 3rd, Ngan, H. Y., & Pecorelli, S. (2000). FIGO staging classifications and clinical practice guidelines in the management of gynecologic cancers. FIGO Committee on Gynecologic Oncology. *Int J Gynaecol Obstet*, *70*(2), 209-262.
- Beral, V., Doll, R., Hermon, C., Peto, R., & Reeves, G. (2008). Ovarian cancer and oral contraceptives: collaborative reanalysis of data from 45 epidemiological studies including 23,257 women with ovarian cancer and 87,303 controls. *Lancet*, *371*(9609), 303-314. https://doi.org/10.1016/s0140-6736(08)60167-1
- Berchuck, A., Iversen, E. S., Lancaster, J. M., Dressman, H. K., West, M., Nevins, J. R., & Marks, J. R. (2004). Prediction of optimal versus suboptimal cytoreduction of advanced-stage serous ovarian cancer with the use of microarrays. *Am J Obstet Gynecol*, *190*(4), 910-925. https://doi.org/10.1016/j.ajog.2004.02.005
- Bernard, A. (2017). Clinical prediction models: a fashion or a necessity in medicine? In *J Thorac Dis* (Vol. 9, pp. 3456-3457). https://doi.org/10.21037/jtd.2017.09.42
- Bian, X., Gao, J., Luo, F., Rui, C., Zheng, T., Wang, D., . . . Cheng, H. (2018). PTEN deficiency sensitizes endometrioid endometrial cancer to compound PARP-PI3K inhibition but not PARP inhibition as monotherapy. *Oncogene*, *37*(3), 341-351. https://doi.org/10.1038/onc.2017.326
- Bookman, M. A., Brady, M. F., McGuire, W. P., Harper, P. G., Alberts, D. S., Friedlander, M., . . . Roth, L. M. (2009). Evaluation of new platinum-based treatment regimens in advanced-stage ovarian cancer: a Phase III Trial of the Gynecologic Cancer Intergroup. *J Clin Oncol*, *27*(9), 1419-1425. https://doi.org/10.1200/jco.2008.19.1684
- Borley, J., Wilhelm-Benartzi, C., Yazbek, J., Williamson, R., Bharwani, N., Stewart, V., . . . Ghaem-Maghami, S. (2015). Radiological predictors of cytoreductive outcomes in patients with advanced ovarian cancer. *Bjog*, *122*(6), 843-849. https://doi.org/10.1111/1471-0528.12992
- Bristow, R. E., Chang, J., Ziogas, A., Campos, B., Chavez, L. R., & Anton-Culver, H. (2015). Impact of National Cancer Institute Comprehensive Cancer Centers on ovarian cancer treatment and survival. *J Am Coll Surg*, *220*(5), 940-950. https://doi.org/10.1016/j.jamcollsurg.2015.01.056
- Bristow, R. E., & Chi, D. S. (2006). Platinum-based neoadjuvant chemotherapy and interval surgical cytoreduction for advanced ovarian cancer: a meta-analysis. *Gynecol Oncol*, *103*(3), 1070-1076. https://doi.org/10.1016/j.ygyno.2006.06.025
- Bristow, R. E., Karian, B. Y., & Chi, D. Surgery for Ovarian Cancer 3rd Edition. In.

- Bristow, R. E., Karian, B. Y., & Chi, D. (2016). Surgery for Ovarian Cancer 3rd Edition. In.
- Bristow, R. E., Tomacruz, R. S., Armstrong, D. K., Trimble, E. L., & Montz, F. J. (2002). Survival effect of maximal cytoreductive surgery for advanced ovarian carcinoma during the platinum era: a meta-analysis. *J Clin Oncol*, *20*(5), 1248-1259. https://doi.org/10.1200/jco.2002.20.5.1248
- Buys, S. S., Partridge, E., Black, A., Johnson, C. C., Lamerato, L., Isaacs, C., . . . Berg, C. D. (2011). Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Randomized Controlled Trial. *Jama*, *305*(22), 2295-2303. https://doi.org/10.1001/jama.2011.766
- Cabasag, C. J., Butler, J., Arnold, M., Rutherford, M., Bardot, A., Ferlay, J., . . . Soerjomataram, I. (2020). Exploring variations in ovarian cancer survival by age and stage (ICBP SurvMark-2): A population-based study. *Gynecol Oncol*, *157*(1), 234-244. https://doi.org/10.1016/j.ygyno.2019.12.047
- Capoluongo, E., Ellison, G., López-Guerrero, J. A., Penault-Llorca, F., Ligtenberg, M. J. L., Banerjee, S., . . . de Castro, D. G. (2017). Guidance Statement On BRCA1/2 Tumor Testing in Ovarian Cancer Patients. *Semin Oncol*, *44*(3), 187-197. https://doi.org/10.1053/j.seminoncol.2017.08.004
- Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., . . . Schultz, N. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*, *2*(5), 401-404. https://doi.org/10.1158/2159-8290.cd-12-0095
- Chang, C., Chiang, A. J., Chen, W. A., Chang, H. W., & Chen, J. (2016). A joint model based on longitudinal CA125 in ovarian cancer to predict recurrence. *Biomarkers in Medicine*, *10*(1), 53-61. https://doi.org/http://dx.doi.org/10.2217/bmm.15.110
- Chang, L. C., Huang, C. F., Lai, M. S., Shen, L. J., Wu, F. L., & Cheng, W. F. (2018). Prognostic factors in epithelial ovarian cancer: A population-based study. *PLoS One*, *13*(3), e0194993. https://doi.org/10.1371/journal.pone.0194993
- Chen, J. P., Huang, Q. D., Wan, T., Tu, H., Gu, H. F., Cao, J. Y., & Liu, J. H. (2019). Combined score of pretreatment platelet count and CA125 level (PLT-CA125) stratified prognosis in patients with FIGO stage IV epithelial ovarian cancer. *J Ovarian Res*, *12*(1), 72. https://doi.org/10.1186/s13048-019-0544-y
- Chen, L. (2020). Overview of clinical prediction models. In *Ann Transl Med* (Vol. 8, pp. 71). https://doi.org/10.21037/atm.2019.11.121
- Chern, J. Y., & Curtin, J. P. (2016). Appropriate Recommendations for Surgical Debulking in Stage IV Ovarian Cancer. *Curr Treat Options Oncol*, *17*(1), 1. https://doi.org/10.1007/s11864-015-0380-2
- Chesnais, M., Lecuru, F., Mimouni, M., Ngo, C., Fauconnier, A., & Huchon, C. (2017). A pre-operative predictive score to evaluate the feasibility of complete

cytoreductive surgery in patients with epithelial ovarian cancer. *PLoS One*, *12*(11), e0187245. https://doi.org/10.1371/journal.pone.0187245

- Chi, D. S., Eisenhauer, E. L., Lang, J., Huh, J., Haddad, L., Abu-Rustum, N. R., . . . Barakat, R. R. (2006). What is the optimal goal of primary cytoreductive surgery for bulky stage IIIC epithelial ovarian carcinoma (EOC)? *Gynecologic Oncology*, *103*(2), 559-564. https://doi.org/10.1016/j.ygyno.2006.03.051
- Chi, D. S., Eisenhauer, E. L., Zivanovic, O., Sonoda, Y., Abu-Rustum, N. R., Levine, D. A., . . . Barakat, R. R. (2009). Improved progression-free and overall survival in advanced ovarian cancer as a result of a change in surgical paradigm. *Gynecol Oncol*, *114*(1), 26-31. https://doi.org/10.1016/j.ygyno.2009.03.018
- Chi, D. S., Ramirez, P. T., Teitcher, J. B., Mironov, S., Sarasohn, D. M., Iyer, R. B., . . . Barakat, R. R. (2007). Prospective study of the correlation between postoperative computed tomography scan and primary surgeon assessment in patients with advanced ovarian, tubal, and peritoneal carcinoma reported to have undergone primary surgical cytoreduction to residual disease 1 cm or less. *J Clin Oncol*, *25*(31), 4946-4951. https://doi.org/10.1200/jco.2007.12.2317
- Chiang, Y. C., Lin, P. H., & Cheng, W. F. (2021). Homologous Recombination Deficiency Assays in Epithelial Ovarian Cancer: Current Status and Future Direction. *Front Oncol*, *11*, 675972. https://doi.org/10.3389/fonc.2021.675972
- Choi, K. U., Yun, J. S., Lee, I. H., Heo, S. C., Shin, S. H., Jeon, E. S., . . . Kim, J. H. (2011). Lysophosphatidic acid-induced expression of periostin in stromal cells: Prognoistic relevance of periostin expression in epithelial ovarian cancer. *Int J Cancer*, *128*(2), 332-342. https://doi.org/10.1002/ijc.25341
- Chong, G. O., Jeong, S. Y., Lee, Y. H., Lee, H. J., Lee, S. W., Han, H. S., . . . Lee, Y. S. (2019a). The ability of whole-body SUVmax in F-18 FDG PET/CT to predict suboptimal cytoreduction during primary debulking surgery for advanced ovarian cancer. *J Ovarian Res*, *12*(1), 12. https://doi.org/10.1186/s13048-019-0488-2
- https://doi.org/https://dx.doi.org/10.1186/s13048-019-0488-2
- Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*, *110*, 12-22. https://doi.org/10.1016/j.jclinepi.2019.02.004
- Chudecka-Glaz, A. M., Cymbaluk-Ploska, A. A., Menkiszak, J. L., Sompolska-Rzechula, A. M., Toloczko-Grabarek, A. I., & Rzepka-Gorska, I. A. (2014). Serum HE4, CA125, YKL-40, bcl-2, cathepsin-L and prediction optimal debulking surgery, response to chemotherapy in ovarian cancer. *J Ovarian Res*, *7*, 62. https://doi.org/10.1186/1757-2215-7-62
- Coakley, F. V., Choi, P. H., Gougoutas, C. A., Pothuri, B., Venkatraman, E., Chi, D., . . . Hricak, H. (2002). Peritoneal metastases: detection with spiral CT in

patients with ovarian cancer. *Radiology*, *223*(2), 495-499. https://doi.org/10.1148/radiol.2232011081

- Collins, G. S., de Groot, J. A., Dutton, S., Omar, O., Shanyinde, M., Tajar, A., . . . Altman, D. G. (2014). External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol*, *14*, 40. https://doi.org/10.1186/1471-2288-14-40
- Colombo, N., Guthrie, D., Chiari, S., Parmar, M., Qian, W., Swart, A. M., . . . Bonazzi, C. (2003). International Collaborative Ovarian Neoplasm trial 1: a randomized trial of adjuvant chemotherapy in women with early-stage ovarian cancer. *J Natl Cancer Inst*, *95*(2), 125-132. https://doi.org/10.1093/jnci/95.2.125
- Colombo, N., Sessa, C., du Bois, A., Ledermann, J., McCluggage, W. G., McNeish, I., . . . Querleu, D. (2019). ESMO-ESGO consensus conference recommendations on ovarian cancer: pathology and molecular biology, early and advanced stages, borderline tumours and recurrent disease†. *Ann Oncol*, *30*(5), 672-705. https://doi.org/10.1093/annonc/mdz062
- CRUK. (2018). *Cancer research UK*. Retrieved 14/02/2018 from
- Damia, G., & Broggini, M. (2019). Platinum Resistance in Ovarian Cancer: Role of DNA Repair. *Cancers*, *11*(1), 119.
- Danforth, K. N., Tworoger, S. S., Hecht, J. L., Rosner, B. A., Colditz, G. A., & Hankinson, S. E. (2007). A prospective study of postmenopausal hormone use and ovarian cancer risk. *Br J Cancer*, *96*(1), 151-156. https://doi.org/10.1038/sj.bjc.6603527
- de Jong, D., Eijkemans, M. J., Fong, S. L., Gerestein, C. G., Kooi, G. S., Baalbergen, A., . . . Ansink, A. C. (2007). Preoperative predictors for residual tumor after surgery in patients with ovarian carcinoma. *Oncology*, *72*(5-6), 293-301. https://doi.org/10.1159/000113051
- Delgado, G., Oram, D. H., & Petrilli, E. S. (1984). Stage III epithelial ovarian cancer: the role of maximal surgical reduction. *Gynecol Oncol*, *18*(3), 293-298. https://doi.org/10.1016/0090-8258(84)90040-4
- Diniz Bizzo, S. M., Meira, D. D., Lima, J. M., Mororo Jda, S., Casali-da-Rocha, J. C., & Ornellas, M. H. (2010). Peritoneal VEGF burden as a predictor of cytoreductive surgery outcome in women with epithelial ovarian cancer. *Int J Gynaecol Obstet*, *109*(2), 113-117. https://doi.org/10.1016/j.ijgo.2009.11.021
- Doufekas, K., & Olaitan, A. (2014). Clinical epidemiology of epithelial ovarian cancer in the UK. *Int J Womens Health*, *6*, 537-545. https://doi.org/10.2147/ijwh.s40894
- Dowdy, S. C., Loewen, R. T., Aletti, G., Feitoza, S. S., & Cliby, W. (2008). Assessment of outcomes and morbidity following diaphragmatic peritonectomy for women with ovarian carcinoma. *Gynecol Oncol*, *109*(2), 303-307. https://doi.org/10.1016/j.ygyno.2008.02.012

- Dubeau, L. (1999). The cell of origin of ovarian epithelial tumors and the ovarian surface epithelium dogma: does the emperor have no clothes? *Gynecol Oncol*, *72*(3), 437-442. https://doi.org/10.1006/gyno.1998.5275
- Economou, M., Schöni, L., Hammer, C., Galván, J. A., Mueller, D. E., & Zlobec, I. (2014). Proper paraffin slide storage is crucial for translational research projects involving immunohistochemistry stains. In *Clin Transl Med* (Vol. 3, pp. 4). https://doi.org/10.1186/2001-1326-3-4
- Eisenhauer, E. L., Abu-Rustum, N. R., Sonoda, Y., Levine, D. A., Poynor, E. A., Aghajanian, C., . . . Chi, D. S. (2006). The addition of extensive upper abdominal surgery to achieve optimal cytoreduction improves survival in patients with stages IIIC-IV epithelial ovarian cancer. *Gynecol Oncol*, *103*(3), 1083-1090. https://doi.org/10.1016/j.ygyno.2006.06.028
- Eisenkop, S. M., Spirtos, N. M., Friedman, R. L., Lin, W. C., Pisani, A. L., & Perticucci, S. (2003). Relative influences of tumor volume before surgery and the cytoreductive outcome on survival for patients with advanced ovarian cancer: a prospective study. *Gynecol Oncol*, *90*(2), 390-396. https://doi.org/10.1016/s0090-8258(03)00278-6
- Enshaei, A., Robson, C. N., & Edmondson, R. J. (2015). Artificial Intelligence Systems as Prognostic and Predictive Tools in Ovarian Cancer. *Ann Surg Oncol*, *22*(12), 3970-3975. https://doi.org/10.1245/s10434-015-4475-6
- Eriksen, M., & Frandsen, T. (2018). The impact of patient, intervention, comparison, outcome (PICO) as a search strategy tool on literature search quality: a systematic review. *Journal of the Medical Library Association : JMLA*, *106*, 420-431. https://doi.org/10.5195/jmla.2018.345
- Ezzati, M., Abdullah, A., Shariftabrizi, A., Hou, J., Kopf, M., Stedman, J. K., . . . Shahabi, S. (2014). Recent Advancements in Prognostic Factors of Epithelial Ovarian Carcinoma. *Int Sch Res Notices*, *2014*, 953509. https://doi.org/10.1155/2014/953509
- FA., P. (1949). Carcinoma of the ovary. *American Journal of Obstetrics and Gynaecology*, *58*(4), 640-653.
- Fago-Olsen, C. L., Ottesen, B., Christensen, I. J., Hogdall, E., Lundvall, L., Nedergaard, L., . . . Hogdall, C. (2014). Biomarkers for predicting complete debulking in ovarian cancer: lessons to be learned [Research Support, Non-U.S. Gov't]. *Anticancer Research*, *34*(2), 679-682.
- Fagotti, A., Ferrandina, G., Fanfani, F., Ercoli, A., Lorusso, D., Rossi, M., & Scambia, G. (2006). A laparoscopy-based score to predict surgical outcome in patients with advanced ovarian carcinoma: a pilot study. *Ann Surg Oncol*, *13*(8), 1156-1161. https://doi.org/10.1245/aso.2006.08.021
- Fagotti, A., Vizzielli, G., Fanfani, F., Costantini, B., Ferrandina, G., Gallotta, V., . . . Scambia, G. (2013). Introduction of staging laparoscopy in the management of advanced epithelial ovarian, tubal and peritoneal cancer: impact on prognosis in a single institution experience. *Gynecol Oncol*, *131*(2), 341-346. https://doi.org/10.1016/j.ygyno.2013.08.005

- Fathalla, M. F. (1971). Incessant ovulation--a factor in ovarian neoplasia? *Lancet*, *2*(7716), 163. https://doi.org/10.1016/s0140-6736(71)92335-x
- Feng, L.-Y., Liao, S.-B., & Li, L. (2020). Preoperative serum levels of HE4 and CA125 predict primary optimal cytoreduction in advanced epithelial ovarian cancer: a preliminary model study. *Journal of ovarian research*, *13*(1), 17-17. https://doi.org/10.1186/s13048-020-0614-1
- Feng, Z., Wen, H., Bi, R., Duan, Y., Yang, W., & Wu, X. (2016). Thrombocytosis and hyperfibrinogenemia are predictive factors of clinical outcomes in high-grade serous ovarian cancer patients. *BMC Cancer*, *16*, 43. https://doi.org/10.1186/s12885-016-2070-2
- Feng, Z., Wen, H., Jiang, Z., Liu, S., Ju, X., Chen, X., . . . Wu, X. (2018). A triage strategy in advanced ovarian cancer management based on multiple predictive models for R0 resection: a prospective cohort study. *Journal of Gynecologic Oncology*, *29*(5), e65. https://doi.org/https://dx.doi.org/10.3802/jgo.2018.29.e65
- Fleissig, A., Jenkins, V., Catt, S., & Fallowfield, L. (2006). Multidisciplinary teams in cancer care: are they effective in the UK? *Lancet Oncol*, *7*(11), 935-943. https://doi.org/10.1016/s1470-2045(06)70940-8
- Forse, C. L., Pinnaduwage, D., Bull, S. B., Mulligan, A. M., & Andrulis, I. L. (2019). Fresh Cut Versus Stored Cut Paraffin-embedded Tissue: Effect on Immunohistochemical Staining for Common Breast Cancer Markers. *Appl Immunohistochem Mol Morphol*, *27*(3), 231-237. https://doi.org/10.1097/pai.0000000000000579
- Fotopoulou, C., Hall, M., Cruickshank, D., Gabra, H., Ganesan, R., Hughes, C., . . . Sundar, S. (2017). British Gynaecological Cancer Society (BGCS) epithelial ovarian/fallopian tube/primary peritoneal cancer guidelines: recommendations for practice. *Eur J Obstet Gynecol Reprod Biol*, *213*, 123-139. https://doi.org/10.1016/j.ejogrb.2017.04.016
- Frampton, G. M., Fichtenholtz, A., Otto, G. A., Wang, K., Downing, S. R., He, J., . . . Yelensky, R. (2013). Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat Biotechnol*, *31*(11), 1023-1031. https://doi.org/10.1038/nbt.2696
- Frey, M. K., & Pothuri, B. (2017). Homologous recombination deficiency (HRD) testing in ovarian cancer clinical practice: a review of the literature. *Gynecol Oncol Res Pract*, *4*, 4. https://doi.org/10.1186/s40661-017-0039-8
- Fujwara, K., Yoshino, K., Enomoto, T., Fujita, M., Ueda, Y., Miyatake, T., . . . Hori, M. (2011). Usefulness of computed tomography in predicting cytoreductive surgical outcomes for ovarian cancer. *Arch Gynecol Obstet*, *284*(6), 1501-1507. https://doi.org/10.1007/s00404-011-1864-3
- Funston, G., Van Melle, M., Baun, M.-L. L., Jensen, H., Helsper, C., Emery, J., . . . Walter, F. M. (2019). Variation in the initial assessment and investigation for ovarian cancer in symptomatic women: a systematic review

of international guidelines. *BMC Cancer*, *19*(1), 1028. https://doi.org/10.1186/s12885-019-6211-2

- Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., . . . Schultz, N. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*, *6*(269), pl1. https://doi.org/10.1126/scisignal.2004088
- Ge, L. N., & Wang, F. (2018). Prognostic significance of preoperative serum albumin in epithelial ovarian cancer patients: a systematic review and dose-response meta-analysis of observational studies. *Cancer Manag Res*, *10*, 815-825. https://doi.org/10.2147/cmar.s161876
- Gee, M. E., Faraahi, Z., McCormick, A., & Edmondson, R. J. (2018). DNA damage repair in ovarian cancer: unlocking the heterogeneity. *J Ovarian Res*, *11*(1), 50. https://doi.org/10.1186/s13048-018-0424-x
- Gemer, O., Gdalevich, M., Ravid, M., Piura, B., Rabinovich, A., Gasper, T., . . . Lavie, O. (2009). A multicenter validation of computerized tomography models as predictors of non- optimal primary cytoreduction of advanced epithelial ovarian cancer. *Eur J Surg Oncol*, *35*(10), 1109-1112. https://doi.org/10.1016/j.ejso.2009.03.002
- Geomini, P., Kruitwagen, R., Bremer, G. L., Cnossen, J., & Mol, B. W. (2009). The accuracy of risk scores in predicting ovarian malignancy: a systematic review. *Obstet Gynecol*, *113*(2 Pt 1), 384-394. https://doi.org/10.1097/AOG.0b013e318195ad17
- Gerestein, C. G., Eijkemans, M. J., Bakker, J., Elgersma, O. E., Van der Burg, M. E. L., Kooi, G. S., & Burger, C. W. (2011). Nomogram for Suboptimal Cytoreduction at Primary Surgery for Advanced Stage Ovarian Cancer. *Anticancer Research*, *31*(11), 4043-4049.
- Gill, S. E., McGree, M. E., Weaver, A. L., Cliby, W. A., & Langstraat, C. L. (2017). Optimizing the treatment of ovarian cancer: Neoadjuvant chemotherapy and interval debulking versus primary debulking surgery for epithelial ovarian cancers likely to have suboptimal resection. *Gynecol Oncol*, *144*(2), 266-273. https://doi.org/10.1016/j.ygyno.2016.11.021
- Gillan, L., Matei, D., Fishman, D. A., Gerbin, C. S., Karlan, B. Y., & Chang, D. D. (2002). Periostin secreted by epithelial ovarian carcinoma is a ligand for alpha(V)beta(3) and alpha(V)beta(5) integrins and promotes cell motility. *Cancer Res*, *62*(18), 5358-5364.
- González-Martín, A., Pothuri, B., Vergote, I., DePont Christensen, R., Graybill, W., Mirza, M. R., . . . Monk, B. J. (2019). Niraparib in Patients with Newly Diagnosed Advanced Ovarian Cancer. *N Engl J Med*, *381*(25), 2391-2402. https://doi.org/10.1056/NEJMoa1910962
- Graeser, M., McCarthy, A., Lord, C. J., Savage, K., Hills, M., Salter, J., . . . Turner, N. C. (2010). A marker of homologous recombination predicts pathologic complete response to neoadjuvant chemotherapy in primary

breast cancer. *Clin Cancer Res*, *16*(24), 6159-6168. https://doi.org/10.1158/1078-0432.ccr-10-1027

- Griffiths, C. T. (1975). Surgical resection of tumor bulk in the primary treatment of ovarian carcinoma. *Natl Cancer Inst Monogr*, *42*, 101-104.
- Griffiths, C. T., Parker, L. M., & Fuller, A. F., Jr. (1979). Role of cytoreductive surgical treatment in the management of advanced ovarian cancer. *Cancer Treat Rep*, *63*(2), 235-240.
- Gu, B., Xia, L., Ge, H., & Liu, S. (2020). Preoperative PET/CT score can predict complete resection in advanced epithelial ovarian cancer: a prospective study. *Quant Imaging Med Surg*, *10*(3), 743-753. https://doi.org/10.21037/qims.2020.02.19
- Gu, Y., Qin, M., Jin, Y., Zuo, J., Li, N., Bian, C., . . . Pan, L. Y. (2020). A Prediction Model for Optimal Primary Debulking Surgery Based on Preoperative Computed Tomography Scans and Clinical Factors in Patients With Advanced Ovarian Cancer: A Multicenter Retrospective Cohort Study. *Front Oncol*, *10*, 611617. https://doi.org/10.3389/fonc.2020.611617
- Gupta, D., & Lis, C. G. (2009). Role of CA125 in predicting ovarian cancer survival - a review of the epidemiological literature. *J Ovarian Res*, *2*, 13. https://doi.org/10.1186/1757-2215-2-13
- Hacker, N. F., Berek, J. S., Lagasse, L. D., Nieberg, R. K., & Elashoff, R. M. (1983). Primary cytoreductive surgery for epithelial ovarian cancer. *Obstet Gynecol*, *61*(4), 413-420.
- Hacker, N. F., & Rao, A. (2017). Surgery for advanced epithelial ovarian cancer. *Best Pract Res Clin Obstet Gynaecol*, *41*, 71-87. https://doi.org/10.1016/j.bpobgyn.2016.10.007
- Hamilton, C. A., Miller, A., Miller, C., Krivak, T. C., Farley, J. H., Chernofsky, M. R., . . . Maxwell, G. L. (2011). The impact of disease distribution on survival in patients with stage III epithelial ovarian cancer cytoreduced to microscopic residual: a Gynecologic Oncology Group study. *Gynecol Oncol*, *122*(3), 521-526. https://doi.org/10.1016/j.ygyno.2011.04.041
- Harrell, F. E., Jr., Lee, K. L., & Mark, D. B. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*, *15*(4), 361-387. https://doi.org/10.1002/(sici)1097-0258(19960229)15:4<361::aid-sim168>3.0.co;2-4
- Haward, R. A. (2006). The Calman-Hine report: a personal retrospective on the UK's first comprehensive policy on cancer services. *Lancet Oncol*, *7*(4), 336-346. https://doi.org/10.1016/s1470-2045(06)70659-3
- Hawarden, A., Russell, B., Gee, M. E., Kayali, F., Clamp, A., Crosbie, E. J., & Edmondson, R. J. (2021). Factors determining ultra-short-term survival and the commencement of active treatment in high-grade serous ovarian cancer: a case comparison study. *BMC Cancer*, *21*(1), 378. https://doi.org/10.1186/s12885-021-08019-9

- Hayden, J. A., van der Windt, D. A., Cartwright, J. L., Côté, P., & Bombardier, C. (2013). Assessing bias in studies of prognostic factors. *Ann Intern Med*, *158*(4), 280-286. https://doi.org/10.7326/0003-4819-158-4-201302190-00009
- Heitz, F., Kommoss, S., Tourani, R., Grandelis, A., Uppendahl, L., Aliferis, C., . . . du Bois, A. (2020). Dilution of Molecular-Pathologic Gene Signatures by Medically Associated Factors Might Prevent Prediction of Resection Status After Debulking Surgery in Patients With Advanced Ovarian Cancer. *Clin Cancer Res*, *26*(1), 213-219. https://doi.org/10.1158/1078-0432.ccr-19-1741
- Horowitz, N. S., Larry Maxwell, G., Miller, A., Hamilton, C. A., Rungruang, B., Rodriguez, N., . . . Bookman, M. A. (2018). Predictive modeling for determination of microscopic residual disease at primary cytoreduction: An NRG Oncology/Gynecologic Oncology Group 182 Study. *Gynecologic Oncology*, *148*(1), 49-55. https://doi.org/http://dx.doi.org/10.1016/j.ygyno.2017.10.011
- Horowitz, N. S., Miller, A., Rungruang, B., Richard, S. D., Rodriguez, N., Bookman, M. A., . . . Maxwell, G. L. (2015). Does aggressive surgery improve outcomes? Interaction between preoperative disease burden and complex surgery in patients with advanced-stage ovarian cancer: an analysis of GOG 182. *J Clin Oncol*, *33*(8), 937-943. https://doi.org/10.1200/jco.2014.56.3106
- Horowitz, N. S., Miller, A., Rungruang, B. J., Krivak, T. C., Richard, S. D., Hamilton, C. A., . . . Maxwell, G. L. (2014). Predictive model for preoperative determination of microscopic residual disease at the time of primary cytoreduction in patients with advanced-stage epithelial ovarian cancer: A Gynecologic Oncology Group (GOG) 182 analysis [Conference Abstract]. *Gynecologic Oncology*, *133*, 24. https://doi.org/http://dx.doi.org/10.1016/j.ygyno.2014.03.080
- Huang, K. L., Mashl, R. J., Wu, Y., Ritter, D. I., Wang, J., Oh, C., . . . Ding, L. (2018). Pathogenic Germline Variants in 10,389 Adult Cancers. *Cell*, *173*(2), 355-370.e314. https://doi.org/10.1016/j.cell.2018.03.039
- Huang, Z., Gao, Y., Wen, W., Li, H., Zheng, W., Shu, X. O., & Beeghly-Fadiel, A. (2015). Contraceptive methods and ovarian cancer risk among Chinese women: A report from the Shanghai Women's Health Study. *Int J Cancer*, *137*(3), 607-614. https://doi.org/10.1002/ijc.29412
- Hudson, T. J., Anderson, W., Artez, A., Barker, A. D., Bell, C., Bernabé, R. R., . . . Wainwright, B. J. (2010). International network of cancer genome projects. *Nature*, *464*(7291), 993-998. https://doi.org/10.1038/nature08987
- Huhtinen, K., Suvitie, P., Hiissa, J., Junnila, J., Huvila, J., Kujari, H., . . . Perheentupa, A. (2009). Serum HE4 concentration differentiates malignant ovarian tumours from ovarian endometriotic cysts. *Br J Cancer*, *100*(8), 1315-1319. https://doi.org/10.1038/sj.bjc.6605011
- Hunt, C. R., Gupta, A., Horikoshi, N., & Pandita, T. K. (2012). Does PTEN loss impair DNA double-strand break repair by homologous recombination? *Clin Cancer Res*, *18*(4), 920-922. https://doi.org/10.1158/1078-0432.ccr-11-3131

- Hunter, R. W., Alexander, N. D., & Soutter, W. P. (1992). Meta-analysis of surgery in advanced ovarian carcinoma: is maximum cytoreductive surgery an independent determinant of prognosis? *Am J Obstet Gynecol*, *166*(2), 504-511.
- Integrated genomic analyses of ovarian carcinoma. (2011). *Nature*, *474*(7353), 609-615. https://doi.org/10.1038/nature10166
- Jacobs, I., Oram, D., Fairbanks, J., Turner, J., Frost, C., & Grudzinskas, J. G. (1990). A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer. *Br J Obstet Gynaecol*, *97*(10), 922-929. https://doi.org/10.1111/j.1471-0528.1990.tb02448.x
- Jacobs, I. J., Menon, U., Ryan, A., Gentry-Maharaj, A., Burnell, M., Kalsi, J. K., . . . Skates, S. J. (2016). Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. *Lancet*, *387*(10022), 945-956. https://doi.org/10.1016/s0140-6736(15)01224-6
- Janco, J. M., Glaser, G., Kim, B., McGree, M. E., Weaver, A. L., Cliby, W. A., . . . Bakkum-Gamez, J. N. (2015a). Development of a prediction model for residual disease in newly diagnosed advanced ovarian cancer. *Gynecol Oncol*, *138*(1), 70-77. https://doi.org/10.1016/j.ygyno.2015.04.013
- Janssen-Heijnen, M. L., Houterman, S., Lemmens, V. E., Louwman, M. W., & Coebergh, J. W. (2005). Age and co-morbidity in cancer patients: a population-based approach. *Cancer Treat Res*, *124*, 89-107. https://doi.org/10.1007/0-387-23962-6_5
- Jasen, P. (2009). From the "silent killer" to the "whispering disease": ovarian cancer and the uses of metaphor. *Med Hist*, *53*(4), 489-512. https://doi.org/10.1017/s0025727300000521
- Jones, N. L., Chen, L., Chatterjee, S., Tergas, A. I., Burke, W. M., Hou, J. Y., . . . Wright, J. D. (2018). National Trends in Extended Procedures for Ovarian Cancer Debulking Surgery. *Int J Gynecol Cancer*, *28*(1), 19-25. https://doi.org/10.1097/igc.0000000000001132
- Jung, D. C., Kang, S., Kim, S. C., Kim, J. W., Nam, J. H., Ryu, S. Y., . . . Kim, B. G. (2013). Use of complex surgical procedures, patterns of tumor spread, and CA-125 predicts a risk of incomplete cytoreduction: A Korean Gynecologic Oncology Group study (KGOG-3022). *Gynecologic Oncology*, *131*(2), 336-340. https://doi.org/http://dx.doi.org/10.1016/j.ygyno.2013.07.110
- Kallogjeri, D., Gaynor, S. M., Piccirillo, M. L., Jean, R. A., Spitznagel, E. L., Jr., & Piccirillo, J. F. (2014). Comparison of comorbidity collection methods. *J Am Coll Surg*, *219*(2), 245-255. https://doi.org/10.1016/j.jamcollsurg.2014.01.059
- Kang, S., Kim, T. J., Nam, B. H., Seo, S. S., Kim, B. G., Bae, D. S., & Park, S. Y. (2010). Preoperative serum CA-125 levels and risk of suboptimal cytoreduction in ovarian cancer: a meta-analysis. *J Surg Oncol*, *101*(1), 13-17. https://doi.org/10.1002/jso.21398

- Karlan, B. Y., Dering, J., Walsh, C., Orsulic, S., Lester, J., Anderson, L. A., . . . Slamon, D. (2014). POSTN/TGFBI-associated stromal signature predicts poor prognosis in serous epithelial ovarian cancer. *Gynecol Oncol*, *132*(2), 334-342. https://doi.org/10.1016/j.ygyno.2013.12.021
- Karlsen, M. A., Fago-Olsen, C., Hogdall, E., Schnack, T. H., Christensen, I. J., Nedergaard, L., . . . Hogdall, C. (2016). A novel index for preoperative, non-invasive prediction of macro-radical primary surgery in patients with stage IIIC-IV ovarian cancer-a part of the Danish prospective pelvic mass study. *Tumor Biology*, *37*(9), 12619-12626. https://doi.org/http://dx.doi.org/10.1007/s13277-016-5166-z
- Kehoe, S., Hook, J., Nankivell, M., Jayson, G. C., Kitchener, H., Lopes, T., . . . Swart, A. M. (2015). Primary chemotherapy versus primary surgery for newly diagnosed advanced ovarian cancer (CHORUS): an open-label, randomised, controlled, non-inferiority trial. *Lancet*, *386*(9990), 249-257. https://doi.org/10.1016/s0140-6736(14)62223-6
- Khoja, L., Nolan, K., Mekki, R., Milani, A., Mescallado, N., Ashcroft, L., . . . Jayson, G. C. (2016). Improved Survival from Ovarian Cancer in Patients Treated in Phase III Trial Active Cancer Centres in the UK. *Clin Oncol (R Coll Radiol)*, *28*(12), 760-765. https://doi.org/10.1016/j.clon.2016.06.011
- Kirwan, J. M., Tincello, D. G., Herod, J. J., Frost, O., & Kingston, R. E. (2002). Effect of delays in primary care referral on survival of women with epithelial ovarian cancer: retrospective audit. *Bmj*, *324*(7330), 148-151. https://doi.org/10.1136/bmj.324.7330.148
- Koirala, P., Moon, A. S., & Chuang, L. (2020). Clinical Utility of Preoperative Assessment in Ovarian Cancer Cytoreduction. *Diagnostics (Basel)*, *10*(8). https://doi.org/10.3390/diagnostics10080568
- Konstantinopoulos, P. A., Norquist, B., Lacchetti, C., Armstrong, D., Grisham, R. N., Goodfellow, P. J., . . . Annunziata, C. M. (2020). Germline and Somatic Tumor Testing in Epithelial Ovarian Cancer: ASCO Guideline. *J Clin Oncol*, *38*(11), 1222-1245. https://doi.org/10.1200/jco.19.02960
- Kuhn, E., Kurman, R. J., & Shih, I. M. (2012). Ovarian Cancer Is an Imported Disease: Fact or Fiction? *Curr Obstet Gynecol Rep*, *1*(1), 1-9. https://doi.org/10.1007/s13669-011-0004-1
- Kujawa, K. A., Zembala-Nożyńska, E., Cortez, A. J., Kujawa, T., Kupryjańczyk, J., & Lisowska, K. M. (2020). Fibronectin and Periostin as Prognostic Markers in Ovarian Cancer. *Cells*, *9*(1), 149. https://doi.org/10.3390/cells9010149
- Kumar, A., Sheedy, S., Kim, B., Suidan, R., Sarasohn, D. M., Nikolovski, I., . . . Cliby, W. A. (2019). Models to predict outcomes after primary debulking surgery: Independent validation of models to predict suboptimal cytoreduction and gross residual disease [Research Support, N.I.H., Extramural]. *Gynecologic Oncology*, *154*(1), 72-76. https://doi.org/https://dx.doi.org/10.1016/j.ygyno.2019.04.011

- Kurman, R. J., & Shih Ie, M. (2010). The origin and pathogenesis of epithelial ovarian cancer: a proposed unifying theory. *Am J Surg Pathol*, *34*(3), 433-443. https://doi.org/10.1097/PAS.0b013e3181cf3d79
- Kurman, R. J., & Shih Ie, M. (2011). Molecular pathogenesis and extraovarian origin of epithelial ovarian cancer--shifting the paradigm. *Hum Pathol*, *42*(7), 918-931. https://doi.org/10.1016/j.humpath.2011.03.003
- Kurman, R. J., & Shih Ie, M. (2016). The Dualistic Model of Ovarian Carcinogenesis: Revisited, Revised, and Expanded. *Am J Pathol*, *186*(4), 733-747. https://doi.org/10.1016/j.ajpath.2015.11.011
- Ledermann, J. A., Raja, F. A., Fotopoulou, C., Gonzalez-Martin, A., Colombo, N., & Sessa, C. (2013). Newly diagnosed and relapsed epithelial ovarian carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*, *24 Suppl 6*, vi24-32. https://doi.org/10.1093/annonc/mdt333
- Leibman, A. J., Kruse, B., & McSweeney, M. B. (1988). Transvaginal sonography: comparison with transabdominal sonography in the diagnosis of pelvic masses. *AJR Am J Roentgenol*, *151*(1), 89-92. https://doi.org/10.2214/ajr.151.1.89
- Lengyel, E. (2010). Ovarian cancer development and metastasis. *Am J Pathol*, *177*(3), 1053-1064. https://doi.org/10.2353/ajpath.2010.100105
- Lheureux, S., Braunstein, M., & Oza, A. M. (2019). Epithelial ovarian cancer: Evolution of management in the era of precision medicine. *CA Cancer J Clin*, *69*(4), 280-304. https://doi.org/10.3322/caac.21559
- Li, X., Zhao, L., & Meng, T. (2020). Upregulated CXCL14 is associated with poor survival outcomes and promotes ovarian cancer cells proliferation. *Cell Biochem Funct*. https://doi.org/10.1002/cbf.3516
- Liberati, A., Altman, D., Tetzlaff, J., Mulrow, C., Gøtzsche, P., Ioannidis, J., . . . Moher, D. (2009). The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *Journal of clinical epidemiology*, *62*, e1-34. https://doi.org/10.1016/j.jclinepi.2009.06.006
- Lin, H. W., Tu, Y. Y., Lin, S. Y., Su, W. J., Lin, W. L., Lin, W. Z., . . . Lai, Y. L. (2011). Risk of ovarian cancer in women with pelvic inflammatory disease: a population-based study. *Lancet Oncol*, *12*(9), 900-904. https://doi.org/10.1016/s1470-2045(11)70165-6
- Lisio, M.-A., Fu, L., Goyeneche, A., Gao, Z.-H., & Telleria, C. (2019). High-Grade Serous Ovarian Cancer: Basic Sciences, Clinical and Therapeutic Standpoints. *International journal of molecular sciences*, *20*(4), 952. https://doi.org/10.3390/ijms20040952
- Llueca, A., Serra, A., Rivadulla, I., Gomez, L., Escrig, J., & group, M. w. (2018). Prediction of suboptimal cytoreductive surgery in patients with advanced ovarian cancer based on preoperative and intraoperative determination of the

peritoneal carcinomatosis index. *World Journal of Surgical Oncology*, *16*(1), 37. https://doi.org/https://dx.doi.org/10.1186/s12957-018-1339-0

- Lord, C. J., & Ashworth, A. (2012). The DNA damage response and cancer therapy. *Nature*, *481*(7381), 287-294. https://doi.org/10.1038/nature10760
- Lu, J., Chatterjee, M., Schmid, H., Beck, S., & Gawaz, M. (2016). CXCL14 as an emerging immune and inflammatory modulator. *J Inflamm (Lond)*, *13*, 1. https://doi.org/10.1186/s12950-015-0109-9
- Macintyre, G., Goranova, T. E., De Silva, D., Ennis, D., Piskorz, A. M., Eldridge, M., . . . Brenton, J. D. (2018). Copy number signatures and mutational processes in ovarian carcinoma. *Nat Genet*, *50*(9), 1262-1270. https://doi.org/10.1038/s41588-018-0179-8
- MacKintosh, M. L., Rahim, R., Rajashanker, B., Swindell, R., Kirmani, B. H., Hunt, J., . . . Clayton, R. D. (2014). CT scan does not predict optimal debulking in stage III-IV epithelial ovarian cancer: a multicentre validation study. *J Obstet Gynaecol*, *34*(5), 424-428. https://doi.org/10.3109/01443615.2014.899330
- Macnab, W. a. G. R. a. S. S. a. S. A. H. a. B. K. A. a. M. I. a. R. N. S. a. S. N. a. (2018). Outcomes of women not receiving treatment for advanced ovarian cancer in the West of Scotland: A retrospective analysis. *39*, 58-62.
- Mahmood, R. D., Morgan, R. D., Edmondson, R. J., Clamp, A. R., & Jayson, G. C. (2020). First-Line Management of Advanced High-Grade Serous Ovarian Cancer. *Curr Oncol Rep*, *22*(6), 64. https://doi.org/10.1007/s11912-020-00933-8
- Manvelyan, V., Khemarangsan, V., Huang, K.-G., Adlan, A.-S., & Lee, C.-L. (2016). Port-site metastasis in laparoscopic gynecological oncology surgery: An overview. *Gynecology and Minimally Invasive Therapy*, *5*(1), 1-6. https://doi.org/https://doi.org/10.1016/j.gmit.2015.06.009
- Markman, M. (2003). Optimizing primary chemotherapy in ovarian cancer. *Hematol Oncol Clin North Am*, *17*(4), 957-968, viii. https://doi.org/10.1016/s0889-8588(03)00058-3
- Markman, M. (2008). Antineoplastic agents in the management of ovarian cancer: current status and emerging therapeutic strategies. *Trends Pharmacol Sci*, *29*(10), 515-519. https://doi.org/10.1016/j.tips.2008.07.007
- Matondo, A., Jo, Y. H., Shahid, M., Choi, T. G., Nguyen, M. N., Nguyen, N. N. Y., . . . Kim, S. S. (2017). The Prognostic 97 Chemoresponse Gene Signature in Ovarian Cancer. *Scientific reports*, *7*(1), 9689-9689.
- McGuire, W. P., Hoskins, W. J., Brady, M. F., Kucera, P. R., Partridge, E. E., Look, K. Y., . . . Davidson, M. (1996). Cyclophosphamide and cisplatin compared with paclitaxel and cisplatin in patients with stage III and stage IV ovarian cancer. *N Engl J Med*, *334*(1), 1-6. https://doi.org/10.1056/nejm199601043340101
- Miller, D. S., Blessing, J. A., Krasner, C. N., Mannel, R. S., Hanjani, P., Pearl, M. L., . . . Boardman, C. H. (2009). Phase II evaluation of pemetrexed in the treatment of recurrent or persistent platinum-resistant ovarian or primary

peritoneal carcinoma: a study of the Gynecologic Oncology Group. *J Clin Oncol*, *27*(16), 2686-2691. https://doi.org/10.1200/jco.2008.19.2963

- Ministry of housing, c. a. l. G. (2015). *English Indices of deprivation 2015*. GOV.UK
- Montagnana, M., Lippi, G., Danese, E., Franchi, M., & Guidi, G. C. (2009). Usefulness of serum HE4 in endometriotic cysts. In *Br J Cancer* (Vol. 101, pp. 548). https://doi.org/10.1038/sj.bjc.6605119
- Moons, K. G., de Groot, J. A., Bouwmeester, W., Vergouwe, Y., Mallett, S., Altman, D. G., . . . Collins, G. S. (2014). Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med*, *11*(10), e1001744. https://doi.org/10.1371/journal.pmed.1001744
- Moons, K. G., Kengne, A. P., Grobbee, D. E., Royston, P., Vergouwe, Y., Altman, D. G., & Woodward, M. (2012). Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*, *98*(9), 691-698. https://doi.org/10.1136/heartjnl-2011-301247
- Moons, K. G., Kengne, A. P., Woodward, M., Royston, P., Vergouwe, Y., Altman, D. G., & Grobbee, D. E. (2012). Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart*, *98*(9), 683-690. https://doi.org/10.1136/heartjnl-2011-301246
- Moons, K. G. M., Wolff, R. F., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., . . . Mallett, S. (2019). PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann Intern Med*, *170*(1), W1-w33. https://doi.org/10.7326/m18-1377
- Moore, K., Colombo, N., Scambia, G., Kim, B. G., Oaknin, A., Friedlander, M., . . . DiSilvestro, P. (2018). Maintenance Olaparib in Patients with Newly Diagnosed Advanced Ovarian Cancer. *N Engl J Med*, *379*(26), 2495-2505. https://doi.org/10.1056/NEJMoa1810858
- Moore, R. G., Brown, A. K., Miller, M. C., Skates, S., Allard, W. J., Verch, T., . . . Bast, R. C., Jr. (2008). The use of multiple novel tumor biomarkers for the detection of ovarian carcinoma in patients with a pelvic mass. *Gynecol Oncol*, *108*(2), 402-408. https://doi.org/10.1016/j.ygyno.2007.10.017
- Moore, R. G., McMeekin, D. S., Brown, A. K., DiSilvestro, P., Miller, M. C., Allard, W. J., . . . Skates, S. J. (2009). A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass. *Gynecol Oncol*, *112*(1), 40-46. https://doi.org/10.1016/j.ygyno.2008.08.031
- Mukhopadhyay, A., Elattar, A., Cerbinskaite, A., Wilkinson, S. J., Drew, Y., Kyle, S., . . . Curtin, N. J. (2010). Development of a functional assay for homologous recombination status in primary cultures of epithelial ovarian tumor and correlation with sensitivity to poly(ADP-ribose) polymerase inhibitors. *Clin Cancer Res*, *16*(8), 2344-2351. https://doi.org/10.1158/1078-0432.ccr-09-2758

- Mukhopadhyay, A., Plummer, E. R., Elattar, A., Soohoo, S., Uzir, B., Quinn, J. E., . . . Edmondson, R. J. (2012). Clinicopathological features of homologous recombination-deficient epithelial ovarian cancers: sensitivity to PARP inhibitors, platinum, and survival. *Cancer Res*, *72*(22), 5675-5682. https://doi.org/10.1158/0008-5472.can-12-0324
- Munnell, E. W. (1968). The changing prognosis and treatment in cancer of the ovary. A report of 235 patients with primary ovarian carcinoma 1952-1961. *Am J Obstet Gynecol*, *100*(6), 790-805.
- Myers, P. D., Ng, K., Severson, K., Kartoun, U., Dai, W., Huang, W., . . . Stultz, C. M. (2020). Identifying unreliable predictions in clinical risk models. *npj Digital Medicine*, *3*(1), 8. https://doi.org/10.1038/s41746-019-0209-7
- Narod, S. (2016). Can advanced-stage ovarian cancer be cured? *Nat Rev Clin Oncol*, *13*(4), 255-261. https://doi.org/10.1038/nrclinonc.2015.224
- National Collaborating Centre for, C. (2011). National Institute for Health and Clinical Excellence: Guidance. In *Ovarian Cancer: The Recognition and Initial Management of Ovarian Cancer*. National Collaborating Centre for Cancer (UK)
- Copyright © 2011, National Collaborating Centre for Cancer.
- NCIN. (cited 2018). *Short term ovarian cancer mortality*.
- Newmark, J. J. (2016). 'No ordinary meeting': Robert McWhirter and the decline of radical mastectomy. *J R Coll Physicians Edinb*, *46*(1), 43-48. https://doi.org/10.4997/jrcpe.2016.110
- Nolen, B., Velikokhatnaya, L., Marrangoni, A., De Geest, K., Lomakin, A., Bast, R. C., Jr., & Lokshin, A. (2010). Serum biomarker panels for the discrimination of benign from malignant cases in patients with an adnexal mass. *Gynecol Oncol*, *117*(3), 440-445. https://doi.org/10.1016/j.ygyno.2010.02.005
- O'Malley, C. D., Shema, S. J., Cress, R. D., Bauer, K., Kahn, A. R., Schymura, M. J., . . . Stewart, S. L. (2012). The implications of age and comorbidity on survival following epithelial ovarian cancer: summary and results from a Centers for Disease Control and Prevention study. *J Womens Health (Larchmt)*, *21*(9), 887-894. https://doi.org/10.1089/jwh.2012.3781
- Obermair, A., Petru, E., Windbichler, G., Peters-Engl, C., Graf, A. H., Stummvoll, W., . . . Sevelda, P. (2000). Significance of pretreatment serum hemoglobin and survival in epithelial ovarian cancer. *Oncol Rep*, *7*(3), 639-644. https://doi.org/10.3892/or.7.3.639
- Oken, M. M., Creech, R. H., Tormey, D. C., Horton, J., Davis, T. E., McFadden, E. T., & Carbone, P. P. (1982). Toxicity and response criteria of the Eastern Cooperative Oncology Group. *Am J Clin Oncol*, *5*(6), 649-655.
- Paclitaxel plus carboplatin versus standard chemotherapy with either single-agent carboplatin or cyclophosphamide, doxorubicin, and cisplatin in women with ovarian cancer: the ICON3 randomised trial. (2002). *Lancet*, *360*(9332), 505-515. https://doi.org/10.1016/s0140-6736(02)09738-6

- Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol*, *49*(12), 1373-1379. https://doi.org/10.1016/s0895-4356(96)00236-3
- Pellegrino, B., Mateo, J., Serra, V., & Balmaña, J. (2019). Controversies in oncology: are genomic tests quantifying homologous recombination repair deficiency (HRD) useful for treatment decision making? In *ESMO Open* (Vol. 4, pp. e000480). https://doi.org/10.1136/esmoopen-2018-000480
- Pennington, K. P., Walsh, T., Harrell, M. I., Lee, M. K., Pennil, C. C., Rendi, M. H., . . . Swisher, E. M. (2014). Germline and somatic mutations in homologous recombination genes predict platinum response and survival in ovarian, fallopian tube, and peritoneal carcinomas. *Clin Cancer Res*, *20*(3), 764-775. https://doi.org/10.1158/1078-0432.ccr-13-2287
- Pergialiotis, V., Thomakos, N., Frountzas, M., Haidopoulos, D., Loutradis, D., & Rodolakis, A. (2020). Perioperative blood transfusion and ovarian cancer survival rates: A meta-analysis based on univariate, multivariate and propensity score matched data. *European journal of obstetrics, gynecology, and reproductive biology*, *252*, 137-143. https://doi.org/10.1016/j.ejogrb.2020.06.013
- Petrillo, M., Vizzielli, G., Fanfani, F., Gallotta, V., Cosentino, F., Chiantera, V., . . . Fagotti, A. (2015). Definition of a dynamic laparoscopic model for the prediction of incomplete cytoreduction in advanced epithelial ovarian cancer: proof of a concept. *Gynecologic Oncology*, *139*(1), 5-9. https://doi.org/https://dx.doi.org/10.1016/j.ygyno.2015.07.095
- Phillips, A., Kehoe, S., Singh, K., Elattar, A., Nevin, J., Balega, J., . . . Sundar, S. (2019). Socioeconomic differences impact overall survival in advanced ovarian cancer (AOC) prior to achievement of standard therapy. *Arch Gynecol Obstet*, *300*(5), 1261-1270. https://doi.org/10.1007/s00404-019-05269-8
- Piek, J. M., van Diest, P. J., Zweemer, R. P., Jansen, J. W., Poort-Keesom, R. J., Menko, F. H., . . . Verheijen, R. H. (2001). Dysplastic changes in prophylactically removed Fallopian tubes of women predisposed to developing ovarian cancer. *J Pathol*, *195*(4), 451-456. https://doi.org/10.1002/path.1000
- Piver, M. S., Lele, S. B., Marchetti, D. L., Baker, T. R., Tsukada, Y., & Emrich, L. J. (1988). The impact of aggressive debulking surgery and cisplatin-based chemotherapy on progression-free survival in stage III and IV ovarian carcinoma. *J Clin Oncol*, *6*(6), 983-989. https://doi.org/10.1200/jco.1988.6.6.983
- Prat, J. (2015). Staging Classification for Cancer of the Ovary, Fallopian Tube, and Peritoneum: Abridged Republication of Guidelines From the International Federation of Gynecology and Obstetrics (FIGO). *Obstet Gynecol*, *126*(1), 171-174. https://doi.org/10.1097/aog.0000000000000917

- Puls, L. E., Duniho, T., Hunter, J. E., Kryscio, R., Blackhurst, D., & Gallion, H. (1996). The prognostic implication of ascites in advanced-stage ovarian cancer. *Gynecol Oncol*, *61*(1), 109-112. https://doi.org/10.1006/gyno.1996.0106
- Raab, S. S. (2000). The cost-effectiveness of immunohistochemistry. *Arch Pathol Lab Med*, *124*(8), 1185-1191. https://doi.org/10.1043/0003-9985(2000)124<1185:tceoi>2.0.co;2
- Rausei, S., Uccella, S., D'Alessandro, V., Gisone, B., Frattini, F., Lianos, G., . . . Ghezzi, F. (2019). Aggressive surgery for advanced ovarian cancer performed by a multidisciplinary team: A retrospective analysis on a large series of patients. *Surg Open Sci*, *1*(1), 43-47. https://doi.org/10.1016/j.sopen.2019.05.005
- Reade, C. J., McVey, R. M., Tone, A. A., Finlayson, S. J., McAlpine, J. N., Fung-Kee-Fung, M., & Ferguson, S. E. (2014). The fallopian tube as the origin of high grade serous ovarian cancer: review of a paradigm shift. *J Obstet Gynaecol Can*, *36*(2), 133-140. https://doi.org/10.1016/s1701-2163(15)30659-9
- Reid, B. M., Permuth, J. B., & Sellers, T. A. (2017). Epidemiology of ovarian cancer: a review. *Cancer Biol Med*, *14*(1), 9-32. https://doi.org/10.20892/j.issn.2095-3941.2016.0084
- Riester, M., Wei, W., Culhane, A. C., Trippa, L., Michor, F., Huttenhower, C., . . . Birrer, M. (2014). Risk prediction for late-stage ovarian cancer by meta-analysis of 1,525 patient samples [Conference Abstract]. *Cancer Research. Conference: 105th Annual Meeting of the American Association for Cancer Research, AACR*, *74*(19 SUPPL. 1). https://doi.org/http://dx.doi.org/10.1158/1538-7445.AM2014-2355
- Riggs, M. J., Pandalai, P. K., Kim, J., & Dietrich, C. S. (2020). Hyperthermic Intraperitoneal Chemotherapy in Ovarian Cancer. *Diagnostics (Basel)*, *10*(1). https://doi.org/10.3390/diagnostics10010043
- Riopel, M. A., Ronnett, B. M., & Kurman, R. J. (1999). Evaluation of diagnostic criteria and behavior of ovarian intestinal-type mucinous tumors: atypical proliferative (borderline) tumors and intraepithelial, microinvasive, invasive, and metastatic carcinomas. *Am J Surg Pathol*, *23*(6), 617-635. https://doi.org/10.1097/00000478-199906000-00001
- Risum, S., Hogdall, E., Engelholm, S. A., Fung, E., Lomas, L., Yip, C., . . . Hogdall, C. (2009). A proteomics panel for predicting optimal primary cytoreduction in stage III/IV ovarian cancer. *Int J Gynecol Cancer*, *19*(9), 1535-1538. https://doi.org/10.1111/IGC.0b013e3181a840f5
- Robinson, D. H., & Toledo, A. H. (2012). Historical development of modern anesthesia. *J Invest Surg*, *25*(3), 141-149. https://doi.org/10.3109/08941939.2012.690328
- Romanidis, K., Nagorni, E. A., Halkia, E., & Pitiakoudis, M. (2014). The role of cytoreductive surgery in advanced ovarian cancer: the general surgeon's perspective. *J buon*, *19*(3), 598-604.

- Rose, P., Rubin, G., Perera, R., Almberg, S., Barisic, A., Dawes, M., . . . Hamilton, W. (2015). Explaining variation in cancer survival between 11 jurisdictions in the International Cancer Benchmarking Partnership: a primary care vignette survey. *BMJ open*, *5*, e007212. https://doi.org/10.1136/bmjopen-2014-007212
- Rose, P. G., Nerenstone, S., Brady, M. F., Clarke-Pearson, D., Olt, G., Rubin, S. C., . . . Small, J. M. (2004). Secondary surgical cytoreduction for advanced ovarian carcinoma. *N Engl J Med*, *351*(24), 2489-2497. https://doi.org/10.1056/NEJMoa041125
- Rutten, I. J., van de Laar, R., Kruitwagen, R. F., Bakers, F. C., Ploegmakers, M. J., Pappot, T. W., . . . Van Gorp, T. (2016). Prediction of incomplete primary debulking surgery in patients with advanced ovarian cancer: An external validation study of three models using computed tomography. *Gynecol Oncol*, *140*(1), 22-28. https://doi.org/10.1016/j.ygyno.2015.11.022
- Systematic Review]. *International Journal of Gynecological Cancer*, *25*(3), 407-415. https://doi.org/https://dx.doi.org/10.1097/IGC.0000000000000368
- Rutten, M. J., van de Vrie, R., Bruining, A., Spijkerboer, A. M., Mol, B. W., Kenter, G. G., & Buist, M. R. (2015b). 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. *Int J Gynecol Cancer*, *25*(3), 407-415. https://doi.org/10.1097/igc.0000000000000368
- Sayasneh, A., Tsivos, D., & Crawford, R. (2011). Endometriosis and ovarian cancer: a systematic review. *ISRN Obstet Gynecol*, *2011*, 140310. https://doi.org/10.5402/2011/140310
- Schildkraut, J. M., Schwingl, P. J., Bastos, E., Evanoff, A., & Hughes, C. (1996). Epithelial ovarian cancer risk among women with polycystic ovary syndrome. *Obstet Gynecol*, *88*(4 Pt 1), 554-559. https://doi.org/10.1016/0029-7844(96)00226-8
- Schorge, J. O., McCann, C., & Del Carmen, M. G. (2010). Surgical debulking of ovarian cancer: what difference does it make? *Rev Obstet Gynecol*, *3*(3), 111-117.
- Scott, R., Hawarden, A., Russell, B., & Edmondson, R. J. (2020). Decision-Making in Gynaecological Oncology Multidisciplinary Team Meetings: A Cross-Sectional, Observational Study of Ovarian Cancer Cases. *Oncol Res Treat*, *43*(3), 70-77. https://doi.org/10.1159/000504260
- Shah, A. A., Frierson, H. F., Jr., & Cathro, H. P. (2012). Analysis of immunohistochemical stain usage in different pathology practice settings. *Am J Clin Pathol*, *138*(6), 831-836. https://doi.org/10.1309/ajcpagvtckdxkk0x
- Shah, C. A., Lowe, K. A., Paley, P., Wallace, E., Anderson, G. L., McIntosh, M. W., . . . Drescher, C. W. (2009). Influence of ovarian cancer risk status on the diagnostic performance of the serum biomarkers mesothelin, HE4, and CA125.

Cancer Epidemiol Biomarkers Prev, *18*(5), 1365-1372. https://doi.org/10.1158/1055-9965.epi-08-1034

- Son, H. M., Kim, S. H., Kwon, B. R., Kim, M. J., Kim, C. S., & Cho, S. H. (2017). Preoperative prediction of suboptimal resection in advanced ovarian cancer based on clinical and CT parameters. *Acta Radiol*, *58*(4), 498-504. https://doi.org/10.1177/0284185116658683
- Stashwick, C., Post, M. D., Arruda, J. S., Spillman, M. A., Behbakht, K., Davidson, S. A., & Kelly, M. G. (2011). Surgical risk score predicts suboptimal debulking or a major perioperative complication in patients with advanced epithelial ovarian, fallopian tube, or primary peritoneal cancer. *International journal of gynecological cancer : official journal of the International Gynecological Cancer Society*, *21*(8), 1422-1427.
- Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., . . . Higgins, J. P. (2016). ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *Bmj*, *355*, i4919. https://doi.org/10.1136/bmj.i4919
- Su, J., Barbera, L., & Sutradhar, R. (2015). Do repeated assessments of performance status improve predictions for risk of death among patients with cancer? A population-based cohort study. *Palliat Med*, *29*(6), 547-553. https://doi.org/10.1177/0269216314568231
- Suidan, R. S., Ramirez, P. T., Sarasohn, D. M., Teitcher, J. B., Iyer, R. B., Zhou, Q., . . . Chi, D. S. (2017). A multicenter assessment of the ability of preoperative computed tomography scan and CA-125 to predict gross residual disease at primary debulking for advanced epithelial ovarian cancer. *Gynecol Oncol*, *145*(1), 27-31. https://doi.org/10.1016/j.ygyno.2017.02.020
- Suidan, R. S., Ramirez, P. T., Sarasohn, D. M., Teitcher, J. B., Mironov, S., Iyer, R. B., . . . Chi, D. S. (2014b). Surgical risk score predicts suboptimal debulking or a major perioperative complication in patients with advanced epithelial ovarian, fallopian tube, or primary peritoneal cancer. *Gynecol Oncol*, *134*(3), 455-461. https://doi.org/10.1016/j.ygyno.2014.07.002
- Sun, G. W., Shook, T. L., & Kay, G. L. (1996). Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *J Clin Epidemiol*, *49*(8), 907-916. https://doi.org/10.1016/0895-4356(96)00025-x
- Sundar, S. S., Knott, C., Paley, L., Jones, A., Wakefield, C., Platt, M.-C., . . . Nordin, A. (2020). 604 Significant variation in treatment and survival outcomes in stage 2–4 ovarian cancer in england: results from the national ovarian cancer feasibility audit pilot. *International Journal of Gynecologic Cancer*, *30*(Suppl 4), A133. https://doi.org/10.1136/ijgc-2020-ESGO.232
- Tang, M., Liu, B., Bu, X., & Zhao, P. (2018). Cross-talk between ovarian cancer cells and macrophages through periostin promotes macrophage recruitment. *Cancer science*, *109*(5), 1309-1318. https://doi.org/10.1111/cas.13567
- TCGA. (2011). Integrated genomic analyses of ovarian carcinoma. *Nature*, *474*(7353), 609-615. https://doi.org/10.1038/nature10166

- Telli, M. L., Timms, K. M., Reid, J., Hennessy, B., Mills, G. B., Jensen, K. C., . . . Richardson, A. L. (2016). Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer. *Clin Cancer Res*, *22*(15), 3764-3773. https://doi.org/10.1158/1078-0432.ccr-15-2477
- Terada, K. Y., Ahn, H. J., & Kessel, B. (2016). Differences in risk for type 1 and type 2 ovarian cancer in a large cancer screening trial. *J Gynecol Oncol*, *27*(3), e25. https://doi.org/10.3802/jgo.2016.27.e25
- Thomas, Q. D., Boussere, A., Classe, J.-M., Pomel, C., Costaz, H., Rodrigues, M., . . . Fiteni, F. (2022). Optimal timing of interval debulking surgery for advanced epithelial ovarian cancer: A retrospective study from the ESME national cohort. *Gynecologic Oncology*. https://doi.org/https://doi.org/10.1016/j.ygyno.2022.08.005
- Tothill, R. W., Tinker, A. V., George, J., Brown, R., Fox, S. B., Lade, S., . . . Bowtell, D. D. (2008). Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin Cancer Res*, *14*(16), 5198-5208. https://doi.org/10.1158/1078-0432.ccr-08-0196
- Tucker, S. L., Gharpure, K., Herbrich, S. M., Unruh, A. K., Nick, A. M., Crane, E. K., . . . Sood, A. K. (2014). Molecular biomarkers of residual disease after surgical debulking of high-grade serous ovarian cancer. *Clin Cancer Res*, *20*(12), 3280-3288. https://doi.org/10.1158/1078-0432.ccr-14-0445
- Tumiati, M., Hietanen, S., Hynninen, J., Pietilä, E., Färkkilä, A., Kaipio, K., . . . Kauppi, L. (2018). A Functional Homologous Recombination Assay Predicts Primary Chemotherapy Response and Long-Term Survival in Ovarian Cancer Patients. *Clin Cancer Res*, *24*(18), 4482-4493. https://doi.org/10.1158/1078-0432.ccr-17-3770
- Tumiati, M., Hietanen, S., & Kauppi, L. (2018). Time to go functional! Determining tumors' DNA repair capacity ex vivo. *Oncotarget*, *9*(96), 36826-36827.
- Tworoger, S. S., Fairfield, K. M., Colditz, G. A., Rosner, B. A., & Hankinson, S. E. (2007). Association of oral contraceptive use, other contraceptive methods, and infertility with ovarian cancer risk. *Am J Epidemiol*, *166*(8), 894-901. https://doi.org/10.1093/aje/kwm157
- Urban, N., Thorpe, J., Karlan, B. Y., McIntosh, M. W., Palomares, M. R., Daly, M. B., . . . Drescher, C. W. (2012). Interpretation of single and serial measures of HE4 and CA125 in asymptomatic women at high risk for ovarian cancer. *Cancer Epidemiol Biomarkers Prev*, *21*(11), 2087-2094. https://doi.org/10.1158/1055-9965.epi-12-0616
- Urban, R. R., He, H., Alfonso, R., Hardesty, M. M., Gray, H. J., & Goff, B. A. (2016). Ovarian cancer outcomes: Predictors of early death. *Gynecol Oncol*, *140*(3), 474-480. https://doi.org/10.1016/j.ygyno.2015.12.021
- van de Vrie, R., Rutten, M. J., Asseler, J. D., Leeflang, M. M., Kenter, G. G., Mol, B. W. J., & Buist, M. (2019). Laparoscopy for diagnosing resectability of disease

in women with advanced ovarian cancer [Systematic Review]. *Cochrane Database of Systematic Reviews*, *3*, CD009786. https://doi.org/https://dx.doi.org/10.1002/14651858.CD009786.pub3

- van de Vrie, R., van Meurs, H. S., Rutten, M. J., Naaktgeboren, C. A., Opmeer, B. C., Gaarenstroom, K. N., . . . Buist, M. R. (2017). Cost-effectiveness of laparoscopy as diagnostic tool before primary cytoreductive surgery in ovarian cancer. *Gynecol Oncol*, *146*(3), 449-456. https://doi.org/10.1016/j.ygyno.2017.06.019
- van der Burg, M. E., van Lent, M., Buyse, M., Kobierska, A., Colombo, N., Favalli, G., . . . Pecorelli, S. (1995). The effect of debulking surgery after induction chemotherapy on the prognosis in advanced epithelial ovarian cancer. Gynecological Cancer Cooperative Group of the European Organization for Research and Treatment of Cancer. *N Engl J Med*, *332*(10), 629-634. https://doi.org/10.1056/nejm199503093321002
- van Meurs, H. S., Tajik, P., Hof, M. H., Vergote, I., Kenter, G. G., Mol, B. W., . . . Bossuyt, P. M. (2013). Which patients benefit most from primary surgery or neoadjuvant chemotherapy in stage IIIC or IV ovarian cancer? An exploratory analysis of the European Organisation for Research and Treatment of Cancer 55971 randomised trial. *Eur J Cancer*, *49*(15), 3191-3201. https://doi.org/10.1016/j.ejca.2013.06.013
- van Wijk, L. M., Vermeulen, S., Meijers, M., van Diest, M. F., Ter Haar, N. T., de Jonge, M. M., . . . Vreeswijk, M. P. G. (2020). The RECAP Test Rapidly and Reliably Identifies Homologous Recombination-Deficient Ovarian Carcinomas. *Cancers (Basel)*, *12*(10). https://doi.org/10.3390/cancers12102805
- Vanderstichele, A., Busschaert, P., Olbrecht, S., Lambrechts, D., & Vergote, I. (2017). Genomic signatures as predictive biomarkers of homologous recombination deficiency in ovarian cancer. *Eur J Cancer*, *86*, 5-14. https://doi.org/10.1016/j.ejca.2017.08.029
- Vergote, I., Marquette, S., Amant, F., Berteloot, P., & Neven, P. (2005). Port-site metastases after open laparoscopy: a study in 173 patients with advanced ovarian carcinoma. *Int J Gynecol Cancer*, *15*(5), 776-779. https://doi.org/10.1111/j.1525-1438.2005.00135.x
- Vergote, I., Trope, C. G., Amant, F., Kristensen, G. B., Ehlen, T., Johnson, N., . . . Reed, N. S. (2010). Neoadjuvant chemotherapy or primary surgery in stage IIIC or IV ovarian cancer. *N Engl J Med*, *363*(10), 943-953. https://doi.org/10.1056/NEJMoa0908806
- Vásárhelyi, B. a. A. D. L. (2017). Lab Test Findings in the Elderly. *Ejifcc*, *28*, 328-332.
- Wahner Hendrickson, A. E., Hawthorne, K. M., Goode, E. L., Kalli, K. R., Goergen, K. M., Bakkum-Gamez, J. N., . . . Maurer, M. J. (2015). Assessment of published models and prognostic variables in epithelial ovarian cancer at Mayo Clinic [Research Support, N.I.H., Extramural]. *Gynecologic Oncology*, *137*(1), 77-85. https://doi.org/https://dx.doi.org/10.1016/j.ygyno.2015.01.539

- Wang, Y., Liu, P., Xu, Y., Zhang, W., Tong, L., Guo, Z., & Ni, H. (2015). Preoperative neutrophil-to-lymphocyte ratio predicts response to first-line platinum-based chemotherapy and prognosis in serous ovarian cancer. *Cancer Chemother Pharmacol*, *75*(2), 255-262. https://doi.org/10.1007/s00280-014-2622-6
- Warner, L. L., Dowdy, S. C., Martin, J. R., Lemens, M. A., McGree, M. E., Weaver, A. L., . . . Bakkum-Gamez, J. N. (2013). The impact of perioperative packed red blood cell transfusion on survival in epithelial ovarian cancer. *Int J Gynecol Cancer*, *23*(9), 1612-1619. https://doi.org/10.1097/01.IGC.0000436089.03581.6b
- Webb, P. M., & Jordan, S. J. (2017). Epidemiology of epithelial ovarian cancer. *Best Pract Res Clin Obstet Gynaecol*, *41*, 3-14. https://doi.org/10.1016/j.bpobgyn.2016.08.006
- Winter, W. E., 3rd, Maxwell, G. L., Tian, C., Sundborg, M. J., Rose, G. S., Rose, P. G., . . . McGuire, W. P. (2008). Tumor residual after surgical cytoreduction in prediction of clinical outcome in stage IV epithelial ovarian cancer: a Gynecologic Oncology Group Study. *J Clin Oncol*, *26*(1), 83-89. https://doi.org/10.1200/jco.2007.13.1953
- Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., . . . Mallett, S. (2019). PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med*, *170*(1), 51-58. https://doi.org/10.7326/m18-1376
- Worboys, M. (2013). Joseph Lister and the performance of antiseptic surgery. *Notes Rec R Soc Lond*, *67*(3), 199-209. https://doi.org/10.1098/rsnr.2013.0028
- Wright, A. A., Bohlke, K., Armstrong, D. K., Bookman, M. A., Cliby, W. A., Coleman, R. L., . . . Edelson, M. I. (2016). Neoadjuvant chemotherapy for newly diagnosed, advanced ovarian cancer: Society of Gynecologic Oncology and American Society of Clinical Oncology Clinical Practice Guideline. *Gynecol Oncol*, *143*(1), 3-15. https://doi.org/10.1016/j.ygyno.2016.05.022
- Yang, H. S., Yoon, C., Myung, S. K., & Park, S. M. (2011). Effect of obesity on survival of women with epithelial ovarian cancer: a systematic review and meta-analysis of observational studies. *Int J Gynecol Cancer*, *21*(9), 1525-1532. https://doi.org/10.1097/IGC.0b013e31822eb5f8
- Zhou, Q., Hong, L., Zuo, M. Z., & He, Z. (2017). Prognostic significance of neutrophil to lymphocyte ratio in ovarian cancer: evidence from 4,910 patients. *Oncotarget*, *8*(40), 68938-68949. https://doi.org/10.18632/oncotarget.20196
- Zhou, Y., Chlebowski, R., LaMonte, M. J., Bea, J. W., Qi, L., Wallace, R., . . . Irwin, M. L. (2014). Body mass index, physical activity, and mortality in women diagnosed with ovarian cancer: results from the Women's Health Initiative. *Gynecol Oncol*, *133*(1), 4-10. https://doi.org/10.1016/j.ygyno.2014.01.033

10 Appendices

10.1 Appendix A. Publication resulting from this project

PLOS ONE

A three protein signature fails to externally validate as a biomarker to predict surgical outcome in high-grade epithelial ovarian cancer

--Manuscript Draft--

Manuscript Number:	
Article Type:	Research Article
Full Title:	A three protein signature fails to externally validate as a biomarker to predict surgical outcome in high-grade epithelial ovarian cancer
Short Title:	surgical biomarkers in ovarian cancer
Corresponding Author:	Richard Edmondson University of Manchester Manchester, UNITED KINGDOM
Keywords:	ovarian cancer; surgery; POSTN; CXCL14; pSmad2/3
Abstract:	Introduction For patients with advanced epithelial ovarian cancer, complete surgical cytoreduction remains the strongest predictor of outcome. However, identifying patients who are likely to benefit from such surgery remains elusive and to date few surgical outcome prediction tools have been validated. Here we attempted to externally validate a promising three protein signature, which had previously shown strong association with suboptimal surgical debulking (AUC 0.89, accuracy 92.8%), 1 . Methods 238 high-grade epithelial ovarian cancer samples were collected from patients who participated in a large multicentre trial (ICON5). Samples were collected at the time of initial surgery and before randomisation. Surgical outcome data were collated from prospectively collected study records. Immunohistochemical scores were generated by two independent observers for the three proteins in the original signature (POSTN, CXCL14 and pSmad2/3). Predictive values were generated for individual and combination protein signatures. Results When assessed individually, none of the proteins showed any evidence of predictive affinity for suboptimal surgical outcome in our cohort (AUC POSTN 0.55, pSmad 2/3 0.53, CXCL 14 0.62). The combined signature again showed poor predictive ability with an AUC 0.58. Conclusions Despite showing original promise, when this protein signature is applied to a large external cohort, it is unable to accurately predictive surgical outcomes. This could be attributed to overfitting of the original model, or differences in surgical practice between cohorts.
Order of Authors:	Richard Edmondson
	Amy Hawarden
	Marcus Price
	Bryn Russell
	Godfrey Wilson
	Laura Farrelly
	Andrew Embleton-Thirsk
	Mahesh Parmar
Opposed Reviewers:	Michael Birrer

Abstract

Introduction

For patients with advanced epithelial ovarian cancer, complete surgical cytoreduction remains the strongest predictor of outcome. However, identifying patients who are likely to benefit from such surgery remains elusive and to date few surgical outcome prediction tools have been validated. Here we attempted to externally validate a promising three protein signature, which had previously shown strong association with suboptimal surgical debulking (AUC 0.89, accuracy 92.8%), [1].

Methods

238 high-grade epithelial ovarian cancer samples were collected from patients who participated in a large multicentre trial (ICON5). Samples were collected at the time of initial surgery and before randomisation. Surgical outcome data were collated from prospectively collected study records. Immunohistochemical scores were generated by two independent observers for the three proteins in the original signature (POSTN, CXCL14 and pSmad2/3). Predictive values were generated for individual and combination protein signatures.

Results

When assessed individually, none of the proteins showed any evidence of predictive affinity for suboptimal surgical outcome in our cohort (AUC POSTN 0.55, pSmad 2/3 0.53, CXCL 14 0.62). The combined signature again showed poor predictive ability with an AUC 0.58.

Conclusions

Despite showing original promise, when this protein signature is applied to a large external cohort, it is unable to accurately predictive surgical outcomes. This could be attributed to overfitting of the original model, or differences in surgical practice between cohorts.

Introduction

Epithelial ovarian cancer (EOC) accounts for an estimated 239,000 new cases and 152,000 deaths worldwide annually [2]. Survival outcomes remain poor with the five-year survival for all stages being just 35%[3]. The majority of EOCs present with advanced disease, reflecting disease spread outside of the pelvis[4].

Treatment for EOC combines surgical resection of disease, platinum-based chemotherapy and more recently individualised maintenance therapies including PARP inhibitors[5]. Complete cytoreduction (no visible remaining disease following surgery) is the over-riding goal of surgical treatment, with overall survival (OS) being compromised if this is not achieved[6, 7]. Surgery can occur first-line, as primary debulking surgery (PDS) or if disease is deemed unresectable at the time of diagnosis as a second line treatment following chemotherapy, termed interval debulking surgery (IDS) [8-10]. Complete debulking at the time of PDS may hold a slight survival advantage over complete debulking at the time of IDS [11, 12]. However, tools to predict outcome remain elusive.

Many surgical prediction models have been published in the literature. These models utilise a combination of many different data modalities, including patient demographics, biochemical factors, radiological factors, genomic factors, and diagnostic laparoscopy. To date, only laparoscopy has externally validated with enough success to be considered for clinical use[13] but has not been widely adopted, in part because it is an invasive surgical procedure with an associated morbidity. Failure to develop and validate biomarkers of surgical outcome has been identified as a major deficit to the management of patients with advanced ovarian cancer[14].

An ideal biomarker would be simple, non-invasive, carry no additional morbidity for the patient, and would have a high degree of accuracy. Such a candidate biomarker was proposed by Riester *et al*, in 2014[15]. The proposed model used expression of three proteins in stage III and IV high grade epithelial ovarian cancer tumours to predict suboptimal surgical debulking status with a sensitivity of 92.8% and an area under the Receiver Operating Characteristic curve (AUC) of 0.89[15].

The model utilises immunohistochemistry, a method that is used in all specialist histopathology labs[16] and could be applied using image guided biopsies which are now the standard diagnostic material for advanced ovarian cancer. Immunohistochemistry is therefore readily available, fast and cost effective[17].

Here we describe the external validation of the three-protein signature, using an independent cohort of patients recruited in the ICON 5 trial[18].

Methods

Samples were accrued prospectively from patients enrolled in the MRC ICON5 clinical trial[18]. Patients with stage III or IV epithelial ovarian cancer were enrolled into ICON5 following primary cytoreductive surgery between 2002 and 2004. Patients were all WHO performance status 0-2 and had sufficient bone marrow, kidney and neurological function to be considered for chemotherapy. Following enrolment into ICON 5, patients were randomised to one of five chemotherapy arms but critically for the current study randomisation took place following surgery and tissue collection. Surgical outcome was therefore independent of allocated treatment arm. All clinical data were recorded prospectively as part of the clinical trial protocol but were kept blinded to the laboratory team until scoring and analysis had been completed. Included tumours had ≥ 3 available slides to allow for staining for each of the three proteins and for negative controls

All patients consented to donation and use of tissue samples at the time of enrolment into the ICON 5 study which had appropriate regulatory and ethical approval (London REC; MREC/02/2/3)

Immunohistochemistry

Each tumour sample underwent immunohistochemical staining for POSTN (Anti- POSTN 1.25μg/mL Oxford biosystems (RD18104050)), CXCL 14 (Anti- CXCL 14 2.5μg/mL Abcam, Cambridge, UK ab46010) and pSmad 2/3 (Anti- phosphor- Smad2, cell signalling Tech (3108S)) via the Bond-III automated IHC stainer following antibody concentration optimisation via hand staining. Deparaffinised sections were subjected to antigen retrieval (citrate buffer, pH = 6, in microwave for 2x5 mins), incubated with each primary antibody overnight at 4°C, visualised with a three-layer avidin-biotin technique and 3,3' - diaminobenzidine, and counterstained with Mayer's hematoxylin. Slides were scanned and images captured using a Leica SCN 400.

Scoring

For the purposes of validation, methods and statistical analysis described by Reister *et al* were replicated exactly, taking details from the original paper and contacting the authors for clarification where necessary[15]. Each slide was scored in three separate 1mm^2 pre-determined areas by two independent scorers (AH and MP), using QuPath – 0.2.0 – m8 software (Queens University Belfast, N.I.) at 20 times magnification. Each region was given a score based on the difference of staining intensity between tumour and stroma of 1, 2, or 3 (mild, moderate and strong respectively), multiplied by the percentage of tumour cells within that region displaying this staining intensity, represented as a score of 0 to 4 (<5%, 5-25%, 26-75%, and >75% respectively) Both scorers were blinded to clinical outcomes. Discrepancies in scoring were resolved by a third party (GW).

All slide scoring data was collated in MS Excel. Inter-scorer variability was determined using Spearman's Rank Coefficient, and differences in cohorts when data is not paired was determined using Mann-Whitney-U test, both performed using Graphpad Prism version 8.4.3 (471). Validation of both individual proteins predictive value, and combined score value was firstly calculated via simple logistic regression and subsequent creation of receiver operator characteristic curves (ROC curves) in Graphpad Prism version 8.4.3 (GraphPad Software, San Diego, CA). The code originally used by Riester *et al* in R was available from supplementary materials and was re-run on the validation dataset to ensure consistency. The multivariable prediction model was created using logistic regression in WEKA, an open source machine learning software[19]. A p value of <0.05 was used to determine significance for all statistics.

Results

238 patient samples were identified from the ICON5 clinical trial. All samples comprised a block or >= 3 formalin fixed paraffin embedded slides, Figure 1.

Antibody concentrations were determined by hand staining and confirmed using the automated staining platform, Figure 2. Antibodies for POSTN and CXCL 14 produced adequate staining at the same concentrations used in the original study (1:800, 1:400 respectively), however Anti- pSmad 2/3 was required at a more concentrated dilution (1:50) in order to achieve adequate staining levels.

Of the 238 patient samples included for analysis, all were from a high-grade epithelial subtype and originated from FIGO stage III and IV tumours. 202/238 (85%) patients underwent primary debulking surgery, with a suboptimal rate of 46%, a higher percentage than reported for the whole trial cohort (30%). Thus the number of events (suboptimal cytoreduction) was 109 for this validation cohort.

Slides were scored as per the methods section above. The scoring for each protein was assessed, and all three showed a strong positive association between the two scorers ($p > 0.001$), table 1. When a sum of the scores was calculated for each of the three proteins, again a strong positive correlation was demonstrated between the two scorers ($R_{2=}$ 0.8025, 95% CI 0.7506 – 0.8445, $p < 0.0001$)

Logistic regression was utilised to externally validate four models (three individual protein models and one combined score model) in the total cohort of 238 cases, firstly taking each protein in turn and finally combining the individual proteins scores to create a combined score. All four models were associated with poor performance to predict suboptimal cytoreduction (POSTN AUC 0.55 p=0.174, pSmad2/3 AUC 0.53 p=0.437, CXCL 14 AUC 0.62 p=0.0012, combined score AUC 0.59 p=0.0131).

As the original model contained only primary tumours, and our validation cohort contained 85% primary tumours (n=202) and 15% tumours taken at the time of interval debulking surgery (n=36), the analysis was repeated limiting to PDS samples only. Excluding the 36 IDS tumours from the analysis, there was marginal improvement in predictive ability for the four models (POSTN AUC 0.56 pSmad 2/3 AUC 0.54, CXCL 14 AUC 0.63, combined scores AUC 0.58).

A comparison of these results, alongside results from the original study are shown in table 2.

Discussion

Despite showing early promise on an internal validation cohort, the predictive affinity of this three protein signature is not replicated when applied to an external cohort of patients. This failure of validation may be attributed to overfitting of the original model and/or differing surgical practice between centres.

The accuracy levels achieved (AUC $\leq$ 0.621) for all models in this validation are scarcely more than chance, and therefore would not be acceptable for use in clinical practice.

The external validation was performed with care to ensure the techniques used to create the original model were replicated as closely as possible in the validation set. The conditions in which the IHC were undertaken were as similar as possible with the exception that the original model used hand staining, in contrast to this validation study which used automated staining. Despite this difference, hand staining was also successfully performed for the optimisation of antibodies in the validation model and all other materials and methods were kept consistent. Although the concentrations of the antibody anti-pSmad 2/3 did differ between the two studies, the working concentrations of anti-pSmad 2/3 used in the original study were not known. Differences may be accounted for by batch inconsistency.

There was very strong positive association between the two scorers in the validation cohort for all protein stains, which gives confidence in the consistency of the scoring.

Both studies used historical slides that had been stored between 4 – 20 years before IHC was undertaken. The validation cohort were stored at room temperature in a pre-cut paraffin fixed state. The method of storing used in the original study is not known. There are very few studies exploring the relationship between the time fixed slides are stored and the accuracy of IHC results. Some studies have suggested that longer storage time may be detrimental to antigenicity in tumour samples, resulting in false negative findings[20]. Conversely, other studies have contradicted this thinking, with Forse et al reporting

adequate staining of breast cancer tissue via IHC following 12 years of storage[21], although these slides were stored at -80°C and not at room temperature. Consensus does agree however, that if slides are to be stored over prolonged periods, they must be paraffin fixed, as they were in this validation study. Both studies also included successful negative control, suggesting that a positive result was indeed a true positive. Staining was also reviewed by an experienced consultant histopathologist (GW), who confirmed that despite their age, the slides have stained adequately.

Many previously published surgical prediction models have also failed to successfully validate when applied to external cohorts, and this is often attributed to the differences between the cohorts, with this study being no exception. Most notably, the original study cohort underwent surgery in a single institution, whereas the validation cohort were made up of patients from multiple different centres internationally. Variation in surgical practice within centres is well established, and surgeon heterogeneity between centres is vast[22-24]. This variation in practice may explain some of the differences seen between the two cohorts. However, the ICON5 study recruited patients from many centres in the UK and thus more likely represents clinical practice.

This external validation was conducted with adequate power and replicated the methods and materials used in the previous internal validation. Despite this, the three-protein prediction model failed to accurately predict suboptimal surgical outcome in this cohort. Furthermore, given the poor predictive accuracy seen here, it is unlikely that the addition of further cohorts would change this finding significantly.

Future work should therefore focus on identifying different biomarkers that may offer more accurate prediction for the outcome of surgery.

Tables

Table 1 correlations between scorer 1 and scorer 2 for each protein expression assay

	R_2 value (95% CI)	p value
POSTN	0.749 (0.68 – 0.80)	<0.001
CXCL 14	0.700 (0.62 – 0.76)	<0.001
pSmad 2/3	0.836 (0.79 – 0.87)	<0.0001

Table 2 Comparison of AUC between each study cohort and for each model

Model	Area under Curve (AUC)		
	Reister et al	All cases (n=238)	Limited to PDS (n=202)
POSTN	0.81	0.55	0.56
pSmad2/3	0.79	0.53	0.54
CXCL 14	0.79	0.62	0.63
Combined model	0.87	0.59	0.58

References

1. Riester, M. et al. Risk prediction for late-stage ovarian cancer by meta-analysis of 1525 patient samples. *Journal of the National Cancer Institute* **106** (2014).

2. Reid, B.M., Permuth, J.B. & Sellers, T.A. Epidemiology of ovarian cancer: a review. *Cancer Biol Med* **14**, 9-32 (2017).
3. Arnold, M. et al. Progress in cancer survival, mortality, and incidence in seven high-income countries 1995-2014 (ICBP SURVMARK-2): a population-based study. *The Lancet. Oncology* **20**, 1493-1505 (2019).
4. Ferlay, J. et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* **136**, E359-386 (2015).
5. Fotopoulou, C. et al. British Gynaecological Cancer Society (BGCS) epithelial ovarian/fallopian tube/primary peritoneal cancer guidelines: recommendations for practice. *Eur J Obstet Gynecol Reprod Biol* **213**, 123-139 (2017).
6. Rose, P.G. et al. Secondary surgical cytoreduction for advanced ovarian carcinoma. *N Engl J Med* **351**, 2489-2497 (2004).
7. Horowitz, N.S. et al. Does aggressive surgery improve outcomes? Interaction between preoperative disease burden and complex surgery in patients with advanced-stage ovarian cancer: an analysis of GOG 182. *J Clin Oncol* **33**, 937-943 (2015).
8. Vergote, I. et al. Neoadjuvant Chemotherapy or Primary Surgery in Stage IIIC or IV Ovarian Cancer. *New England Journal of Medicine* **363**, 943-953 (2010).
9. Kehoe, S. et al. Primary chemotherapy versus primary surgery for newly diagnosed advanced ovarian cancer (CHORUS): an open-label, randomised, controlled, non-inferiority trial. *Lancet (London, England)* (2015).
10. Fagotti, A. et al. Randomized trial of primary debulking surgery versus neoadjuvant chemotherapy for advanced epithelial ovarian cancer (SCORPION-NCT01461850). *International journal of gynecological cancer : official journal of the International Gynecological Cancer Society* **30**, 1657-1664 (2020).
11. van Meurs, H.S. et al. Which patients benefit most from primary surgery or neoadjuvant chemotherapy in stage IIIC or IV ovarian cancer? An exploratory analysis of the European Organisation for Research and Treatment of Cancer 55971 randomised trial. *European journal of cancer (Oxford, England : 1990)* **49**, 3191-3201 (2013).
12. Chern, J.Y. & Curtin, J.P. Appropriate Recommendations for Surgical Debulking in Stage IV Ovarian Cancer. *Curr Treat Options Oncol* **17**, 1 (2016).
13. Fagotti, A. et al. Prospective validation of a laparoscopic predictive model for optimal cytoreduction in advanced ovarian carcinoma. *American Journal of Obstetrics and Gynecology* **199**, 642.e641-642.e646 (2008).
14. Bowtell, D.D. et al. Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. *Nat Rev Cancer* **15**, 668-679 (2015).
15. Riester, M. et al. Risk prediction for late-stage ovarian cancer by meta-analysis of 1,525 patient samples. *Cancer Research. Conference: 105th Annual Meeting of the American Association for Cancer Research, AACR* **74** (2014).
16. Shah, A.A., Frierson, H.F., Jr. & Cathro, H.P. Analysis of immunohistochemical stain usage in different pathology practice settings. *Am J Clin Pathol* **138**, 831-836 (2012).
17. Raab, S.S. The cost-effectiveness of immunohistochemistry. *Arch Pathol Lab Med* **124**, 1185-1191 (2000).
18. Bookman, M.A. et al. Evaluation of new platinum-based treatment regimens in advanced-stage ovarian cancer: a Phase III Trial of the Gynecologic Cancer

Figure Legends

Figure 1- consort diagram describing tumour sample selection process for inclusion in study

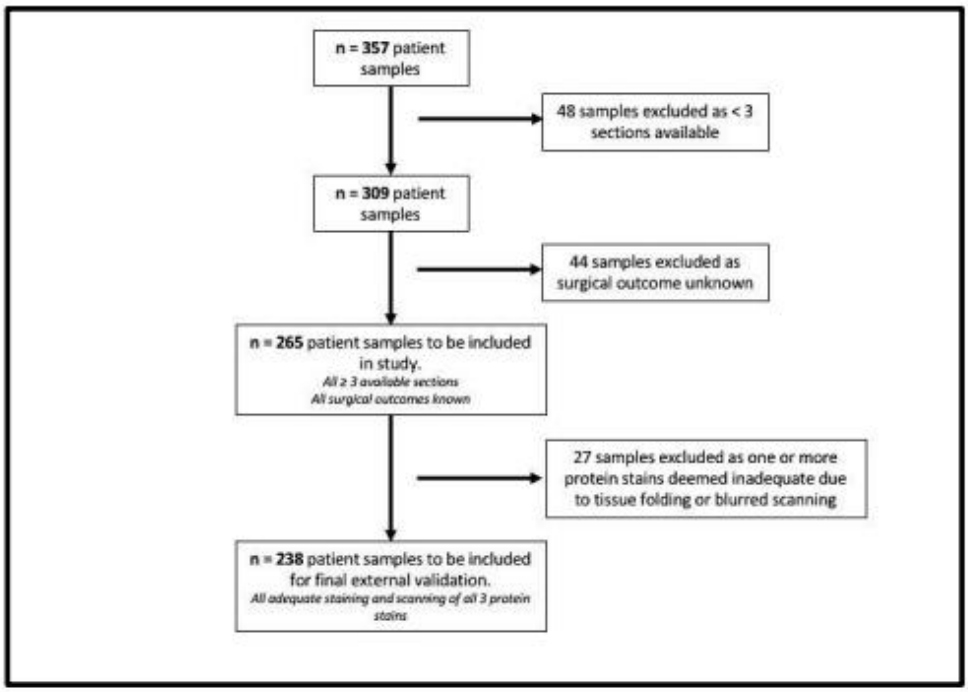
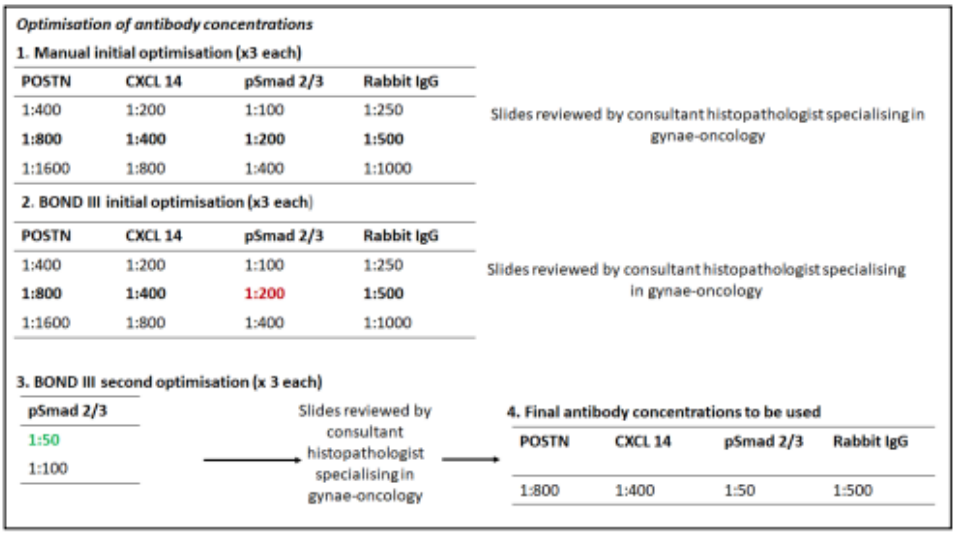
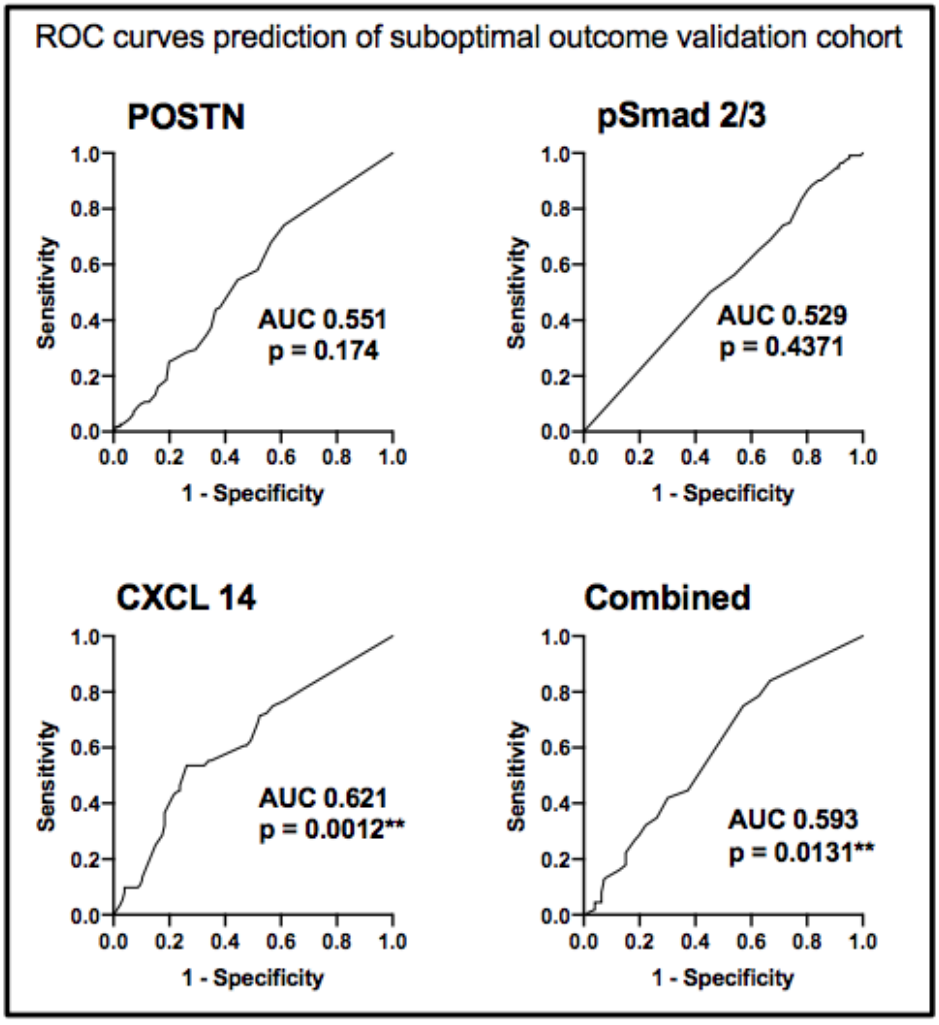


Figure 2. Optimisation process for antibody selection. Step one shows in bold antibody dilutionsused in original study. Three slides per dilution were stained for concentrations more and less dilute than the original. Following hand staining, stained slides were reviewed by the author and a consultant histopatologist with a specialty in gynae oncology, to ensure adequate staining.
Step two again describes the range of dilutions stained on the automated platform. Review agreed adequate staining for POSTN and CXCL 14, however psMAD 2/3 appeared under-stained at dilutions used in the original paper. For this reason a further optimisation step was performed and a dilution of 1:50 was then agreed to result in adequate staining.
Step four highlights antibody concentrations used in final IHC of whole validation cohort.



372 Figure 3- Receiver Operating Characteristics curves demonstrating poor predictive affinity
373 for suboptimal debulking rates in the validation cohort, n=238
374
375



376

Consent form allowing further study of trial data

[INSTITUTION-HEADED PAPER]

Supplementary Patient Consent Form - Biological Samples

ICON5 - A multi-arm clinical trial in ovarian or primary peritoneal cancer

Please initial boxes

1. I have read and understood the patient information sheet for the above study and have had the opportunity to ask questions and discuss it with my doctor. ☐
2. I agree to give a *sample of blood* for research in the above project. I understand how the sample will be collected, that giving a sample is voluntary and that I am free to withdraw my approval for use of the sample at any time without giving a reason and without my medical care or legal rights being affected. ☐
3. I agree to a *sample of my cancer*, removed during surgery, being used for research in the above project. I understand that I am free to withdraw my approval for use of the sample at any time without giving a reason and without my medical care or legal rights being affected. ☐
4. I agree that the samples and information collected about me will be stored on behalf of the Medical Research Council (MRC) for use in future projects, as described in the information sheet. I understand that some of these projects may be carried out by researchers other than the MRC, including researchers working for commercial companies. ☐
5. I understand that future research using the samples I give may include genetic research aimed at understanding the genetic influences on ovarian or primary peritoneal cancer, but the results of these investigations are unlikely to have any implications for me personally. ☐
6. I understand that I shall not benefit financially if future research leads to the development of new treatments or medical tests. ☐

.....	.....	.....
Name of patient (BLOCK CAPITALS)	Date (dd/mm/yyyy)	Signature
.....	.....	.....
Name of researcher (BLOCK CAPITALS)	Date (dd/mm/yyyy)	Signature

1 copy to be kept by the patient, a copy to be kept with hospital notes, and a copy kept in the local investigator's file.

10.3 Appendix C. Patient consent form for inclusion in MOCHR database

0995

NHS
Manchester University
NHS Foundation Trust

MFT BIOBANK

CONSENT FORM

Thank you for reading the information about giving blood, tissue and other body samples to the Manchester University NHS Foundation Trust (MFT) Biobank for biomedical research. **Please note that this process is separate from that related to any treatment you will receive and is also separate from that related to any clinical trials that you may also be asked to join.** If you would like to take part, please give us your consent by initialling the six questions below. Then sign the form and have it witnessed at the same time as you sign it.

	PLEASE INITIAL BOX
1. I have read and understood the MFT Biobank patient information leaflet entitled 'Consent to storage and use of samples for research' (version___), and have had the opportunity to ask questions. These questions have been answered clearly and satisfactorily and I understand the risks and benefits of giving my samples to the MFT Biobank.	☐
2. I give permission for my blood, tissue and other body samples to be donated to the MFT Biobank. I understand that the MFT Biobank will be custodian of these samples. I consent to their storage by the MFT Biobank and future use in regulated medical research, including genetic analysis, in the UK and overseas.	☐
3. I understand how the samples will be collected, that giving samples for research is voluntary and that I am free to withdraw my approval for use of the samples at any time without giving a reason and without my medical treatment or legal rights being affected. I understand that my samples would then be destroyed and my personal data erased from the MFT Biobank database.	☐
4. I give permission for information about me, provided by me or found in my medical records, to be supplied to and stored by the MFT Biobank for research purposes. I understand that some information may be held at different sources such as the NHS Information Centre or disease registries. I understand that the MFT Biobank will keep this information confidential at all times and will only pass information to researchers in an anonymous way that protects my identity.	☐
5. I understand that the samples may be used in studies carried out by researchers working for universities, hospitals and private/commercial organisations.	☐
6. I understand that I will not personally benefit, financially or otherwise, from my gift of blood, tissue and other body samples. This includes the circumstances of my samples being involved in research resulting in the development of a new treatment or medical test.	☐

Thank you for agreeing to make this gift to help research.

Donor name: _____ Signature: _____ Date: _____

I have explained to the donor the reasons for collecting, storing and using samples for research.
I am satisfied the donor signing this form understands the content and purpose of this consent form.

Person taking consent: _____ Signature: _____ Date: _____

CM10117 REV2 02/18

MFT Biobank consent form V7.0 (23/01/2018)
White copy: Biobank. Green: Patient. Pink: Pathology. Blue: Patient records.

10.4 Appendix D. Data collection guide

Introduction

Over the last few years we have compiled a comprehensive dataset for all patients with a diagnosis of Ovarian cancer at St Mary's, Manchester between 2013-2018 inclusive. This guide aims to help the replication of this dataset in other units, to allow for comparison and combination of data for analysis.

Also attached with this guide;

- An Excel spreadsheet with the required headers
- An excel spreadsheet with some data collected for reference
- A PDF with the ACE-27 calculator

Please be sure to fill your spreadsheet in the required format- this takes no extra time for the collector but means a lot of time saved when it comes to data comparison and interpretation.

Any questions regarding any data collection in this guide please direct to

Amy.hawarden@manchester.ac.uk OR
Richard.edmondson@manchester.ac.uk

The guide is in order of the data collection on the spreadsheet.

If a data field has been searched for, and the data is not available please fill the field with NA, rather than leaving it blank. This will inform that the data is not available, rather than incompletely collected.

General patient information

NHS number

10 digit NHS number

DOB

Date of birth dd/mm/yyyy

DOD

Date of death dd/mm/yyyy

Postcode

The post code of the current registered address the hospital holds.

IMD centile

The IMD score provides a decile ranking of deprivation for each geographical area of 1500 residents in the UK, where 1 is the most deprived and 10 is the least deprived. The score encompasses income, employment, education, health, including access to healthcare, crime, barriers to housing and services, and living environment to give an overall marker of deprivation 1. The decile is calculated by inputting the postcode into the tool as shown below. Tool can be accessed via the following link;

IMD calculator

It is possible to input several postcodes at the same time by copy and pasting the list straight from excel.
Click on 'get deprivation data' and the tool then outputs an excel sheet with lots of information- the column that is required is the column highlighted below and will be a number between one and ten.

Height, weight and BMI

Height must be measured in meters not cm, and weight collected in kg.
BMI can be calculated by inputting the formula =(weight/height^2) into the first cell in the BMI column. If this cell is then clicked on it will be surrounded by a green box. By clicking the small green box in the bottom left of the cell and dragging down to include all required cells the formula will be applied to all cells highlighted.

WHO Performance Status (PS)

Performance status (PS) is a WHO recognised tool widely used as a measure of fitness for treatment in oncology patients. It is useful to assess the acute fitness of a patient, but does not take into account co-existing co-morbidities. It is graded between 0-5, 0 being fully active and 5 being dead. 2

The required performance status is the one at the time of presentation. This is the earliest recorded PS that can be found in the patient notes. This is to attempt to gain an idea of fitness as soon in the patient journey as possible. For our data set this was often recorded on the referral letter, or on the MDT discussion on Somerset cancer registry.

Adult comorbidity evaluation-27 index (ACE-27)

The ACE-27 score quantifies co-morbidities present at the time of diagnosis. The score ranges from grade 0 (no comorbidities) to 3 (severe comorbidities)3. This score does not take into account the current acute state of the patient, but instead acts as a background marker of fitness.

The required ACE score is the one at the time of presentation. This is the earliest recorded ACE that can be found in the patient notes. For our data set this was often recorded on the referral letter, the outpatient letter with the gynae-oncology surgeon, or on the oncology outpatient review discussion. If the ACE score was not recorded, but the co-morbidites had been recorded free hand, then this was calculated by the data collector, using the tool attached as a PDF alongside this guide.

It is important to note that when calculating for our oncology patients, as it is used as a marker of non-cancer related fitness, our patient will not score for their CURRENT cancer. They will however score for previous diagnoses of cancer not related to the current episode.

In the first column please record the numerical score (between 0 and 3) and in the next column named 'comorbidities' please record in free text the condition that resulted in the score (the highest scoring condition).

Genetic test, material tested and Mutation

For these three columns, we are interested in genetic testing for BRCA mutations mainly, however if testing for other mutations have been undertaken for any reason please record these also.

Genetic test	Meaning
Yes	Testing has been performed, regardless of the outcome. This includes both germline and somatic testing.
No	Recorded evidence that no test has been performed OR no mention of genetic testing in patient notes- this can then be assumed as a no.

Material tested	Meaning
Germline	The patient themselves have been tested
Somatic	The patient's tumour itself tested
NA	'no' recorded in previous column

Mutation	Meaning
BRCA 1	
BRCA 2	
Other	Other mutation
None	Testing was performed but NO mutation was identified.
NA	'no' recorded in previous column

Evidence of testing at any point along the patient's journey can be recorded here. Often this was found in our dataset in the oncology notes, as well as outpatient letters with the genetics team.

Tumour information- Histology, grade, stage, substage, diagnosis date, method of diagnosis.

Histology and grade to be recorded as per the histology report. High grade serous and low grade serous can be recorded as HGS and LGS but for all other diagnoses please write in full. Grade should be recorded as high or low, again as per the histology report.

Stage is to be recorded as I-IV and substage as per the most up to date 2014 FIGO staging. Current link found below.

FIGO staging

It is the final staging of the cancer that is required, once all investigations are completed.

The date of diagnosis (dd/mm/yyyy) is the first recorded evidence that the patient has an ovarian cancer, but the following hierarchy applies;

1. Histology report- includes both biopsy results (either radiologically or via laparoscopy) and post operative histology reports. Record method of diagnosis as HISTOLOGY
2. If no histology available, cytology reports- earliest available. Record method of diagnosis as CYTOLOGY
3. If no histology or cytology and the diagnosis is made on a clinical basis then the date that this assumed diagnosis is recorded. Record method of diagnosis as CLINICAL.

MDT dates

Date of referral to MDT (dd/mm/yyyy)- The date that the referral was received by the tertiary team. This referral can be either from primary or more likely secondary care. This information is often recorded in Somerset cancer registry.

Date of first discussion at MDT (dd/mm/yyyy)- The date that the first MDT discussion of this patient took place. This is regardless of the outcome, for example even if if no treatment decision was made this date should still be recorded.

Date of MDT treatment decision (dd/mm/yyyy)- The date that the MDT decided which treatment pathway to send the patient down (i.e. surgical management, neoadjuvant chemotherapy, palliative treatment).

Blood parameters

Blood tests are required for collection at up to two different time points (P and PN as described below) depending on the clinical scenario (see table below). CA 125 is required only for time point P for all patients.

Clinical scenario	Bloods required
Primary surgery	P only
NACT and IDS	P and PN
NACT no IDS	P only

No surgery or chemo P only

Two time points of bloods are required

1. Presentation bloods (Hb P, Alb P etc)- should be the earliest recorded bloods for the period of illness attributed to the Ovarian cancer diagnosis. The aim is to identify the blood tests taken as early into the patient journey as possible i.e before interventions. During our data collection it was sometimes required to contact DGH units to access these. This was done via the Macmillan cancer team.
If test results are only available after interventions such as blood transfusions, albumin infusions, or ascitic drains, then the earliest bloods that are available should be used.
ALL diagnosis blood tests should be before the primary main treatment (primary surgery or starting neoadjuvant chemo therapy).
If no bloods are available pre-primary treatment then please input NA into the appropriate cells.
2. Post Neo-adjuvant chemotherapy bloods (PN Hb, PN alb etc)- should be the last set of bloods taken AFTER completion of the 3 cycles of chemotherapy, but BEFORE having Interval debulking surgery.

CT scan results

The CT scan referred to in the CT scan performed before primary surgery or neo-adjuvant chemotherapy. If a CT scan report is available before the primary treatment state yes, if not state no.
If no CT scan pre-primary treatment is available please do not include CT scans done post operatively, or done after chemotherapy.
In our unit our pre-treatment CT scans are reviewed by a consultant radiologist who specialises in gynae-oncology. Please try to use a report by a radiologist with a specialism in gynae oncology.
The suggestion of disease at any of the listed sites is denoted with a 1 in the corresponding cell. Absence of a suggestion of disease is denoted with a 0. If no CT was available please populate these cells with NA. This will make it clear which patients do not have a scan, and those who do not have disease reported in that site.

Surgical treatment

Surgical treatment refers to an operation that is performed for the treatment of the ovarian cancer.

It can be recorded as;
- Primary- Primary debulking surgery- any surgery performed before chemotherapy.
- IDS- interval debulking surgery- any surgery performed after chemotherapy, regardless of the number of cycles this was
- None- No surgery performed
- NA- unknown as to whether surgery was performed

Surgeries not to be included;

- look-in/diagnostic laparoscopies
- palliative procedures- such as defunctioning ileostomies where the main aim of the surgery was not to remove disease but to symptom relieve in the palliative setting
- Return to theatres- this should be recorded as a complication in the complication column.

Please state the name of the consultant surgeon who performed/supervised the procedure.

Please record the outcome of the surgery as one of the following definitions. This information should be found on the operative note. If not recorded by the surgeon, can be surmised using the definitions below if the appropriate information is available in the free text of the operative note.

Surgical outcome to be recorded	Meaning
Complete	No macroscopically visual residual disease
Optimal	Disease remains largest site ?1cm
Suboptimal	Disease remains largest site >1cm
NA	No surgery performed/information not available

Chemotherapy treatment

In the column 'had chemo', 'yes' denotes that the patient has received chemotherapy of any kind for their ovarian cancer. 'No' denotes that they received no chemotherapy. 'NA' denotes that this is not clear from the patient notes.

If the patient has completed ?6 cycles of chemotherapy this is recorded as yes, if the patient has received <6 cycles then this is recorded as no. If the number of cycles is known this is recorded as NA.

If the patient received PARP inhibitor at any point this is recorded. If it is unknown whether this is received this is recorded as NA.

References

1. Ministry of housing, c.a.l.G. (ed. c.a.l.G. Ministry of housing) (GOV.UK; 2015).
2. Su, J., Barbera, L. & Sutradhar, R. Do repeated assessments of performance status improve predictions for risk of death among patients with cancer? A population-based cohort study. Palliat Med 29, 547-553 (2015).
3. Kallogjeri, D. et al. Comparison of comorbidity collection methods. J Am Coll Surg 219, 245-255 (2014).

10.5 Appendix E. Full list of genes included in study, see table 4.2

Taken from (Abdallah, Chon, et al., 2015b)

TABLE 2. List of genes differentially expressed in the optimal cytoreduction group (maximal diameter of RD ≤ 1 cm) when compared to the suboptimal cytoreduction group in the training set

Symbol	*P*	Fold Change	Name
RRAGC	4.50E−06	0.80	Ras-related GTP binding C
MT2A	1.32E−05	0.61	Metallothionein 2A
FAAH	2.70E−05	0.80	Fatty acid amide hydrolase
MAP3K11	4.77E−05	0.85	Mitogen-activated protein kinase kinase kinase 11
KHSRP	6.87E−05	0.81	KH-type splicing regulatory protein
TRIP10	9.52E−05	0.84	Thyroid hormone receptor interactor 10
MT1P2	1.43E−04	0.68	Metallothionein 1 pseudogene 2
MT1H	1.56E−04	0.63	Metallothionein 1H
MT1G	1.62E−04	0.63	Metallothionein 1G
MT1X	1.64E−04	0.63	Metallothionein 1X
TYMP	1.69E−04	0.75	Thymidine phosphorylase
SLC26A6	1.82E−04	0.86	Solute carrier family 26, member 6
RPS6KA4	1.84E−04	0.91	Ribosomal protein S6 kinase, 90 kDa, polypeptide 4
MT1E	2.01E−04	0.60	Metallothionein 1E
IRF3	2.44E−04	0.83	Interferon regulatory factor 3
VPS39	2.84E−04	0.90	Vacuolar protein sorting 39 homolog (*Saccharomyces cerevisiae*)
RRAS	3.00E−04	0.83	Related RAS viral (r-ras) oncogene homolog
NNMT	3.19E−04	0.48	Nicotinamide *N*-methyltransferase
SIPA1	3.22E−04	0.88	Signal-induced proliferation-associated 1
PLEKHA4	3.30E−04	0.83	Pleckstrin homology domain containing, family A (phosphoinositide binding specific) member 4
MICAL2	3.39E−04	0.65	Microtubule associated monoxygenase, calponin and LIM domain containing 2
PADI2	3.61E−04	0.83	Peptidyl arginine deiminase, type II
GPR172A	3.81E−04	0.82	G protein–coupled receptor 172A
FBXL6	3.94E−04	0.78	F-box and leucine-rich repeat protein 6
BGN	4.25E−04	0.67	Biglycan
B3GAT3	4.39E−04	0.88	Beta-1,3-glucuronyltransferase 3 (glucuronosyltransferase I)
NUCB1	4.42E−04	0.84	Nucleobindin 1
VASP	4.47E−04	0.85	Vasodilator-stimulated phosphoprotein
SNRNP70	4.59E−04	0.83	Small nuclear ribonucleoprotein 70 kDa (U1)
ABCA7	6.06E−04	0.83	ATP-binding cassette, subfamily A (ABC1), member 7
INHBA	6.55E−04	0.55	Inhibin, beta A
FMNL1	6.65E−04	0.88	Formin-like 1
FAM50A	6.92E−04	0.76	Family with sequence similarity 50, member A
SYMPK	7.42E−04	0.87	Symplekin
GSDMD	8.85E−04	0.80	Gasdermin D
MBD3	9.28E−04	0.86	Methyl-CpG binding domain protein 3
MT1F	9.64E−04	0.66	Metallothionein 1F
PSMD7	9.58E−04	1.21	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 7
MAK	9.24E−04	1.20	Male germ cell–associated kinase

(Continued on next page)

TABLE 2. *(Continued)*

Symbol	*P*	Fold Change	Name
ADNP	8.54E−04	1.33	Activity-dependent neuroprotector homeobox
RPS15A	8.01E−04	1.16	Ribosomal protein S15a
ALK	6.80E−04	1.24	Anaplastic lymphoma receptor tyrosine kinase
CSDE1	6.72E−04	1.25	Cold shock domain containing E1, RNA-binding
C7	6.33E−04	1.67	Complement component 7
C6orf130	6.01E−04	1.26	Chromosome 6 open reading frame 130
TM9SF2	5.59E−04	1.35	Transmembrane 9 superfamily member 2
BAMBI	5.24E−04	1.31	BMP and activin membrane-bound inhibitor homolog (*Xenopus laevis*)
C1D	2.96E−04	1.24	C1D nuclear receptor corepressor
PUM2	2.46E−04	1.17	Pumilio homolog 2 (*Drosophila*)
MKL2	2.32E−04	1.30	MKL/myocardin-like 2
PEG3	1.95E−04	1.84	Paternally expressed 3
UBXN8	1.68E−04	1.27	UBX domain protein 8
KCTD12	1.26E−04	1.47	Potassium channel tetramerisation domain containing 12
PIGN	9.50E−05	1.19	Phosphatidylinositol glycan anchor biosynthesis, class N
STK24	8.78E−05	1.23	Serine/threonine kinase 24
RBMX	8.66E−05	1.24	RNA binding motif protein, X-linked
CLCN3	5.38E−05	1.29	Chloride channel 3
COLEC11	1.41E−05	1.68	Collectin subfamily member 11