

From the Department of Clinical Science and Education,
Karolinska Institutet, Stockholm, Sweden

**THE VALUE OF NEW SONOGRAPHIC
MODALITIES, DEMOGRAPHIC,
BIOMETRIC AND PROLIFERATIVE
VARIABLES FOR THE RISK PREDICTION
IN ENDOMETRIAL CANCER**

Rasmus W Green



**Karolinska
Institutet**

Stockholm 2020

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

Printed by Universitetservice AB

© Rasmus W Green, 2020

ISBN 978-91-7831-802-5

The value of new sonographic modalities, demographic, biometric and proliferative variables for the risk prediction in endometrial cancer

THESIS FOR DOCTORAL DEGREE (Ph.D.)

Publicly defended in Halldin lecture hall, elevator P, floor 0, Karolinska Institutet, Södersjukhuset, Stockholm

Friday 29th May at 9:00 am.

By

Rasmus W Green

Principal Supervisor:

Associate Professor Elisabeth Epstein
Karolinska Institutet
Department of Clinical Science and Education
Division of Obstetrics and Gynecology

Co-supervisor(s):

Associate Professor Joseph Carlson
Karolinska Institutet
Department of Oncology and Pathology

Professor Bo Baldetorp
Lund University
Department of Clinical Science
Division of Oncology and Pathology

Opponent:

Professor Elisabete Weiderpass Vainio
International Agency for Research on Cancer
World Health Organization

Examination Board:

Associate Professor Mark Clements
Karolinska Institutet
Department of Medical Epidemiology and
Biostatistics

Professor Ole Mogensen
Aarhus University
Department of Clinical Medicine
Division of Obstetrics and Gynaecology

Professor Daniel Altman
Uppsala University
Department of Women's and Children's Health
Division of Reproductive Health

To my child.

ABSTRACT

Background: Endometrial cancer is the sixth most common type of cancer among women worldwide and constitutes 4.8% of all cancer in women. The incidence is highest in high-income countries, given how age and obesity are the two most important risk factors. The majority of cases present at an early stage, with a favourable prognosis. In those patients, added treatment after surgery does not improve survival rates. In advanced stages, the prognosis is generally poor, and treatment options are few with a modest effect at best. Therefore, correct staging and triaging of patients to low- and high-risk groups are the cornerstone of endometrial cancer treatment. Today, risk stratification is based on imaging with ultrasound and/or magnetic resonance imaging and histologic assessment. These are all based on human interpretation and coding and can thus be more or less reliable. Improving reliability and diagnostic accuracy of our instruments, or finding more objective risk factors to complement or replace the current standard, is thus necessary to further tailor treatment to each patient. This thesis aimed to assess DNA ploidy and S-phase fraction as prognostic markers, measure the degree of rater dependency in ultrasound staging, investigate the possible benefit of adding dynamic contrast-enhanced ultrasound and compare diagnostic performance in assessing local tumour extension with ultrasound as well as the prevalence of known high-risk ultrasound features.

Methods: Study I used a population-based, consecutive cohort of 1140 women with FIGO stage I endometrioid endometrial cancer. Cox regression was used, including age, degree of differentiation, myometrial invasion, DNA ploidy, S-phase fraction and adjuvant treatment as covariates, with endometrial cancer death being the end-point. In study II, fifteen ultrasound experts assessed off-line 2D video clips and 3D volumes from 58 patients with endometrial cancer for deep myometrial invasion and cervical stromal involvement. Kappa statistics and diagnostic performance were calculated, and rater accuracy was correlated to rater experience measures. In study III, the added benefit to the routine ultrasound of dynamic contrast-enhanced ultrasound in diagnosing local tumour extension of endometrial cancer was assessed by comparing the results in a prospectively enrolled study cohort ($n=93$) to a matched control cohort ($n=279$). In study IV, pre- and postmenopausal women from the prospective IETA-4 multicentre cohort ($n=1538$) were compared concerning the prevalence of high-risk sonographic features and the diagnostic performance in assessing local tumour extension with ultrasound.

Results: A high S-phase fraction, but not DNA aneuploidy, was an independent prognostic factor for endometrial cancer death in FIGO stage I endometrioid cancer. Interrater reliability was higher with 2D video clips than with 3D volumes and diagnostic performance was also higher. Diagnostic performance was correlated to the number of cases each rater assessed annually. Dynamic contrast-enhanced ultrasound improved sensitivity in diagnosing deep myometrial invasion, without lowering specificity. Premenopausal women more often had low-risk cancer but still had high-risk sonographic features related to the vascularity of the endometrium, likely related to physiological features of a cycling endometrium. An intact

endometrial-myometrial border suggested low-risk disease in both pre- and postmenopausal women. Local tumour extension was more accurately assessed in pre- compared to postmenopausal women.

Conclusion: The utility and optimal cut-off value for S-phase fraction have to be sought in a new study before it can be introduced into clinical practice. Ultrasound is a reliable imaging modality but should be centralised to high-throughput centres to increase rater experience. Dynamic contrast-enhanced ultrasound is safe and does not require any special preparations and can thus be used as a complement in tricky cases to rule out myometrial invasion. Tumour vascularity in endometrial cancer assessment has to be interpreted in light of menopausal status.

POPULÄRVETENSKAPLIG SAMMANFATTNING

Livmoderkroppscancer, eller endometriecancer, drabbar slemhinnan i livmodern och är i Sverige den sjätte vanligaste cancerformen bland kvinnor. Den drabbar cirka trettio per hundrausen kvinnor årligen. De starkaste riskfaktorerna för livmoderkroppscancer är övervikt och högre ålder. Eftersom befolkningen blir både äldre och mer överviktig är livmoderkroppscancer ett växande problem. Lyckligtvis upptäckts livmoderkroppscancer i åtta av tio fall i ett tidigt stadium, då det vanligaste debutsymtomet är blödning efter klimakteriet, vilket ofta får kvinnor att söka hjälp tidigt. De flesta botas helt genom att livmodern opereras bort. Dock händer det att även bland dem med förmodat tidigt stadium får en del canceråterfall. Utmaningen är att hitta dessa egentliga högriskfall bland de förmodade lågriskfallen, helst redan innan operation. Högriskfall gynnas av att även lymfkörtlarna i bäckenet opereras bort, men lågriskfallen har ingen nytta av detta utan får istället bara biverkningar med kraftigt svullna underben på grund av sämre lymfdränage.

Idag görs detta urval genom att med ultraljud eller magnetkamera bedöma tumörutbredningen och genom att kategorisera olika tumörtyper i vävnadsprov i mikroskop. På detta sätt blir åtta av tio korrekt klassificerade som låg- eller högrisk. Kartläggning av portvaktscörtlar (eng. *sentinel node*) har länge använts inom bröstcancer- och hudcancerkirurgi och har börjat hitta in även till livmoderkroppscancerkirurgin. Dock är denna teknik fortfarande experimentell, tekniskt avancerad, tar längre tid under operation, är inte tillgänglig annat än på högspecialiserade centra och har, hittills, inte visats leda till bättre överlevnad. Därför kommer det, troligtvis, även i framtiden behövas bra metoder att klassificera livmoderkroppscancerpatienter som låg- och högrisk, särskilt då antalet patienter sannolikt kommer öka.

Studie I undersökte om S-fasfraktion och DNA aneuploiditet, två faktorer som relaterar till mängden DNA i celler, kunde förutsäga vem som dör i livmoderkroppscancer med hänsyn taget till tumörtyp, tumörcellernas grad av differentiering (hur ”ful” eller ”normal” en cancercell ser ut), ålder, tumörutbredning och given behandling. Totalt inkluderades 1140 patienter med tumör begränsad till livmodern från Södra Sjukvårdsregionen, vilket var alla fall i det området mellan 2001 och 2007. Vi fann att högre ålder, större tumörutbredning, ”fulare” tumörceller (lägre differentieringsgrad) och högre S-fasfraktion ökade sannolikheten att dö av livmoderkroppscancer, men DNA aneuploiditet hade ingen påverkan. Detta var nytt, då DNA aneuploiditet tidigare ansågs vara en riskfaktor som kvalificerade patienter för utökad kirurgi. Denna studie, bland andra, påverkade det nationella vårdprogrammet för livmoderkroppscancer så att DNA aneuploiditet togs bort som riskfaktor.

För att en diagnostisk metod ska vara användbar och tillförlitlig krävs att alla undersökare får samma resultat när de undersöker samma patient. Så är inte alltid fallet. Graden av tillförlitlighet mellan undersökare mäts med så kallad *kappastatistik*, som är ett mått på hur väl två undersökares resultat stämmer överens, utöver hur väl de hade stämt av ren slump. Skalan för kappa går från 0 till 1 där 0 är ingen överensstämmelse annat än slump och 1 är perfekt överensstämmelse. Två undersökare kan vara tillförlitliga så till vida att de får samma

resultat, men det betyder inte att det resultatet stämmer. Därför måste också den diagnostiska träffsäkerheten utvärderas mot något man bestämmer är ett facit. I **studie II** mätte vi graden av tillförliglighet mellan femton undersökare som fick bedöma tumörutbredningen i två-dimensionella ultraljudsvideoklipp och tre-dimensionella ultraljudsvolymer från femtioåtta patienter. Vi hade också tillgång till den patologiska bedömningen efter kirurgi, där tumörutbredningen kan bedömas exakt i mikroskop, vilket vi bestämde var vårt facit. Vi fann att undersökarna var mer samstämmiga och hade högre diagnostisk träffsäkerhet när de bedömde två-dimensionella videoklipp än tre-dimensionella volymer. Vi såg också att den diagnostiska träffsäkerheten samvarierade med hur många fall undersökaren utvärderade årligen. Vår slutsats blev att tillförlitligheten i ultraljudsbedömning av tumörutbredning är god men undersökningen kan vinna på att centraliseras så att varje undersökare får träffa många patienter och på så vis upprätthålla en hög diagnostisk träffsäkerhet.

Inom bedömning av levertumörer har *kontrastförstärkt ultraljud* använts under lång tid, men inte inom livmoderkroppscancer. Därför ville vi i **studie III** undersöka om tillägg av kontrastförstärkt ultraljud till den rutinmässiga undersökningen kunde öka den diagnostiska träffsäkerheten i bedömning av tumörutbredning. Vi undersökte 93 kvinnor med kontrastförstärkt ultraljud och jämförde mot 279 kvinnor som bara undersökts med det rutinmässiga ultraljudet. Diagnostisk träffsäkerhet för ett fynd kan uttryckas som andelen positiva bland dem som har fyndet (sant positiva), vilket kallas *sensitivitet*, och andelen negativa bland dem som inte har fyndet (sant negativa), vilket kallas *specificitet*. Vi kunde visa att tillägg med kontrastförstärkt ultraljud gav högre sensitivitet utan att specificiteten sjönk. I Sverige får 1300 kvinnor livmoderkroppscancer årligen. Med bara rutinmässigt ultraljud skulle nittioåtta kvinnor få sina bäckenlymfkörtlar bortopererade i onödan, men med tillägg av kontrastförstärkt ultraljud skulle det sjunka till trettionio.

Trots att de flesta fall av livmoderkroppscancer drabbar äldre kvinnor efter klimakteriet, så kallade *postmenopausala*, så utgörs ca. 10% av alla livmoderkroppscancerfall kvinnor som drabbas innan klimakteriet, så kallade *premenopausala*. På grund av denna relativa ovanlighet är livmoderkroppscancer hos premenopausala kvinnor inte särskilt beforskat. Därför ville vi i **studie IV** jämföra en rad olika faktorer mellan pre- och postmenopausala kvinnor; ärftlighet, livsstil, kroppsvikt, medicinska och gynekologiska faktorer och debutsymtom samt förekomst av olika ultraljudsfynd. Vi ville också undersöka om den diagnostiska träffsäkerheten i ultraljudsbedömningen av tumörutbredningen skiljde sig åt mellan grupperna samt utforska om övervikt och midjemått påverkade risken för högriskcancer olika. Vår databas innehöll totalt 1538 kvinnor, varav 161 inte genomgått klimakteriet. Premenopausala kvinnor hade fött färre barn, hade oftare tjocktarmscancer i släkten och hade oftare lågrisktumörer, trots att de hade haft sina symptom lägre tid innan diagnos. Trots att de oftare hade lågrisktumörer så hade de ofta kärlik vävnad som man annars oftast ser hos högrisktumörer. Därför behöver dessa ultraljudsfynd tolkas med försiktighet hos kvinnor innan klimakteriet. Den diagnostiska träffsäkerheten var högre bland kvinnor innan klimakteriet. Bland kvinnor efter klimakteriet var övervikt och högt bukomfång skyddande mot högriskcancer, men inget sådant samband sågs bland kvinnor

innan klimakteriet. Detta kan bero på att övervikt ökar risken för östrogenberoende cancer, som oftare är lågrisk, men övervikt ökar inte risken för östrogenoberoende cancer, som oftare är högrisk. Övervikt sänker också åldern för insjuknande. Därför kan det vara så att överviktiga kvinnor hinner få livmodern bortopererad på grund av en lågriskcancer innan de hinner utveckla en högriskcancer, varför övervikt kommer se ut som en skyddande faktor för högriskcancer.

De viktigaste rönen från denna avhandling är att S-fasfraktion är en potentiell ny kandidat i att förutsäga dålig prognos i livmoderkroppscancer, men det måste bekräftas i fler studier. Ultraljud för bedömning av tumörutbredning är en träffsäker metod med god tillförlitlighet mellan undersökare, men bör göras på dedikerade centra. Kontrastförstärkt ultraljud kan förbättra träffsäkerheten ännu mer, och är relativt enkelt att utföra, och kan således bli ett bra komplement i svårbedömda fall. Ultraljudsfynd ska tolkas i ljuset av om patienten har genomgått klimakteriet eller inte och undersökningen har särskilt hög träffsäkerhet bland premenopausala kvinnor.

LIST OF SCIENTIFIC PAPERS

- I. Green RW, Engblom S, Baldetorp B, Hartman L, Måsbäck A, Bjurberg M

Cell proliferation, measured as flow cytometric S-phase fraction, is a strong prognostic indicator in FIGO stage I endometrioid endometrial carcinoma: a population-based study.

Acta Obstet Gynecol Scand 2015; 94:1064-1073

- II. Green RW, Valentin L, Alcazar JL, Chiappa V, Erdodi B, Franchi D, Frühau F, Fruscio R, Guerriero S, Graupera B, Jakab A, di Legge A, Ludovisi M, Mascilini F, Pascual MA, van den Bosch T, Epstein E

Endometrial cancer off-line staging using two-dimensional transvaginal ultrasound and three-dimensional volume contrast imaging: Intermethod agreement, interrater reliability and diagnostic accuracy.

Gynecol Oncol 2018; 150: 438-445

- III. Green RW, Epstein E

Can dynamic contrast-enhanced ultrasound (DCE-US) improve diagnostic performance in endometrial cancer staging?: A proof of concept.

Ultrasound Obstet Gynecol 2019, e-pub ahead of print, doi:
10.1002/uog.21885

- IV. Green RW, Fischerova D, Testa AC, Franchi D, Frühauf F, Lindqvist PG, di Legge A, Cibula D, Fruscio R, Haak LA, Opolskiene G, Vidal Urbinati AM, Timmerman D, Bourne T, van den Bosch T, Epstein E

Sonographic, demographic and clinical characteristics of pre- and postmenopausal women with endometrial cancer; results from the IETA4 (International Endometrial Tumor Analysis) multicenter cohort.

Manuscript. Submitted to *Journal of Gynecologic Oncology*.

CONTENTS

1	Introduction.....	1
1.1	Epidemiology of endometrial cancer.....	1
1.2	Risk factors for developing endometrial cancer.....	2
1.3	Histopathology and molecular pathogenesis of endometrial cancer.....	2
1.4	Diagnosing endometrial cancer.....	4
1.4.1	Sonographic assessment of abnormal uterine bleeding.....	4
1.4.2	Endometrial sampling.....	4
1.4.3	Reliability of diagnostic techniques.....	5
1.5	Imaging in endometrial cancer.....	6
1.5.1	Ultrasound.....	6
1.5.2	Magnetic resonance imaging.....	10
1.6	Treating endometrial cancer.....	10
1.6.1	Surgical treatment and staging.....	10
1.6.2	Adjuvant treatment.....	10
1.6.3	Complications.....	11
1.6.4	Sentinel node biopsy.....	11
1.6.5	Ovarian preservation and fertility-sparing treatment.....	13
1.7	Prognostic factors in endometrial cancer.....	14
1.7.1	FIGO stage.....	15
1.7.2	Histologic subtype and grade.....	15
1.7.3	Myometrial invasion.....	15
1.7.4	Cervical stromal involvement.....	16
1.7.5	Lymphovascular space invasion.....	16
1.7.6	Age.....	16
1.7.7	DNA ploidy.....	16
1.7.8	S-phase fraction.....	17
1.8	Summary.....	18
2	Aims of the thesis.....	21
2.1	Study I.....	21
2.2	Study II.....	21
2.3	Study III.....	21
2.4	Study IV.....	21
3	Participants and methods.....	23
3.1	Participants.....	23
3.1.1	Study I.....	23
3.1.2	Study II.....	23
3.1.3	Study III.....	23
3.1.4	Study IV.....	23
3.2	Methods.....	24
3.2.1	Study I.....	24
3.2.2	Study II.....	24

3.2.3	Study III	25
3.2.4	Study IV	25
3.3	Statistical analysis	26
3.3.1	Comparing proportions between groups	28
3.3.2	Comparing means and medians between unpaired and paired groups.....	28
3.3.3	Analysing survival time between groups	28
3.3.4	Agreement and reliability	30
3.3.5	Diagnostic performance.....	33
3.3.6	Regression analysis.....	35
3.3.7	Correlations.....	36
3.3.8	False discovery rate methods.....	37
4	Results.....	39
4.1	Study I	39
4.2	Study II.....	39
4.3	Study III	39
4.4	Study IV	40
5	Discussion.....	41
5.1	Results in a clinical context	42
5.2	Methodological considerations.....	43
5.2.1	Selection bias	43
5.2.2	External validity.....	43
5.2.3	Internal validity.....	44
5.2.4	Confounders, mediators, moderators and covariates	44
6	Conclusions.....	46
7	Future perspectives	47
8	Acknowledgements	48
9	Errata.....	50
10	References.....	51

LIST OF ABBREVIATIONS

2D-TVU	Two-dimensional transvaginal ultrasound	LVSI	Lymphovascular space invasion
3D-VCI	Three-dimensional volume contrast imaging	MMR	Mismatch repair protein
AUB	Abnormal uterine bleeding	MMR-D	Mismatch repair protein deficient
BMI	Body mass index, defined as body weight (kilograms) divided by body height (meters) squared.	MRI	Magnetic resonance imaging
CI	Confidence Interval	NCCN	the National Comprehensive Cancer Network
D&C	Dilation and curettage	OR	Odds Ratio
DAG	Directed Acyclic Graph	ProMisE	Proactive Molecular Risk Classifier for Endometrial Cancer
DCE-US	Dynamic contrast-enhanced ultrasound	RCT	Randomized controlled trial
EBRT	External beam radio therapy	ROC-AUC	Receiver operating characteristics area under the curve
ESGO	European Society of Gynecological Oncology	ROI	Region of interest
ESMO	European Society of Medical Oncology	SCNAs	Somatic copy number alterations
ESTRO	European Society for Radiotherapy and Oncology	SLN	Sentinel lymph node
EU	European Union	SOE	Salpingo-oophorectomy
FDR	False discovery rate	TCGA	The Cancer Genome Atlas
FIGO	International Federation of Gynecology and Obstetrics	TIC	Time-intensity curve
FWER	Family-wise error rate	UK	United Kingdom
HR	Hazards ratio	US	United States of America
HRT	Hormone replacement therapy	WHO	World Health Organization
IETA	International Endometrial Tumor Analysis group		
ITCs	Isolated tumor cells		
LNG-IUD	Levonogestrel containing intrauterine device		

1 INTRODUCTION

1.1 EPIDEMIOLOGY OF ENDOMETRIAL CANCER

Endometrial cancer is the sixth most common type of cancer worldwide¹ and constitutes 4.8% of all cancer in women². The incidence rate is highest in Europe and North America with 13-19 per 100,000 women per year while in Africa and South-Eastern Asia it is the lowest with <5 per 100,000 women per year². Considering how age and obesity are the major risk factors for endometrial cancer, these differences in incidence rates reflect differences in longevity and lifestyle in these regions. Although, the largest increases in incidence rates during the last decade are found in countries that have made large socioeconomic transitions, such as South Africa, Brazil, India and China³. Endometrial cancer incidence will likely increase as the number of people over the age of 65 in the world is projected to increase, as shown in Figure 1. Worldwide overweight or obesity prevalence among adult women have increased from 30% to 38% from 1980 to 2013, adding to the endometrial cancer risk, but in developed countries, the increase in the incidence rate of overweight and obesity is slowing down⁴.

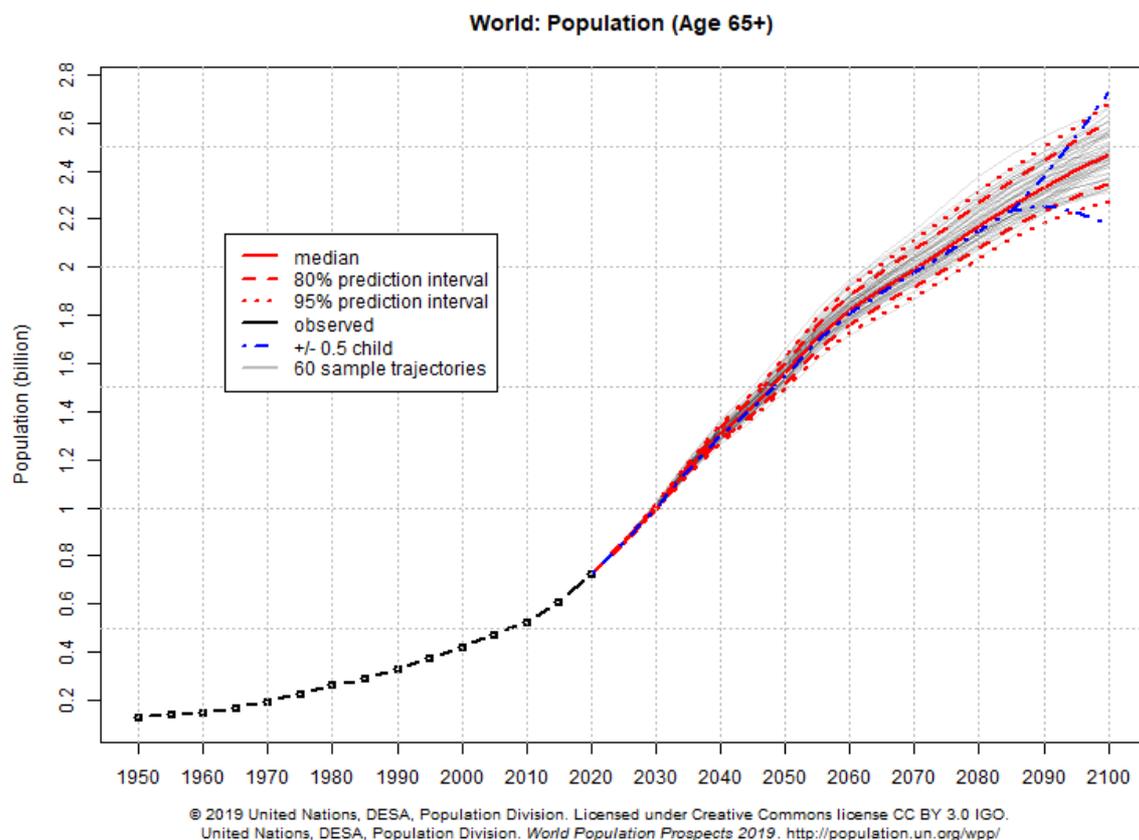


Figure 1. Graph showing the observed and projected number of people in the world aged over 65 for the coming decades. Data from the United Nations.

Mean age at diagnosis is 63 years⁵ and only 3-6% of endometrial cancer cases occur in women younger than 45 years of age, according to national statistics from the US⁶, the UK⁷

and Sweden⁸. Obesity is associated with a lower age at diagnosis for endometrioid endometrial cancer, but not for non-endometrioid endometrial cancer⁹.

1.2 RISK FACTORS FOR DEVELOPING ENDOMETRIAL CANCER

Apart from age and obesity¹⁰, risk factors for endometrial cancer are the metabolic syndrome¹¹, tamoxifen use for breast cancer¹², polycystic ovarian syndrome¹³, unopposed oestrogen in hormone replacement therapy (HRT) for menopausal symptoms^{14,15}, Lynch syndrome¹⁶ and reproductive factors such as early menarche, nulliparity and late menopause^{17,18}. In Lynch syndrome, the cumulative risk of developing endometrial cancer by age 70 is 34%¹⁶ and Lynch syndrome is found in 3% of endometrial cancer cases¹⁹. Apart from genetic factors, risk factors typically have in common that they increase exposure to oestrogen, unopposed by progesterone. Thus, fewer years of menstruation, higher parity and oral combined contraceptives^{18,20}, containing progestins, and intrauterine devices, containing levonorgestrel, reduces the risk of endometrial cancer^{21,22}. The individual contribution of each risk factor is hard to determine, since there is considerable overlap, e.g. between obesity and subfertility and low progesterone²³. Still, the relative risk of developing endometrial cancer is 2.9 per 10 unit increase in BMI (body mass index), which is higher than for any other type of cancer^{10,24}. Given the high background absolute risk, and the commonness of overweight and obesity, the population attributable factor, i.e. the percentage of cases that would not occur if a risk factor were reduced to a minimum, is 40-60% in the UK, the EU and the US population^{10,25}. Few studies exist on risk factors for premenopausal endometrial cancer, but obesity, nulliparity and irregular menses has been identified²⁶⁻²⁸. Risk factors differ between endometrioid and non-endometrioid subtypes of endometrial cancer, with HRT and oral contraceptive use affecting the risk for endometrioid endometrial cancer with little or no effect on the risk of non-endometrioid endometrial cancer²⁹. Obesity is a risk factor for both types, but much stronger for endometrioid endometrial cancer²⁹. A black ethnicity increases the risk of non-endometrioid endometrial cancer but lowers the risk of endometrioid endometrial cancer²⁹.

1.3 HISTOPATHOLOGY AND MOLECULAR PATHOGENESIS OF ENDOMETRIAL CANCER

Originally proposed by Bokhman in 1983, endometrial cancer has been categorized into endometrioid (type 1) and non-endometrioid (type 2)³⁰. Endometrioid endometrial cancer has a preserved glandular architecture with columnar epithelium and can be classified according to the FIGO (International Federation of Gynecology and Obstetrics) grading criteria as grade 1, grade 2 or grade 3 based on the percentage of nonsquamous or nonmorular solid growth pattern¹. Non-endometrioid endometrial cancer is classified as mucinous, serous, clear cell, undifferentiated or of mixed histology^{1,31}. This characterization is paralleled by characteristic genetic alterations where endometrioid tumours have genetic alterations in *PTEN*, *KRAS*, *CTNNB1*, *PIK3CA*, *ARID1A* and *MLH1* while non-endometrioid tumours have genetic alterations in *TP53* and *Her-2/ERBB2*^{32,33}. *PTEN* and *PIK3CA* are both parts of the PI3K-PTEN-AKT-mTOR signalling pathway, which is activated in >80% of all endometrioid

endometrial cancer³⁴ and in 50% of endometrioid endometrial cancer *PIK3CA* is mutated³⁵. Indeed, mutations in this pathway are more common in endometrial cancer than in any other type of cancer within the Cancer Genome Atlas (TCGA)³⁶. *KRAS* mutations are the predominant mechanism of activation of the MEK-ERK pathway in endometrioid endometrial cancer³⁷ and are found in one of five cases³⁸. *FGFR2* mutations can also activate this pathway³⁸. Mutations in *PTEN*, *PIK3CA* and *KRAS* can co-exist and one-third of endometrioid endometrial cancer have two of these three mutated genes³⁵ but *FGFR2* mutations are mutually exclusive to *KRAS* mutations³⁸. Thus, the same pathway can be activated through different gene mutations, one tumour can have more than one pathway activated and there is also crosstalk between these two pathways³⁷. A third molecular pathway in the oncogenesis of endometrioid endometrial cancer is the Wnt/ β -catenin pathway, in endometrial cancer typically activated through mutations in *CTNNB1*³⁶ found in one of five endometrioid endometrial cancers³⁸. *KRAS* and *CTNNB1* mutations are almost mutually exclusive³⁶ but *CTNNB1* and *FGFR2* can coexist³⁸ which suggest that there is also crosstalk between the Wnt/ β -catenin pathway and the MAPK-ERK pathway³⁶, possibly through stabilization of β -catenin and RAS³⁹. Non-endometrioid tumours are dominated by the serous subtype and constitute 5-10% of all endometrial cancer, with >90% having mutations in *TP53*^{36,40} and one of four having mutations in *Her-2/ERBB2*³⁶. Although, there is considerable overlap in genetic alterations between endometrioid and non-endometrioid endometrial cancer. For example, 40% of grade 3 tumours have *TP53* mutations⁴⁰ and serous tumours frequently show mutations in *PIK3CA*. In a landmark study from the TCGA Research Network, 373 cases of endometrial cancer (307 endometrioid, 66 serous and 13 of mixed histology) were analysed using comprehensive integrated genomic, transcriptomic and proteomic techniques³⁶. They were able to identify four distinct subtypes; *POLE* ultramutated, microsatellite instability hypermutated, copy-number low and copy-number high. The identification of *POLE* ultramutated was novel and these tumours had favourable prognosis, despite being found across all subtypes³⁶. *POLE* encodes a subunit of the DNA polymerase responsible for proofreading, thus suppressing or inactivating mutations of this gene results in higher mutation rates⁴¹. Microsatellites are short repeat segments of up to six bases repeated thousands of times over the genome. If there is a mismatch between the leading strand and the daughter strand in DNA replication in these segments, and also a failure to repair these mismatches by mismatch repair (MMR) proteins, this can cause frame-shifts and yield non-functioning protein. Lynch syndrome is characterized by germline mutations in one of four different MMRs⁴². In sporadic endometrial cancer with microsatellite instability, this is typically caused by epigenetic silencing of the MMR protein *MLH1*⁴³. The copy-number low and copy-number high groups are more difficult to characterize. Somatic copy number alterations (SCNAs) are deletions or duplications of a short part or the entire arm of a chromosome and is very common among many different forms of cancer⁴⁴. In TCGA the tumours were grouped into four clusters based on frequencies of SCNAs and the three clusters with least SCNAs was denoted copy-number low and the cluster with most SCNAs was denoted copy-number high. In copy-number low, 97% were of endometrioid histology with oestrogen and progesterone receptor positivity and

frequent mutations in *CTNNB1*. In the copy number high group, 90% had mutations in *TP53* and comprised all tumours of serous histology and one out of three grade 3 endometrioid tumours³⁶. Due to the cumbersome and expensive techniques used by TCGA, the Proactive Molecular Risk Classifier for Endometrial Cancer (ProMisE) was developed to mirror the four subgroups identified by TCGA; MMR-deficient (MMR-D), corresponding to the microsatellite unstable group, *POLE*-mutated, *p53 wt*, corresponding to copy-number low and *p53 abn*, corresponding to copy-number high⁴⁵⁻⁴⁷.

1.4 DIAGNOSING ENDOMETRIAL CANCER

1.4.1 Sonographic assessment of abnormal uterine bleeding

The cardinal symptom of endometrial cancer is abnormal uterine bleeding (AUB), which is the presenting symptom in >90% of cases^{48,49}. The prevalence of AUB in the general population is difficult to estimate due to variances in the definition of AUB, use of objective vs. subjective ways of diagnosing AUB, differences in help-seeking behaviour among women in different populations and varying access to gynaecological care. In premenopausal women >20 years of age, the prevalence of AUB is thus 10-46%^{50,51} and among postmenopausal women, it is 10-15%^{51,52}. Although, even among premenopausal women seeking medical attention for abnormal uterine bleeding the risk of endometrial cancer is only 0.33%⁵³ compared to 9% in women with postmenopausal bleeding⁵⁴. In postmenopausal women with AUB, the endometrial thickness, measured by two-dimensional transvaginal ultrasound (2D-TVU), can be used as an initial screening to triage the need for endometrial sampling. As such, the optimal cut-off should yield high sensitivity and low post-test probability rather than high specificity and high post-test probability, since the consequences of missing endometrial cancer are direr than performing an unnecessary endometrial biopsy. Several meta-analyses have tried to establish this optimal cut-off and results vary from 3-5 mm⁵⁵⁻⁵⁷. The current recommendation from the American College of Obstetricians and Gynecologists is a cut-off of ≤ 4 mm which yields a post-test probability of <1% of endometrial cancer⁵⁸. In premenopausal women the endometrial thickness varies from 5 to 11 mm⁵⁹, depending on cycling day, thus there is no cut-off to rule out endometrial cancer. Endometrial thickness measurements have poor diagnostic performance in ruling out polyps and endometrial hyperplasia⁶⁰.

1.4.2 Endometrial sampling

If endometrial sampling is warranted, the choice of endometrial sampling method is determined on the presence or absence of focal lesions. This is determined by installing saline or gel into the uterine cavity, i.e. hydrosalpingography, which provides a contrast between the cavity and the endometrium⁶¹⁻⁶³. Hydrosalpingography has an overall sensitivity of 95% and an overall specificity of 88% in detecting pathology in the uterine cavity⁶⁴, but in detecting endometrial cancer specifically, the diagnostic performance is heterogeneous, varying from a sensitivity of 28-80% and a specificity of 82-100%⁶⁵⁻⁶⁸ with a poor agreement between raters⁶⁹. Thus, hydrosalpingography, in cases with suspect endometrial cancer, can only be used

to triage for the most suitable endometrial sampling method. Focally growing lesions should be resected at hysteroscopy, to ensure a representative sample, since dilation and curettage (D&C) misses one in ten endometrial cancers and leaves 87% of focal lesions *in situ*⁷⁰. Although, in cases without focal lesions, the agreement between D&C and final pathology is 94%⁷⁰. Blind endometrial aspiration biopsy, using suction catheters (such as Endorette®, Pipelle®, Accurette®, Vabra® and similar devices) can replace D&C in select cases⁷¹, with a pooled positive likelihood ratio of 66, a negative likelihood ratio of 0.14⁷², a sensitivity of 81% and a specificity of 99.9%⁷³. With an endometrial cancer prevalence of 9% among women with postmenopausal bleeding, the absolute risk of having endometrial cancer with a positive biopsy is thus 87-99%, but still 1-2% with a negative biopsy. Thus, all women with AUB, despite reassuring biopsy, should be recommended to seek gynaecological care if a second episode of AUB occurs, since this patient group have a standardized incidence ratio of 2-17 (depending on the type of benign pathology) to develop endometrial cancer over four years⁷⁴.

1.4.3 Reliability of diagnostic techniques

The preoperative biopsy, however sampled, is assessed for presence of endometrial cancer and the histological subtype according to the World Health Organization (WHO) classification is determined. If the subtype is endometrioid it is graded according to FIGO criteria. Pathological evaluation is often the ‘gold standard’ against which other diagnostic modalities are later benchmarked in studies examining diagnostic accuracy, but pathological evaluation is by no means unailing. In situations where no ‘gold standard’ exists, we can at least be sure that when raters disagree at least one of them has to be wrong. Raters can also agree by random chance. This interrater reliability can be assessed using various forms of kappa statistics, which yields a number from -1 to 1, where 0 indicates no better agreement than chance, 1 indicates perfect agreement and -1 indicates perfect disagreement⁷⁵. In classifying endometrial thickness as either more than or less than 4.5 millimetres, Cohen’s kappa is 0.81⁷⁶. In diagnosing pre-operative biopsies as endometrioid or non-endometrioid, Fleiss’ kappa is 0.63⁷⁷. In assessing grade on preoperative biopsy or postoperative hysterectomy samples, Fleiss’ kappa is 0.57-0.65 depending on which grading system is used⁷⁷⁻⁷⁹. Diagnosis on preoperative biopsy does not always agree with the final diagnosis after hysterectomy. The overall pooled agreement between preoperative biopsy and hysterectomy for grade is 67% but increases to 89% if hysteroscopic resection was used⁸⁰. A clinically relevant (from grade 1 or 2 to grade 3 or *vice versa*) downgrading can be expected in 26% of cases, and a clinically relevant upgrading in 8% of cases⁸⁰. In high-grade endometrial cancer, the pathological evaluation is more challenging and a consensus agreement on the major histologic subtype is expected in only two of three cases^{81,82}. Immunohistochemical staining for a panel of markers can improve distinction between grade 3 endometrioid endometrial cancer and serous endometrial cancer^{81,82}. The concordance of ProMisE classification between preoperative biopsy and final hysterectomy is excellent⁸³. In diagnosing deep ($\geq 50\%$) myometrial invasion, Light’s kappa is 0.75⁸⁴. Thus, perfect diagnostic accuracy in diagnosing histologic subtype, risk category or disease spread between

any preoperative imaging modality and final pathological assessment is unattainable since neither imaging nor pathologic evaluation has perfect intramethod reliability between raters.

1.5 IMAGING IN ENDOMETRIAL CANCER

After an endometrial cancer diagnosis has been established, imaging is needed to assess the tumour spread, since this is the most important factor in deciding treatment options.

Localized spread in the uterus and adnexa is assessed by either expert ultrasound and/or magnetic resonance imaging (MRI) and distant spread is assessed with computed tomography of the chest and abdomen³². Deep myometrial invasion and cervical stromal involvement are factors included in endometrial cancer risk stratification and can be assessed with imaging. Not using imaging and classify endometrial cancer as low- or high-risk based on preoperative biopsy alone, will classify three out of four as preoperative low-risk but on postoperative pathological assessment about 42-47% of those will turn out to be high-risk, in need of re-surgery⁸⁵.

1.5.1 Ultrasound

1.5.1.1 Assessment of tumour extension

A transvaginal, two-dimensional, grey-scale, ultrasound (2D-TVU) is the first-line imaging in all women with AUB⁸⁶. It assesses the need for endometrial sampling in cases of thickened endometrium and can diagnose benign conditions that can cause AUB such as fibroids and adenomyosis. When an endometrial cancer diagnosis has been established, ultrasound can diagnose deep myometrial invasion, cervical stromal involvement and local extrauterine spread⁸⁶. In women with endometrial cancer, the ultrasound scan should be performed systematically to ensure the highest possible quality in the examination^{61,62}. A high-end ultrasound system should be used, with a two- or three-dimensional 3-5 to 9-10 MHz transducer. The transvaginal route can be complemented by a transabdominal scan if the entire uterus cannot be visualised. The transrectal route can be used if the transvaginal is inaccessible due to vaginal stenosis or similar conditions⁶². In premenopausal women, the examination should be performed on cycle day 4-6 to ensure that a thickened endometrium is not simply physiological. In women with irregular menses, a short course of progestins can be prescribed to induce endometrial shedding prior to the examination. The examination should be performed in the lithotomy position with an empty bladder. The cervix is assessed first and the image is magnified to include only the cervix and the examiner is looking for cervical stromal involvement. If the tumour is located close to the inner cervical os, the examiner can gently push the probe at the cervix to see if the tumour is sliding relative to the cervical canal, thus not truly invading into the cervix, or not, called a 'sliding sign'. Power Doppler can be used to assess the presence of abnormal vascularity in the region of the inner cervical os, which may indicate invasive tumour growth. If cervical stromal involvement is present, the parametria should be closely examined^{61,62}. The examiner should consider the origin of tumours mainly located in the cervical region since invasive cervical cancer is not treated primarily with surgery but with radiochemotherapy⁸⁷. Then, the uterine corpus is

assessed. The image should be magnified so that the corpus fills two-thirds of the screen and is then scanned in the sagittal plane from cornu to cornu and in the transverse plane from the cervix to the fundus and the anterior-posterior diameter and laterolateral diameter should be recorded. Endometrial thickness is measured in the sagittal plane perpendicular to the endometrial midline (since the transverse plane can be obliquely cut) and both layers are included. In the case of spontaneous fluid, the fluid-layer should be subtracted from the thickness measurement. If intracavitary pathology is present, this should be included in the measurement of endometrial thickness. If the endometrium cannot be clearly visualised, measuring should not be attempted, and thickness should be reported as 'non-measurable'. If the tumour is clearly visible, tumour size and extension should be assessed. Tumour size is measured in three orthogonal planes and volume is estimated using an approximal ellipsoid with $(d_1 \times d_2 \times d_3)/2$. The tumour extension can be expressed as minimal tumour free margin to the serosa in any plane.

1.5.1.2 Assessment of tumour morphology

After this initial assessment, the endometrium and the tumour morphology are described in more detail, using consensus criteria developed by the International Endometrial Tumor Analysis (IETA) study group⁸⁸. As such, the endometrium is assessed for echogenicity, which can be hyperechoic, isoechoic or hypoechoic, relative to the myometrium. Any echogenicity can be 'uniform' if it is homogenous and symmetrical, or 'non-uniform' if it is heterogenous, asymmetrical or cystic. The endometrial midline can be 'linear' if it is straight and uninterrupted, 'non-linear' if it is visible and uninterrupted but not linear, or 'irregular' or 'not defined' if it is interrupted or not visible. Presence or absence of a 'bright edge', a hyperechoic interface between the lesion and the endometrium, is noted. The endometrial-myometrial border is described as 'regular' if it is seen around the entire circumference of the uterine cavity in all planes, 'irregular' if it is seen around the entire circumference, but is irregularly shaped, 'interrupted' if it is seen only partially around the circumference or 'not defined' if it cannot be visualised in any plane. Spontaneous intracavitary fluid is classified as 'anechoic', 'ground-glass' or 'mixed'⁸⁸. Vascularity and blood flow are assessed using power Doppler. The Doppler box should include the endometrium with adjacent myometrium and an ultrasound frequency of 5 MHz or more, a pulse repetition frequency of 0.3-0.9 kHz and gain settings to reduce colour artefacts outside of vessels should be used. A semi-objective colour score from 1 'no flow' to 4 'abundant flow' is noted. Vessels are classified as 'single', if it is a dominant vessel entering the endometrium from the myometrium with or without branching, 'multiple, focal', if it is multiple vessels entering the endometrium from the myometrium but all stems from the same location, 'multiple, multifocal', if there are several vessels entering the endometrium from many locations, 'scattered' if there is vessel signal in the endometrium but without clear origin in the myometrium or 'circular' if there is a ring-shaped pattern surrounding the lesion⁸⁸.

This comprehensive characterisation is useful to distinguish between low- and high-risk features in endometrial cancer, where high-risk endometrial cancer are larger, more often

have ‘irregular’ endometrial-myometrial border, more often have ‘non-uniform’ endometrial echogenicity, higher colour score and more often a ‘multiple, multifocal’ vessel pattern⁸⁹. Although, each sonographic finding has only moderate discriminative abilities in predicting high-risk cancer and these factors also have strong collinearity and thus the independent predictive power of each factor is unclear⁸⁹. There is also limited interrater reliability in the assessment of these variables⁹⁰.

1.5.1.3 Predicting lymph node metastases

Using patient age, duration of AUB, histology on pre-operative biopsy, tumour extension on ultrasound, ultrasound findings consistent with endometrial cancer in cases of benign biopsy and tumour size and an added variable called “undefined tumour with unmeasurable endometrium” (considered an adverse finding), a mixed-effects logistic regression model could predict positive lymph nodes with a receiver operating characteristics area under the curve (ROC-AUC) of 0.73 (95% CI; 0.68-0.78). Compared to using only preoperative biopsy and tumour extension (current standard), the model had better performance at all cut-off values of the model. The authors also provided a decision curve^{91,92}, showing that at any level of the model, using the model provided net benefit compared to the current standard⁹³.

1.5.1.4 Diagnostic performance

In diagnosing deep ($\geq 50\%$) myometrial invasion, 2D-TVU has a pooled sensitivity of 78% (95% CI; 72-83%, 24 studies) and a pooled specificity of 81% (95% CI; 71-87%, 24 studies) in meta-analysis⁹⁴. In detecting cervical stromal involvement, the pooled sensitivity is 63% (95% CI; 51%-74%, 17 studies) and the pooled specificity is 91% (95% CI; 87%-94%, 17 studies)⁹⁵. In a prospective multicentre study, including seventeen European tertiary centres with twenty-six ultrasound examiners and 1538 patients, the sensitivity of subjective assessment in diagnosing deep myometrial invasion was 69.8% (95% CI; 65.8-73.7%) and the specificity was 79.5% (95% CI; 77.1-82.0%). In detecting cervical stromal involvement, the corresponding figures were 49.3% (95% CI; 41.4-57.3%) and 93.9% (95% CI; 92.6-95.1%)⁹⁶. Gynaecologists specialised in ultrasound are more accurate than general gynaecologists and also have higher interrater reliability, with a Fleiss’ kappa of 0.52 in assessing myometrial invasion and 0.58 in assessing cervical stromal involvement⁹⁷.

1.5.1.5 Novel ultrasound techniques

Three-dimensional volume contrast imaging (3D-VCI) is an ultrasound modality in which several two-dimensional serial sections are automatically collected and rendered into a three-dimensional volume post-acquisition by a software. This computer-rendering is sensitive to motion artefacts, thus the image is first optimised in 2D. Compared to 2D-TVU, this allows for off-line manipulation of the volume and analysis in any plane, which allows for more accurate volume measurements and three-dimensional visualisation of the vascular tree as well as automatic calculation of vascular indices. Although, this technology is not superior to 2D-TVU in endometrial cancer assessment, with reported sensitivities of 75% and specificities of 77% in diagnosing deep myometrial invasion^{98,99}.

Dynamic contrast-enhanced ultrasound (DCE-US) is a technique where an intravascular contrast-agent is injected at the ultrasound examination, providing contrast between richly vascularised tissue compared to less vascularized tissue. This contrast-agent consists of microvesicles of a phospholipid monolayer filled with inert gas, so small that they can pass through capillaries. Ultrasound travels poorly in gases, which provides the contrast. The flow of contrast in a region of interest can also be quantified over time, generating a time-intensity curve (TIC), shown in Figure 2. This has long been used in the assessment of hepatic lesions and in cardiology^{100,101}, and thus the safety profile of this procedure is well-known, the only contraindications being cardiac shunts and hypersensitivity. In the assessment of hepatic lesions, more emphasis is put on subjective pattern evaluation than objective TIC measurements¹⁰² and DCE-US can also be used in early evaluation of antiangiogenic therapy response¹⁰³. Other than having an ultrasound system compatible with this technique, and the actual injection, no other preparation is needed such as kidney function tests. In gynaecology, DCE-US can potentially distinguish between benign from malignant lesions in suspected ovarian cancer¹⁰⁴ and pelvic masses¹⁰⁵. It can also be used in the assessment of suspected cervical cancer¹⁰⁶. In malignant tumours, capillary blood flow is often increased, which on a TIC would show as earlier arrival time, shorter time to peak (i.e. steeper wash-in slope) and higher peak intensity compared to normal endometria or benign lesions. Thus, DCE-US can aid in distinguishing benign from malignant endometrial lesions^{107–111} but DCE-US in the preoperative diagnostic work-up of known endometrial cancer is understudied.



Figure 2: TIC curve of three separate region of interests. The green shape is outlining the tumour, the blue square is located in the subjectively highest contrast intensity within the tumour and the red square is located in healthy-looking myometrium. We see that contrast-intensity is higher in the tumour at every time-point. The legend displays automatically calculated TIC indices.

1.5.2 Magnetic resonance imaging

MRI is the first-line imaging modality most centres. MRI has different modalities such as diffusion-weighted imaging (DWI), T2-weighted imaging (T2WI), a combination of the two (DWI-T2WI) and contrast-enhanced imaging, which can be dynamic (DCE-MRI) or static. The current guidelines from the European Society of Urogenital Radiology recommend the combination of T2WI and DCE-MRI, with the option to add DWI¹¹². Meta-analysis comparing these modalities has not been able to tell which is better, but DWI-T2WI seem to have slightly higher specificity (~95% vs. ~86%), with no difference in sensitivity (~86%) in detecting deep myometrial invasion¹¹³. The pooled sensitivity and specificity in detecting cervical stromal involvement is 57% and 95%, respectively¹¹⁴. When using contrast is contraindicated, such as with kidney failure and allergy, DWI alone can be used. The American College of Radiology states that MRI is preferred over ultrasound in the preoperative assessment of endometrial cancer¹¹⁵, but this is unfounded. A meta-analysis of studies comparing MRI to ultrasound in the same set of patients have not shown any differences in sensitivity or specificity in detecting deep myometrial invasion¹¹⁶. Interrater reliability with MRI is only fair in diagnosing deep myometrial invasion, with a Fleiss' kappa of 0.39 (4 raters, 57 cases)¹¹⁷.

1.6 TREATING ENDOMETRIAL CANCER

1.6.1 Surgical treatment and staging

Treatment of endometrial cancer is determined using preoperative risk stratification. The ESMO-ESGO-ESTRO consensus considers FIGO grade 1 or 2 endometrioid tumours with <50% myometrial invasion (MI) as low risk; grade 3 tumours with <50% or grade 1+2 tumours with ≥50% MI as intermediate risk; and grade 3 tumours with ≥50% MI and all non-endometrioid tumours as high risk³². Standard treatment for low- and intermediate-risk cases is surgery with a total hysterectomy and bilateral salpingo-oophorectomy (SOE) without colpectomy, where pelvic and paraaortic lymphadenectomy is reserved for high-risk cases³². In randomized controlled trials (RCTs), lymphadenectomy has not shown any survival benefit in FIGO stage I endometrial cancer¹¹⁸. Although, a well-designed retrospective cohort study on stage I cases where lymphadenectomy was performed on all patients, found that adding paraaortic lymphadenectomy to pelvic lymphadenectomy alone had positive effects on overall-, disease-specific and recurrence-free survival in intermediate and high-risk cases, but not in low-risk cases¹¹⁹.

1.6.2 Adjuvant treatment

Pelvic external beam radiotherapy (ERBT) reduces the risk of locoregional recurrence but has not been found to improve overall- or disease-specific survival in any risk group of FIGO stage I endometrial cancer¹²⁰. In RCTs investigating the benefit of adding chemotherapy to ERBT in FIGO stage I-III cases, added chemotherapy can improve progression-free and disease-specific survival¹²¹. In the PORTEC-3 trial, comparing the addition of chemotherapy to radiotherapy alone in FIGO stage I-III, a *post-hoc* subgroup analysis of only FIGO stage

III cases showed both improved 5-year overall survival and failure-free survival with chemoradiotherapy vs. radiotherapy alone¹²². This group had the greatest absolute benefit of this treatment. It thus makes sense that treatment effect of adjuvant therapy is only found in more advanced stages or high-risk, early-stage, that has a higher likelihood of remaining tumour cells after a total hysterectomy and bilateral SOE, and that correct identification of advanced-stage disease has an impact on adjuvant treatment options.

1.6.3 Complications

Since stage I endometrial cancer tends to have a favourable prognosis, survivors will have time to develop long-term complications from the disease and its treatment. The most common complication of lymphadenectomy is lower limb lymphedema. With lymphadenectomy, the risk of lymphedema increases eight-fold with an absolute risk of 7%¹¹⁸. ERBT is associated with intestinal toxicity, both acute and late, and reduced quality of life scores and rectal and bladder incontinence ten years after treatment¹²⁰. Thus, RCTs on the efficacy of lymphadenectomy, stratified for preoperative risk groups, with survival end-points but also including quality of life measures are lacking but needed.

1.6.4 Sentinel node biopsy

To mitigate the risk of lymphedema, but still be able to assess lymph nodes for metastases to triage for adjuvant therapy, sentinel lymph node (SLN) mapping have been developed also for endometrial cancer surgery. Some centres have adopted this routinely, but it is still considered experimental by the National Comprehensive Cancer Network (NCCN)¹²³. The general idea with SLN mapping is that lymph nodes are connected in series, thus lymph drainage from a region reaches one specific lymph node before reaching other lymph nodes downstream. If this sentinel lymph node is free from metastases, it can be assumed that downstream lymph nodes in the same basin are also free from metastases. Therefore, knowledge of uterine lymphatic anatomy is important. From anatomical autopsies from the late 19th century, three principal lymphatic routes have been identified; (1) an upper, paracervical, route that drains to the internal and external iliac lymph nodes, (2) a lower, paracervical, route that drains to sacral lymph nodes and (3) an infundibulopelvic route, that drains directly to lumbal/paraaortic lymph nodes. These three routes are present on both the right and the left side of the pelvis¹²⁴. The two paracervical routes have many small communicating lymph vessels, and mostly drains the lower parts of the uterus and uterine cervix, but the infundibulopelvic route is more isolated and drains the fundus¹²⁴. These more than a century old anatomical findings have later been confirmed using tracer dye injection techniques that map the lymphatic system¹²⁵. Intuitively, this tracer dye should be injected peritumorally to better reflect the actual drainage, but this is not always technically feasible nor necessary. In patients where both pelvic and paraaortic lymphadenectomy have been performed, the incidence of isolated paraaortic lymph node metastases is only ~1%^{126–129}. Therefore, injecting tracer dye only in the cervix, and thus map mostly the upper and lower paracervical routes only, is acceptable¹²³ and is associated with higher rates of technical success (i.e. at least one SLN detected), higher rates of bilateral detection but lower rates of

paraaortic detection^{125,130}. If bilateral SLN mapping fails, an ipsilateral pelvic lymphadenectomy is recommended^{123,131}. All suspicious-looking or firm nodes should be removed regardless of tracer findings^{123,131}. These recommendations are formalized in the NCCN Clinical Practice Guidelines¹²³ and compliance with these guidelines can reduce the false-negative rate to <5% from 7-15% with the removal of tracer identified lymph nodes only^{132,133}. Although, meta-analysis show that technical success, defined as at least one SLN found in either hemipelvis, is found in only 81% of cases (95% CI; 77-84, 53 studies) and bilateral detection, defined as at least one SLN in each hemipelvis in an individual patient, is 50% (95% CI; 44-56, 36 studies) but can be increased to 75% if indocyanine green is used as tracer dye¹³⁰. In studies where an SLN mapped cohort is compared to a pelvic lymphadenectomy cohort, SLN mapping identifies more positive lymph nodes (14.7% vs. 9.9%, OR 2.03) with no difference in paraaortic detection rates¹³⁴. In studies where both successful SLN mapping and pelvic lymphadenectomy has been performed on all patients, the sensitivity to detect node-positive disease of SLN mapping compared to pelvic lymphadenectomy is 84-98.4%¹³⁵⁻¹³⁸. That SLN mapping can detect *more* lymph node metastases than pelvic lymphadenectomy seems counterintuitive since the positive SLN would likely be included in a pelvic lymphadenectomy and would thus be detected at equal rates regardless of method. But considering that; an adequate pelvic lymphadenectomy is often defined as the removal of only one or two dozen pelvic lymph nodes¹³⁹, an SLN can be detected in basins not usually included in pelvic lymphadenectomy^{136,140}, the SLN is the only positive node in 44-60% of node-positive cases^{136,140,141} and that the possible use of pathological ultrastaging of SLNs (finer sectioning or use of immunohistochemistry) can detect more true positive cases¹³¹, the superiority of SLN mapping can be explained. The prevalence of positive SLNs in apparent intrauterine cases, where the result of SLN mapping could change clinical management, ranges from 11% in preoperative low-risk cases to 20% in preoperative high-risk cases¹³⁴⁻¹³⁷. The NCCN guidelines recommend SLN mapping in all cases of apparent intrauterine disease, which is a paradigm shift compared to the ESMO-ESTRO-ESGO guidelines, where pelvic lymphadenectomy is only recommended in high-risk cases and can be considered in intermediate-risk cases for staging purposes^{32,123}.

It is unclear if SLN mapping improves long-term oncologic outcome. In two non-randomized, comparative cohort studies, with <3.5 years median follow-up time in each group, more positive lymph nodes were found in the SLN groups than in the pelvic lymphadenectomy groups (5.1% vs. 2.6% and 16.7% vs. 7.3%, respectively), yet adjuvant therapy was administered more often in the SLN group in only one of these studies (27.1 vs. 10.8% and 24.1% vs. 35.7%, respectively). Still, the disease-free survival was equal between the groups in both studies (>90% in all groups). Neither study specified any *a priori* statistical power analysis^{142,143}. Thus, it seems clear that SLN mapping can replace pelvic lymphadenectomy in cases where pelvic lymphadenectomy otherwise would be performed (i.e. high-risk cases without apparent extrauterine spread), since this would identify about twice as many positive lymph nodes, yet spare eight out of ten a full pelvic lymphadenectomy when the SLN is negative. Adding SLN mapping in cases where pelvic lymphadenectomy

otherwise would *not* be performed (i.e. low-risk cases) will identify occult metastatic disease in one in ten that would otherwise be missed, that might benefit from adjuvant therapy, but subject 5-40% of low-risk cases to a side-specific pelvic lymphadenectomy when only unilateral SLN mapping is achieved^{137,144}.

Although, SLN identified metastases in low-risk EC are different in clinical behaviour to lymph node metastases identified in high-risk cases. In a study where SLN mapping was performed on 519 patients with presumed early-stage low-risk and high-risk cases, 85 (16.4%) had a positive SLN. Of these 85, 43 (51%) had macrometastasis (i.e. ≥ 2.0 millimetres) but 11 (13%) had micrometastasis (i.e. metastasis < 2.0 millimetres) and 31 (36%) only had isolated tumour cells (ITCs). Adjuvant therapy was given based on positive lymph nodes (ITCs were not considered positive) or on high-risk features in node-negative cases. The 3-year progression-free survival was non-significantly different between those with no metastases, micrometastases and ITCs (85.5-95.5%) but significantly worse in those with macrometastases (58.6%, $p=0.0012$). Only one patient with ITC recurred, despite having had adjuvant chemoradiotherapy due to a deeply invading carcinosarcoma¹³⁸. Similar low recurrence rates among ITC patients have also been found by others¹⁴⁵⁻¹⁴⁷.

1.6.5 Ovarian preservation and fertility-sparing treatment

In premenopausal women, ovarian preservation can be considered in FIGO grade 1 tumours with $< 50\%$ myometrial invasion, if iatrogenic menopause is found unacceptable, after stringent work-up has ruled out synchronous ovarian cancer and/or metastases, *BRCA* mutations and Lynch syndrome³². Premenopausal women who still wish to conceive can be offered fertility-sparing treatment if the tumour histology is atypical hyperplasia with or without atypia or FIGO grade 1 endometrioid endometrial cancer and without overt myometrial invasion^{32,123}. Biopsy should ideally be obtained by hysteroscopy as total resection is possible, or D&C to ensure an adequate sample, and assessed by an expert gynaecologic pathologist. Myometrial invasion should be assessed by MRI or expert ultrasound. Women considering this option should be informed that it is not a standard treatment. Therefore, it should not be offered to premenopausal women routinely and be reserved for pressing cases. They should be referred to tertiary centres. Fertility experts should be consulted considering how premenopausal women with endometrial cancer more often are less fertile and nulliparous for medical reasons. Treatment consists of continuous progestins such as oral megestrol or medroxyprogesterone and/or levonogestrel containing intrauterine devices (LNG-IUDs) The endometrium should be resampled at three to six months. In responders, pregnancy can be encouraged after six months of treatment and surveillance continues as long as a pregnancy is not achieved. Standard surgery is recommended after childbearing, if surveillance shows progression or in non-responders after six to twelve months of initial treatment¹²³. With fertility-sparing treatment with any progestins, the complete response rate is 71% (95% CI; 63-77%, 18 studies) with no difference between oral progestins only compared to LNG-IUD only. Combining oral progestin with LNG-IUD have a complete response rate of 87% (95% CI; 75-93%, four

studies). The live birth rate for any progestins is 20%, and 35% with oral progestins combined with LNG-IUD. However, the relapse rate for any progestins is 29%¹⁴⁸.

1.7 PROGNOSTIC FACTORS IN ENDOMETRIAL CANCER

There are several definitions of low- and high-risk endometrial cancer in the literature. Generally, preoperative risk factors are used to triage for some form of pelvic lymph node sampling and postoperative risk factors are used to identify patients with a higher risk of recurrence or endometrial cancer death that could benefit from adjuvant therapy or require a more extensive follow-up program. Typically, all definitions include histology (endometrioid or non-endometrioid), FIGO grade in endometrioid cases, and myometrial invasion, indicating that these factors have a strong impact on the risk of lymph node metastases and recurrence. When comparing five different risk stratification systems, the “ESMO-modified”, which includes histology, FIGO grade, myometrial invasion and lymphovascular space invasion¹⁴⁹, had the highest performance in predicting recurrence and lymph node metastases, with a concordance index of 0.73 and a ROC-AUC of 0.80¹⁵⁰. When the systems are tested on endometrioid cases only, the ESMO-modified again performs best¹⁵¹. In detecting lymph node metastases, the sensitivity of the ESMO-modified risk stratification system is 91.4% (95% CI; 82.3–96.8) and specificity 63.9% (95% CI; 59.8–67.9)¹⁵¹. All stratification systems are designed to favour high sensitivity at the cost of lower specificity as the consequences of missing occult disease are considered direr than the side-effects of overtreatment. In seven to eight out of ten patients, the tumour is intrauterine, with a 5-year relative survival of ~95%, but in those with extrauterine spread the 5-year relative survival is only 20-60% depending on stage^{152–154}. Thus, the challenge is to identify indirect prognostic markers for extrauterine disease to better tailor treatment.

1.7.1 FIGO stage

Endometrial cancer is staged according to the FIGO criteria, which were revised in 2009¹⁵⁵.

Stage I	Tumor limited to the uterine body
IA	with no or less than 50% myometrial invasion
IB	with more than or 50% myometrial invasion
Stage II	Tumor invading the cervical stroma (but not only the endocervical mucosa or glands), no extrauterine spread
Stage III	Local and/or regional spread
IIIA	Tumor invading the serosa and/or adnexae.
IIIB	Vaginal and/or parametrial involvement
IIIC1	Metastases in pelvic lymphnodes
IIIC2	Metastases in paraaortic lymphnodes with or without metastases in pelvic lymphnodes
Stage IV	Tumor involvement in bladder and/or bowel and/or distant metastases
IVA	Invasion to bladder and/or bowel mucosa
IVB	Distant metastases, including intraabdominal metastases and/or inguinal lymphnodes

Positive wash cytology does not affect stage, but should be reported separately. FIGO grade does not affect stage¹.

Within each FIGO stage, histological subtype and FIGO grade can further identify subgroups with a worse prognosis. Thus, even though histologic subtype and FIGO grade are correlated to FIGO stage, those factors also carry independent prognostic information¹⁵⁶.

1.7.2 Histologic subtype and grade

Uterine papillary serous carcinomas and clear cell carcinomas have poorer prognosis at every stage compared to grade 3 endometrioid carcinomas^{157,158}. Not considering the FIGO stage, the 5-year disease-specific survival rate is ~50% for serous and ~60% for clear cell carcinomas¹⁵⁷. The copy-number high group in the Cancer Genome Atlas, with all serous cancers and many grade 3 endometrioid cancers that had a serous-like genotype, had the worst progression-free survival of the four subtypes identified³⁶.

1.7.3 Myometrial invasion

Deep myometrial invasion is the distinguishing feature between FIGO stage IA and IB, and all cases of IIIA and IVA has deep myometrial invasion. Lymph node positivity is found in 27% of women with deep myometrial invasion, but only in 6% without deep myometrial invasion¹⁵⁹. In multivariate analysis, when adjusting for histology, DNA ploidy and FIGO grade, deep myometrial invasion is associated with a 4-fold increase in relative risk for positive lymph nodes¹⁵⁹ and a 2.8-fold higher odds ratio for distant tumour recurrence when

adjusted for age, histology, FIGO grade and DNA ploidy¹⁶⁰. Interrater reliability among pathologists in assessing deep myometrial invasion is high, with a Light's kappa of 0.75⁸⁴.

1.7.4 Cervical stromal involvement

Cervical stromal involvement is the defining trait of FIGO stage II cancers. Cervical stromal involvement is highly associated with lymphatic spread¹⁶¹. Unfortunately, there is only modest interrater reliability between pathologist in assessing this feature, with a kappa of 0.49¹⁶².

1.7.5 Lymphovascular space invasion

Lymphovascular space invasion (LVSI) is the presence of tumour cells in lymph or blood vessels. It, therefore, makes intuitive sense that this finding correlates to lymph node recurrence, distant spread and poorer survival, even in apparent intrauterine cases^{163–167}. The interrater reliability of pathologists in assessing LVSI is good, with an intraclass correlation coefficient of 0.64¹⁶⁸.

1.7.6 Age

Older age increases the risk of non-endometrioid cancer and can limit treatment options due to co-morbidities and shorter life expectancy. Still, even after adjusting for stage, histology and adjuvant treatment, a one year increase in age confer a 3-4% increase in hazard ratio of death in endometrial cancer¹⁵⁷.

1.7.7 DNA ploidy

DNA ploidy refers to the total amount of DNA in a cell on a chromosomal level. A normal somatic cell in humans has 46 chromosomes, which is called diploid. In germ cells, the chromosome count is 23, which is called haploid. During cell division by mitosis in the M-phase of the cell cycle, the chromosome count is doubled, which is called tetraploid. Anything but an even factor of the normal chromosomal count is called aneuploid. DNA aneuploidy can be found in many different forms of cancer^{44,169}. The prognostic importance of DNA aneuploidy has been debated for about 30 years, much because of differences in inclusion and exclusion criteria and differences in risk grouping between studies, resulting in conflicting results. In general, studies that include all FIGO stages show poorer prognosis in DNA aneuploid cancers^{166,170,171}, while studies including only early-stage generally show no such association^{166,172–175}. Within the ProMisE classification system, DNA aneuploidy can be found in all four subtypes but does not affect overall- or progression-free survival within each subgroup, except for in the MMR-D group¹⁷⁶. DNA aneuploidy is not associated with lymph node metastases^{159,166}. Figure 3 displays the results of 14 select studies on the independent prognostic information of DNA ploidy in multivariate analysis. The publication year is on the x-axis and the lower bound of the confidence interval of the hazard ratio on the y-axis. The included covariates, and the studied endpoints, differs between the studies. Thus, this is not intended as a comprehensive meta-analysis, but as an illustration on why a consensus has been so hard to reach.

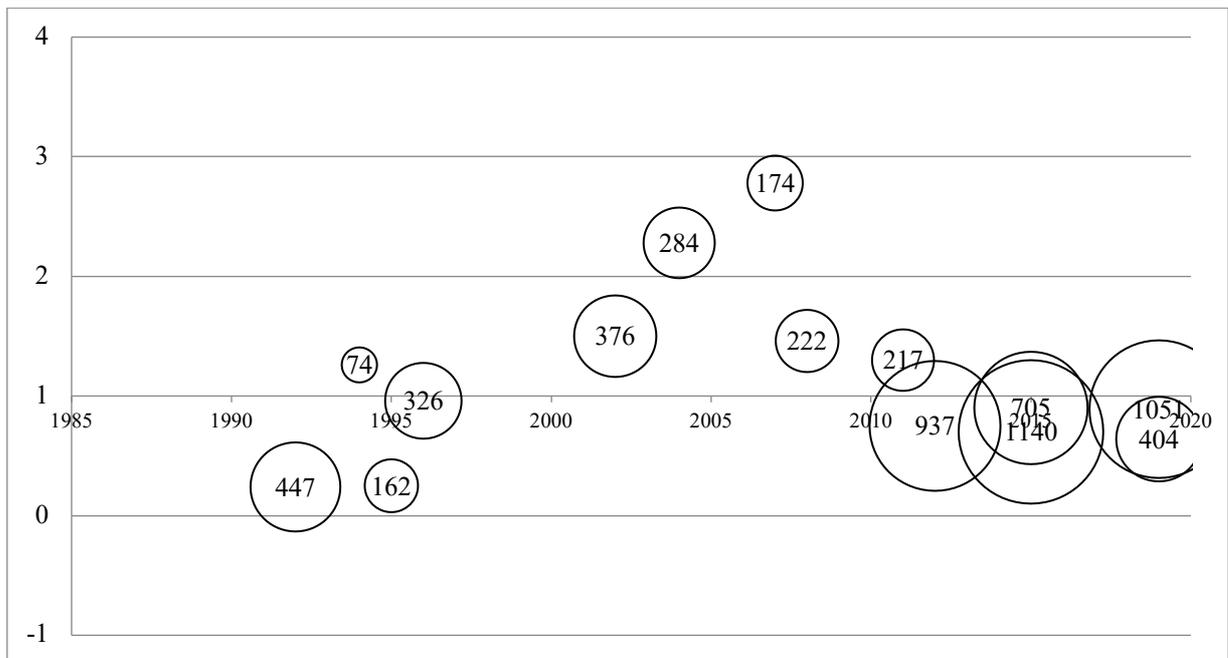


Figure 3: Bubble graph showing publication year on the x-axis and the lowest limit of the confidence interval of the hazard ratio of DNA aneuploidy in multivariate analysis on the end-point of interest (either overall-, relative-, cancer specific-survival or time-to-progression). The size of the bubble, and figure in the bubble, represents the size of the study population. Results from 14 studies from 1992 to 2019^{166,170,182-185,171,172,175,177-181}.

1.7.8 S-phase fraction

The S-phase fraction is the percentage of cells in a sample that is in the S-phase of the cell cycle and is thus a proliferation marker. Like DNA ploidy, S-phase fraction has been debated as a prognostic marker for decades. S-phase fraction and DNA ploidy can be assessed simultaneously with flow cytometry, thus most studies on S-phase fraction also include DNA ploidy. S-phase fraction is mostly reported as an independent prognostic factor in studies from the 1990s, but study population sizes are relatively small, typically around 160-200 patients^{172,177,179,186-191}. Effect sizes are also difficult to compare since some studies treat S-phase fraction as a continuous variable^{177,179,186,190}, some introduce a cut-off value¹⁷² and some does not report the effect size^{187-189,191}. In larger, population-based studies, S-phase fraction is generally not associated with worse outcome in multivariate analysis, but the methodological discrepancies are substantial^{175,184}. For those that treat S-phase fraction as a continuous variable, the hazard ratio for one unit increase in S-phase fraction is typically around 1.1, thus a 10% greater hazard rate per one per cent increase in S-phase fraction. Although, considering how the statistical power of survival analysis is based on the number of events, we can calculate the necessary number of events to detect such an effect. With an α -level of 0.05 and a statistical power of 0.8, an HR of 1.1 per unit increase in S-phase fraction and a standard deviation of S-phase fraction of 4.3 (figure from study I), the necessary number of events is 47, and even more events are needed if there is dependence among included variables in multivariate analysis. The event rate in a low-risk endometrial cancer population is not likely to exceed 10%, thus studies will need to include around 500 patients to detect such an effect. Thus, it seems likely that publication bias is in effect,

considering how few studies have been published with a negative finding on S-phase fraction.

Already in 1995, Jacobus Pfisterer *et al* expressed their frustration with the methodological discrepancies¹⁷²:

Several studies investigating the cellular DNA content and analysing the S-phase fraction of endometrial carcinomas have been published. Increased recurrence rates or decreased survival rates or both were correlated with aneuploidy and an increasing percentage of S-phase fraction. Other investigators could not confirm the latter results. A possible explanation for these conflicting data are large differences in material and laboratory methods. For example, there are different definitions of the term “aneuploid”. [...] The study populations are very heterogenous, and a precise definition of the patient population is not obvious in many cases. The number of patients is often too small to detect a relevant effect of DNA content. The role of therapy given to the patients remain unclear in some studies. Additionally, there are large differences in statistical analysis, e.g., different factors were used in the multivariate models to adjust for the effect of DNA content an S-phase fraction.

1.8 SUMMARY

Endometrial cancer is common and will be an increasing problem. For the individual patient, correct triaging is important to avoid unnecessary morbidity and mortality. For society, more tailored treatment can reduce waste of resources, which will become especially important in low-to-middle income countries where population growth, increased life expectancy and increased obesity rates will all contribute to higher endometrial cancer burden. There is no lack of potential prognostic markers that can aid in identifying low- and high-risk cases, but most have failed to be reproducible by others. The comprehensive integrated genomic characterization of endometrial cancer by the Cancer Genome Atlas Research Network provided a paradigm shift in our understanding of endometrial cancer genomics, and such knowledge is starting to make its way into clinical practice by the advent of the ProMisE classification. Although, despite our best efforts, in all low-risk groups, we observe a subset of patients with a worse-than-expected prognosis and outcome. Considering how many prognostic features, and the recorded outcomes, are subject to human coding and interpretation, we can never remove the risk of misdiagnosis and misclassification, further obscuring the relationship between prognostic factors and the studied outcome. Increased incidence rates in endometrial cancer are mainly driven by increases in endometrioid endometrial cancer, probably due to increases in obesity rates. Increasing the proportion of endometrioid cases, with a generally good prognosis, to non-endometrioid cases, with a generally poor prognosis, will improve survival rates even without improvements in treatments. This shift in distribution between low- and high-risk cancers in the past compared to the present can explain why we can observe an overall survival improvement, without

observing either histology-specific¹⁹² nor stage-specific improvements in survival¹⁹³. Thus, correct staging remains the cornerstone of endometrial cancer treatment. Sentinel-node biopsy has been suggested as a replacement to preoperative triage for lymphadenectomy. It is unclear if this provides any benefit in patients were lymphadenectomy would not be considered after preoperative work-up. Sentinel-node biopsy is a high-end resource; technically challenging, expensive, time-consuming, specialized and not without immediate and long-term risks. Thus, to meet the medical demand of this expanding patient group we will likely continue to rely on preoperative workup and risk stratification to better tailor treatment options.

2 AIMS OF THE THESIS

The overarching aim of this thesis was to increase the fidelity in triaging women with endometrial cancer to low- and high-risk groups by improving diagnostic methods and prognostic factors.

To achieve this the specific aims were:

2.1 STUDY I

To assess the independent prognostic power of S-phase fraction and DNA ploidy flow cytometric measurements in relation to histopathologic factors in FIGO stage I endometrioid endometrial cancer.

2.2 STUDY II

To assess the diagnostic performance of 2D-TVU and 3D-VCI in preoperative tumour extension assessment concerning intermethod agreement, interrater reliability and diagnostic accuracy.

2.3 STUDY III

To investigate if the addition of DCE-US to the routine 2D-TVU ultrasound assessment could improve diagnostic accuracy in tumour extension assessment and to explore if DCE-US subjective pattern recognition and quantitative TIC measurements correlated to high-risk features.

2.4 STUDY IV

To compare the diagnostic accuracy of ultrasound tumour extension assessment between pre- and postmenopausal women with endometrial cancer and to compare sonographic, demographic, anthropometric, lifestyle and clinical characteristics between the two groups. We also assessed if higher BMI and higher waist circumference affected the risk of high-risk cancer differently in pre- and postmenopausal women.

3 PARTICIPANTS AND METHODS

3.1 PARTICIPANTS

3.1.1 Study I

The study used a population-based, consecutive, prospective cohort from historical data with all cases of uterine cancer in the Southern Swedish Health Care Region between January 2001 and December 2007 ($n=1547$). Cases were identified using the Swedish Cancer Registry, with a coverage rate of 96.6% for female genital cancers¹⁹⁴. Data on histology and FIGO stage were obtained from patient records and all with missing data on histology or non-endometrioid ($n=161$) and FIGO stage II-IV ($n=246$ endometrioid) were excluded, leaving 1140 patients for analysis.

3.1.2 Study II

The study cohort was identified from the Stockholm part of the IETA4 cohort. We wanted an even mix of low- and high-risk cases, matching previous studies, since kappa statistics are sensitive to the trait prevalence. To avoid bias, patients were included consecutively in each of these two categories. A total of fifty-eight patients were included, first examined with 2D-TVU and 3D-VCI between March 2011 and December 2015, all by the same examiner. All ultrasound experts within the IETA collaboration were invited as raters to assess these 3D volumes and 2D video clips. Fifteen accepted participation and all completed the assessment.

3.1.3 Study III

This study compared the diagnostic accuracy of the addition of DCE-US to the routine 2D-TVU examination in a study cohort, compared to a routine cohort examined with only 2D-TVU just prior to the study period. All women referred to our centre for a preoperative 2D-TVU endometrial cancer assessment from January 2016 were asked to participate and enrolled consecutively. An *a priori* power analysis gave that we would need one hundred patients to detect an increase in diagnostic accuracy from 80 to 95% with DCE-US compared to routine ultrasound alone¹⁹⁵. One hundred and one patients were enrolled, the last in December 2017. Eight patients had to be excluded because of; cervical cancer ($n=4$), no surgery performed ($n=2$) and poor image quality ($n=2$), leaving 93 patients in the study cohort. The routine cohort was matched 3:1 to the study cohort on FIGO stage, FIGO grade and age, and thus included 279 women.

3.1.4 Study IV

This study used the complete IETA4 cohort. This cohort was collected from 17 gynaecologic oncology centres from seven European countries, enrolled consecutively at each centre between 1st of January 2011 until 31st of December 2015. The details of this cohort have been presented previously⁸⁹. A total of 1714 women were recruited and a total of 176 were excluded, leaving 1538 women after database lock. Exclusions were; hysterectomy not performed, the result of hysterectomy not obtainable or hysterectomy performed after more

than 120 day after ultrasound examination ($n=118$), a final diagnosis other than endometrial cancer or uterine carcinosarcoma ($n=25$), incomplete endometrial morphology assessment due to incorrectly filled study protocol ($n=26$), duplicate entries ($n=5$) and errors in the identification key ($n=2$). One hundred and sixty-one were premenopausal and 1377 postmenopausal.

3.2 METHODS

3.2.1 Study I

Data on age, FIGO stage, degree of differentiation, S-phase fraction, DNA ploidy status and adjuvant treatment were obtained from patient medical records, pathology reports and local laboratory reports. Seventy-one patients had missing data on DNA ploidy, 234 had missing data on S-phase fraction and three patients had missing data on the degree of differentiation, leaving 905 with complete data on all variables. The main end-point was endometrial cancer death, secondary end-points were overall survival and recurrence-free survival. Data on end-points were obtained from the Total Population Register, the Swedish Cause of Death Register and patient records. Data were analysed using cumulative incidence plots where death from other causes than endometrial cancer were considered competing risks. Univariate and multivariate Cox proportional hazards regression were used to assess the individual contribution of each variable on the risk of endometrial cancer death.

3.2.2 Study II

The 3D volume contrast imaging volumes and the 2D-TVU ultrasound video clips were compiled on an encrypted USB flash drive and distributed to the fifteen raters. The 3D volume and the 2D video clip were de-identified and randomised so that they could not be coupled, and raters were also blinded to the result of the pathological examination, which served as the 'gold standard'. Raters were asked to diagnose deep myometrial invasion and presence of cervical stromal involvement. They were also asked to rate their diagnostic certainty and their impression on image quality, using a Visual Analog Scale. Their responses were collected using an online survey form to reduce the risk of typing errors, which can be found at <https://tinyurl.com/w5t5e72>. Within each modality, the reliability of each rater pair was calculated, using Cohen's kappa, and the multirater Fleiss' kappa was calculated for all raters combined. For each rater, the proportion of agreement between the two modalities was calculated. McNemar's test was used to assess if differences in the assessment were systematic or random¹⁹⁶. Each rater's own sensitivity and specificity were calculated and the modalities were compared using the mean sensitivity and specificity. Diagnostic accuracy was then correlated to measures of rater experience, such as years as a second opinion sonographer, years as a board-certified specialist in gynaecology and numbers of endometrial cancer examinations performed annually.

3.2.3 Study III

All women in the study cohort were first examined using 2D-TVU and power Doppler and then DCE-US was added. Since blinding between the modalities was impossible, the joint subjective assessment of both modalities regarding deep myometrial invasion and cervical stromal involvement was recorded. With DCE-US, an intravenous contrast agent (in our case SonoVue®, Bracco International B.V., Amsterdam, the Netherlands) is injected during the ultrasound examination, with an ultrasound system that was compatible with this method is used. The contrast agent consists of microvesicles about one micrometre in diameter, made from a phospholipid monolayer filled with inert gas, making the contrast agent strictly intravascular and able to fill vessels even on a capillary level. At the same time, the contrast agent was injected, a video recording of the region of interest (ROI) began and ran for 180 seconds, which showed the contrast entering and the leaving the ROI. The signal intensity in the ROI was plotted against time post-acquisition, generating a time-intensity curve (TIC). From this TIC, objective measures of contrast flow were taken, such as wash-in slope, time-to-peak, peak intensity, and the area-under-the-TIC. We plotted three curves for each case; ROI1, defined as a 5×5 millimetre square inside the tumour, ROI2, defined as a 5×5 millimetre square in normal-appearing myometrium and ROI3, defined as a free-hand drawn tracing of the whole tumour. ROI1 and ROI3 were then normalized for ROI2. The contrast flow was also described in subjective terms; if the filling pattern and wash-out were ‘focal’ in the tumour only, or ‘global’ in the entire uterus and if the tumour/endometrium filled ‘prior’, ‘simultaneous’ or ‘after’ the surrounding myometrium. We then compared the diagnostic performance, using sensitivity and specificity, in diagnosing deep myometrial invasion and cervical stromal involvement, between this study cohort and the routine cohort. The prevalence of objective and subjective DCE-US parameters was also compared between low- and high-risk cases to see if any of these parameters were high-risk features.

3.2.4 Study IV

Pre- and postmenopausal women were compared regarding demographic (age, parity), presenting symptoms (abnormal bleeding, and duration of abnormal bleeding if present), family history of cancer among 1st-degree relatives (ovarian, breast, colon or other types of cancer), co-morbidities (diabetes mellitus type 2 and essential hypertension), use of hormonal medication, lifestyle (smoking, alcohol and exercise) and anthropometric measures (bra cup size, weight, height, BMI, body constitution and waist circumference). Pre- and postmenopausal women were also compared regarding sonographic variables; presence of a measurable endometrium, presence of a defined tumour, fibroids, adenomyosis, a regular endometrial-myometrial border, endometrial echogenicity, bright edge sign, endometrial midline, colour score, vascular pattern, endometrial thickness and tumour volume, defined according to IETA consensus criteria⁸⁸ and categorised according to findings from the first IETA4 study⁸⁹. Finally, the result of final pathological examination of the hysterectomy specimen was collected; histologic type, FIGO grade among endometrioid cases and FIGO stage, from which we categorized cases into low- and high-risk group according to ESMO-ESGO-ESTRO consensus criteria³². All variables were compared between pre- and

postmenopausal women using Fischer’s exact test and the Mann-Whitney U test, as appropriate. The effect of high BMI and high waist circumference on the risk group classification was investigated using logistic regression analysis, stratified for menopausal status. The sensitivity and specificity in diagnosing deep myometrial invasion and cervical stromal involvement with 2D-TVU were also compared between pre- and postmenopausal women. Considering that many hypotheses were tested, we adjusted the significance level of the p -values using the Benjamini-Yekutieli procedure.

3.3 STATISTICAL ANALYSIS

The statistical analyses used throughout this thesis is presented below in order of appearance. In all studies, continuous parameters in the study population are described using measures of central tendencies such as medians and means, and measures of spread, such as interquartile range and standard deviations and categorical parameters by frequencies and percentages. Subgroup comparisons were done using **Fischer’s exact test** for all categorical parameters, and the **Student’s t -test** or the **Mann-Whitney U test** for unpaired continuous parameters, and the **Wilcoxon Signed Rank test** for paired parameters. The choice of descriptive statistics and hypothesis testing in subgroup comparison for continuous variables was based on the normality of the parameter. If both normal and non-normal parameters were present, non-parametric tests were preferred throughout each study considering how non-parametric tests have almost as good statistical power as parametric tests, even when test assumptions for parametric tests are met, and perform better in small samples¹⁹⁷, as was the case in the included studies. In general, the result of inferential statistics can be summarized as in Table 1.

		The unobserved real world	
		The H_0 is true	The H_0 is false
The result of the statistical test at a certain significance level	Retain the H_0	True negative, probability= $1-\alpha$	Type II error, probability= β
	Reject the H_0	Type I error, probability= α	True positive, probability= $1-\beta$

Table 1. All possible outcomes of inferential statistics hypothesis testing. The type I error rate (α) is typically set at 0.05. The statistical power is the true positive probability ($1-\beta$).

For study I and study III, *a priori* power analyses were conducted to estimate the necessary sample sizes to answer the main study objective. Typically, the statistical power is dependent on the significance level α , the effect size one is interested in detecting and the sample size. If two are known, the third can be calculated. Thus, for study I, we assumed equal group sizes regarding S-phase fraction and 5-year overall survival of 90% in women with low S-phase fraction. For DNA ploidy, we assumed that 80% would have diploid tumours and that the 5-year overall survival in those women would be 80%. We were only interested in

finding hazard ratios between low- and high S-phase fraction, and between diploid and aneuploid tumours, of at least 2. Thus, we saw that we needed 500 and 300 women to achieve 80% statistical power. For study III we wanted to be able to detect an increment in diagnostic accuracy with DCE-US compared to 2D-TVU alone from 80% to 95%, with 90% power, which gave 100 patients in each group¹⁹⁵. In study II and IV, no power analysis was done. For study II, the number of cases was limited to not overburden the participating raters, while still include more cases than all similar, previously published studies. The number of raters was limited by recruitment and eligibility of the raters, while still including more raters than all similar, previously published studies. Study IV was exploratory and similar comparisons between pre- and postmenopausal women had not previously been undertaken, why a power analysis was not performed, and the study was limited to the size of the acquired database. Although, when planning all studies in the IETA collaboration and designing and acquiring the database, the aim was to recruit at least 1500 patients to allow for subgroup analysis, including comparisons between pre- and postmenopausal women⁸⁹.

In study I, we used survival analysis, i.e. time-to-event data analysis. **Cumulative incidence curves**, where death from other causes was considered a **competing event** was used and **Cox proportional hazards regression** was used in obtaining hazard ratios in uni- and multivariate modelling. To calculate *p*-values in Cox regression, the Wald test or the likelihood ratio test were used. In study II, **Cohen's kappa** and **Fleiss' kappa** were used in assessing **interrater reliability**, i.e. to what extent raters obtain the same result when assessing the same trait using the same method. **Proportions of agreement** were calculated to assess agreement between two diagnostic methods, i.e. to what extent two diagnostic tests provide the same result and **McNemar's test** was used to assess if non-agreement was systematically biased. In study II, we also used **Spearman's rank correlation coefficient** to assess if rater accuracy correlated to rater experience measures. Diagnostic performance, where a diagnostic test is compared to a 'gold standard', was assessed by calculating **sensitivities, specificities** and **overall accuracy** in study II-IV. In study III, this was also complemented by calculating **negative** and **positive predictive values**. In study III, we used **logistic regression analysis** to predict adverse binary outcomes, and **ROC curves** were generated for these predictions. In study IV, logistic regression was used to predict binary adverse outcomes from BMI and waist circumference measurements. Paired tumour size measurements with 2D-TVU and DCE-US were compared using **Bland-Altman plots** in study III. Unique for study IV was the use of the **Benjamini-Yekutieli procedure** to control the **false discovery rate** in a setting with multiple hypothesis testing.

For study I, Stata 13.1/SE (StataCorp, College Station, TX, USA) was used. For study II, IBM SPSS v.24 (International Business Machines Corp.; New Orchard Road; Armonk; New York 10504; 914-499-1900), Stata 12.1/IC (StataCorp, College Station, TX, USA) and GraphPad Prism 7 (GraphPad Software, 2365 Northside Dr., Suite 560, San Diego, CA 92108, USA) were used. For study IV, Stata 12.1/IC was used, with the user-written command `multproc` for the Benjamini-Yekutieli procedure.

3.3.1 Comparing proportions between groups

Proportions of categorical variables between groups are compared using Fisher's exact test. Most typically, proportions are presented in a 2×2 table, but Fisher's exact test is valid for any $m \times n$ table. Fisher's exact test is valid for all sample sizes and especially suitable for skewed tables. An alternative to Fisher's exact test is the chi-square test, which approximates the p -value asymptotically from the chi-square distribution, and is thus only valid for larger cell values, but is computationally easier to perform. Now, with modern computational power, there is no real reason not to opt for Fisher's exact test unless the dataset is very large.

3.3.2 Comparing means and medians between unpaired and paired groups

Parameters are often presented with some measure of central tendency, typically the median or mean. Around this measure, a measure of spread is also often reported. Depending on the type of parameter and the distribution, different central tendency and spread measures are preferred. For continuous variables, the mean, with standard deviations, is preferred if the distribution is normal, if it is skewed the median, with interquartile ranges. Groups are often compared on this measure of central tendency, where the null hypothesis is that measures are equal, and the alternative hypothesis that they are not equal. We also need to know if groups are paired or unpaired/independent, as this will affect our choice of hypothesis testing.

The Student's t -test is used to compare the means of normally distributed data. There is a version for paired and one for unpaired/independent samples. If data are not normally distributed and independent, the Mann-Whitney U test is preferred, if its dependent the Wilcoxon signed-rank test is used. The Mann-Whitney U test is valid for several different sets of null and alternative hypotheses, but the most commonly stated hypothesis test for this test is that the distribution of data between two groups are equal versus that the probability of a random value from one group will exceed a random value in the other groups is not equal to 0.5¹⁹⁸. The Mann-Whitney U test is often considered an alternative to Student's t -test when the normality assumption does not hold, but the Mann-Whitney U test thus actually answers a different question relating to distribution rather than comparing measures of central tendency. Thus, it is possible to get a significant Mann-Whitney U test even if population means are identical, which has been shown in simulation studies to happen $>5\%$ of the time¹⁹⁹.

Therefore, in small study samples, even small deviations from normality can have a huge impact on the result from the Student's t -test, and thus the Mann-Whitney U test can perform well, but in large study samples when the research question is a comparison of means, Student's t -test is robust even with deviations from the normality assumption.

3.3.3 Analysing survival time between groups

The methods used to compare survival between groups in this thesis include Cox proportional hazards regression²⁰⁰ and cumulative incidence curves with competing risks. A more general term for survival analysis is time-to-event analysis since the event of interest need not be death. Survival analysis takes into account that different subjects may have different time

under risk for the event, and therefore cannot be compared on the incidence of the event alone as this could bias result if time under risk differs. Thus, the *hazard rate* relates the risk to the time under risk, as such:

$$\text{Hazard rate} = \frac{\text{events}}{\text{time under risk}}$$

For Cox proportional hazards regression, the general formula is;

$$\lambda(t|X) = \lambda_0(t)e^{\beta X}$$

where $\lambda(t|X)$ is the hazard function of an individual at time t given a set of covariates X , which is equal to a baseline hazard $\lambda_0(t)$ multiplied with the exponential function e to the power of a coefficient β multiplied with the value of the covariate X . From this follows;

$$\ln[\lambda(t|X)] = \ln[\lambda_0(t)] + \beta X$$

and

$$\ln[\lambda(t|X)] - \ln[\lambda_0(t)] = \beta X$$

thus, the difference in an individual hazard of covariate X and the baseline hazard at time t is a constant βX over all time points. This is the main assumption of Cox proportional hazards regression, that hazards are proportional over time and that covariates do not change the hazard over time. Cox proportional hazards regression does not make assumptions on the baseline hazard, why Cox proportional hazards regression is sometimes called semi-parametric. Thus, the absolute hazard is not reported, but hazards can be compared by calculating hazard ratios relative to the baseline were all covariates have X equal to zero. In Table 2 we can see the result of a Cox proportional hazards regression model being fitted on 906 patients with FIGO stage I endometrioid endometrial cancer that were censored at 5 years of follow-up, death from any cause was the event and 129 had the event.

Covariate	Hazards	95%	p
S-phase	1.04	1.01-1.07	0.005
Age	1.09	1.07-1.11	<0.001

Table 2: Result of a Cox proportional hazards regression analysis. Adapted from study I.

Here, the hazards ratio is 1.04 for every unit increase in S-phase fraction and the hazards ratio is 1.09 for every year increase in age. The proportional hazards regression assumption gives that these increased hazards should be equally large across all levels of S-phase fraction and age, i.e. the increase in hazards is the same with an increase from 2 to 3 in S-phase fraction as for 4 to 5, and the same for 45-year-olds to 55-year-olds as between 65-year-olds and 75-year-olds. However, the validity of this assumption is dependent on the situation and the research question. In this example, death from any cause is the event, and as a 5-year increase in age is likely related to a larger risk of death for a 75-year-old than a 45-year-old, this assumption is unlikely to be valid. We can fix this by categorizing age, which allows us to

compare the hazard ratios to a reference group (typically the group with the lowest hazard, in this case, the lowest age). Now we see that the oldest (≥ 75 years old) have a much higher hazard ratio of dying from any cause (HR=6.00) than the second youngest (HR=1.25), relative to the youngest age group (≤ 60 years old). In fact, the 61-70 age group has a non-significantly different hazard ratio relative to the youngest age group, confirming our suspicion that the effect of an increase in age is different in different age groups.

Covariate	Hazard ratio	95% confidence interval	<i>p</i>
S-phase	1.05	1.01-1.08	0.003
Age			
≤60	Ref.		
61-70	1.25	0.63-2.50	0.515
71-75	2.55	1.27-5.13	0.009
≥75	6.00	3.31-10.8	<0.001

Table 3: Results of Cox proportional hazards regression with age as a categorized variable. Adapted from study I.

Careful consideration has to be taken when covariates (X) and the effects on the hazard of the covariates (β) can change over time since this time-varying effect is not accounted for in a standard Cox proportional hazard regression analysis, even though such adaptations are available.

In survival analysis, the survival time is censored when we no longer can obtain data on the outcome. Patients thus contribute with survival time up to the censoring event at which we know they did not have the event. Censored patients are accounted for as non-events up to that time point and assumed to have the same risk of experiencing the event like those that we can still follow. Thus, they contribute with time under risk but not any events, thereby reducing the hazard rate. This may not pose a problem in our interpretation of hazard ratios if censoring is non-informative, i.e. random, but if the reason for censoring is differential to the covariates we are comparing in our survival analysis, our hazard ratios can be biased. One such example could be censoring due to migration, where healthier patients with better prognosis are more likely to migrate than sicker patients with a worse prognosis. In standard survival analysis, death from other reasons (heart attacks, shark attacks) than our study interest (endometrial cancer) is also treated as a censoring event, which will bias results since once a patient have died from another cause the patient has a zero risk of later dying of endometrial cancer, and not an equal risk as those that we can still follow. This *competing risk* has to be taken into account, and there are various methods do to so²⁰¹, which is beyond the scope of this method summary.

3.3.4 Agreement and reliability

The terms ‘agreement’ and ‘reliability’ are often used interchangeably in the literature but are two distinct concepts. Agreement relates to what extent different raters assign the same rating

at repeated measures, where repeated measures can be two or more raters using the same rating method, one rater using two different rating methods, or one rater using one rating method but at two different time points. In this case, the absolute measurement error is of interest, but the variability between subjects or the prevalence of the examined trait does not matter. Reliability is instead related to the ratio of variability between ratings on the same subjects to the total variability between all study subjects and thus answers how well the instrument can distinguish between subjects^{202,203}. Thus, it is possible to have a high agreement but low reliability, if the variability between measures is very low or if the trait prevalence is very low or very high. If measurement variability is very low, compared to the total variability, subjects will look more alike than they actually are, making them hard to distinguish. If the total variability is very low, subjects are indeed very much alike, and thus hard to distinguish. In both these scenarios, reliability will be low even if the agreement can be high. This is termed *the prevalence problem* with reliability²⁰⁴. Therefore, it is difficult to compare reliability between studies if the prevalence of the examined trait differs. Calculating agreement in a binary classification is straightforward and shown in Table 4.

		3D-VCI	
		# negatives	# positives
2D-TVU	# negatives	12, <i>a</i>	17, <i>b</i>
	# positives	4, <i>c</i>	25, <i>d</i>

Table 4. The results of one rater, examining the same set of patients with two different methods. Adapted from study II, Table 1, Rater 3.

$$\text{Overall agreement} = \frac{\text{concordant cases } (a + d)}{\text{all cases } (a + b + c + d)} = \frac{12 + 25}{12 + 17 + 4 + 25} \approx 64\%$$

$$\begin{aligned} \text{Negative agreement} &= \frac{\text{negatives with each method } (a + a)}{\text{negatives with either method } (a + b + a + c)} \\ &= \frac{12 * 2}{12 + 17 + 12 + 4} \approx 53\% \end{aligned}$$

$$\begin{aligned} \text{Positive agreement} &= \frac{\text{positives with each method } (d + d)}{\text{positives with either method } (c + d + b + d)} \\ &= \frac{25 * 2}{4 + 25 + 17 + 25} \approx 70\% \end{aligned}$$

There are several ways of assessing reliability. For binary classification, kappa statistics are typically used, and for continuous measures intraclass correlation coefficients. The most commonly used kappa statistics is Cohen's kappa²⁰⁵. Kappa is a relation between the observed agreement and the expected agreement by chance, and calculated as such:

$$\kappa = \frac{p_{\text{observed}} - p_{\text{expected}}}{1 - p_{\text{expected}}}$$

		Rater 4		
		# negatives	# positives	
Rater 1	# negatives	34	10	g_1
	# positives	3	11	g_2
		f_1	f_2	N

Table 5. Cross-tabulation of Rater 1 and Rater 4 when diagnosing deep myometrial invasion with 2D-TVU. Adapted from study II, Table 2.

$$p_{observed} = overall\ agreement = \frac{34 + 11}{58} = 0.776$$

$$p_{expected} = \frac{f_1 g_1 + f_2 g_2}{N^2} = \frac{(37 * 44) + (22 * 14)}{58^2} = \frac{1628 + 308}{3364} = \frac{1939}{3364} = 0.576$$

$$\kappa = \frac{0.776 - 0.576}{1 - 0.576} = \frac{0.2}{0.424} = 0.47$$

A kappa close to zero indicates no better reliability between raters than would be expected by chance and a kappa close to 1 indicates perfect reliability, but what constitutes an adequate kappa is dependent on the situation being assessed. Cohen's kappa can only be used with two raters (or one rater at two points in time). For multiple raters, Fleiss' kappa can be used.

McNemar's test is used on paired nominal data to check for *marginal homogeneity*. Looking at Table 4, marginal homogeneity would give that $a+b = a+c$, and $c+d = b+d$. Since the "double negatives" (a) and "double positives" (d) are part of both sides of the equation these cancel out and McNemar's test thus only utilizes the discordant cases.

$$\chi^2 = \frac{(b - c)^2}{b + c}$$

$$\chi^2 = \frac{(17 - 4)^2}{17 + 4} = \frac{168}{21} = 8$$

For a sufficiently large number, χ^2 has a chi-square distribution with 1 degree of freedom. From a chi-square table, we can see that a value of 8 will result in a p -value of 0.004678. Thus, we can assess if errors or discrepancies between two diagnostic tests, or a diagnostic test compared to a 'gold standard' reference, are systematic or random. Here, McNemar's test showed that Rater 3 systematically overdiagnosed deep myometrial invasion with 3D-VCI relative to 2D-TVU (17:4).

With continuous variables measured by two different methods, a Bland-Altman plot can be drawn²⁰⁶, assessing the agreement between the methods. Bland and Altman recognised that measurements can be perfectly correlated but still not agree if systematic bias exists between the methods. In a Bland-Altman plot, the x -coordinates are the mean of the measurements across pairs, and the y -coordinates are the difference of the measurements across pairs. The x -axis intersects the y -axis at the mean of all calculated differences across pairs and *limits of*

agreement is drawn at ± 1.96 standard deviation of the difference from the x -axis. Thus, 95% of measurements are expected to fall within this range.

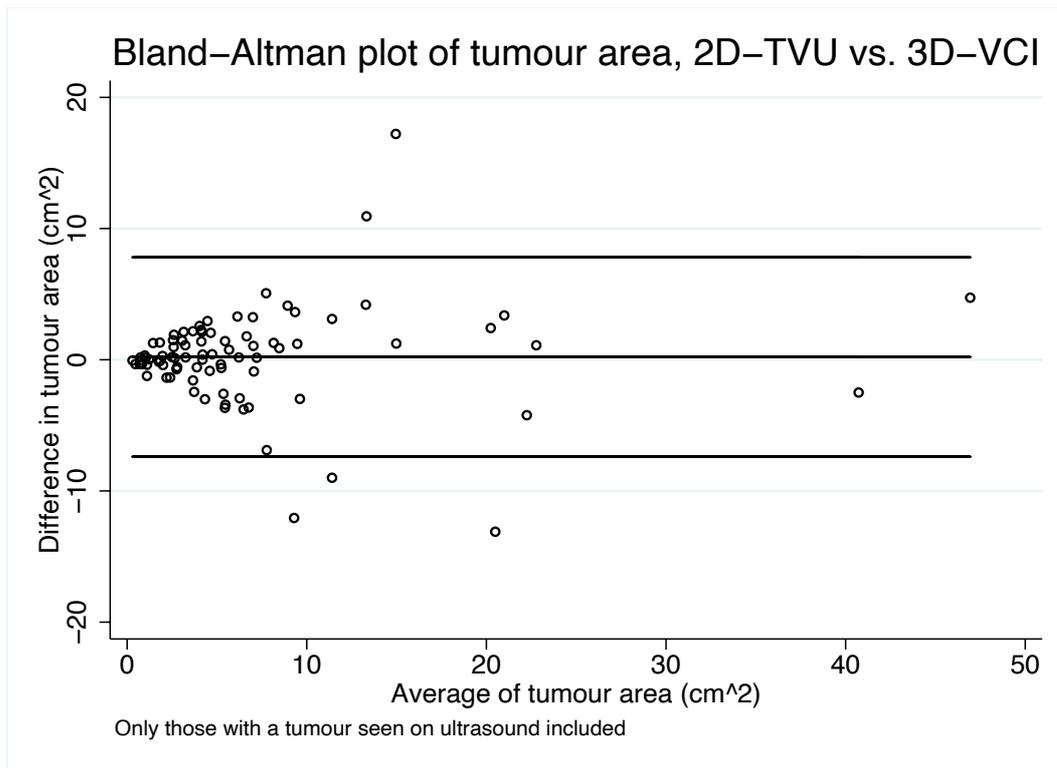


Figure 4: Bland-Altman plot of tumour area, measured by 2D-TVU and 3D-VCI. The mean difference was 0.21 cm², the limits of agreement ranging from -7.4 to 7.8.

Looking at Figure 4, we see that the mean difference is very close to 0 and that differences are randomly spread around the x -axis and that the differences in measurements are larger for larger tumours than for smaller tumours.

3.3.5 Diagnostic performance

The most commonly used measures of diagnostic performance of a diagnostic test are sensitivity, specificity and overall accuracy, which are calculated from a 2×2 table as shown in Table 6.

		Result of reference test/'gold standard'	
		Positive	Negative
Result of the diagnostic test	Positive	True positive	False positive
	Negative	False negative	True negative

Table 6. All possible outcomes of a diagnostic test compared to a reference test

$$\text{Sensitivity} = TP / (TP + FN)$$

$$\text{Specificity} = TN / (FP + TN)$$

$$\text{Accuracy} = (TP + TN) / (TP + FP + FN + TN)$$

Sensitivity and specificity are independent of the prevalence of the examined trait in the population being studied. In most clinical scenarios the dichotomization into a ‘positive’ or a ‘negative’ test is based on some cut-off value on a continuous variable, such as endometrial thickness, with a cut-off of 4 mm for endometrial sampling. Lowering this cut-off would categorize more women as ‘positive’, and fewer as ‘negative’, and more endometrial samplings would be performed. Among the new biopsies performed, most would be benign (increasing false positives), since they are sampled from a low-risk population, but a few would still be malignant (increasing true positives) that would otherwise have been missed. Since the actual prevalence of endometrial cancer (TP+FN), and thus also the number of benign cases (FP+TN), is constant, lowering the cut-off threshold only shifts some cases from false negatives to true positives, and from true negatives to false positives. This increases the numerator and decreases the denominator in our sensitivity calculation, while the opposite happens in our specificity calculation, thus increasing sensitivity at the expense of lowering specificity. This trade-off can be illustrated in a ROC-curve (Figure 5), where sensitivity is plotted against 1-specificity. In this example, a sensitivity of 0.5 would yield a 1-specificity of 0.25, thus a specificity of 0.75. A sensitivity of 0.88 would instead yield a 1-specificity of 0.75, thus a specificity of 0.25.

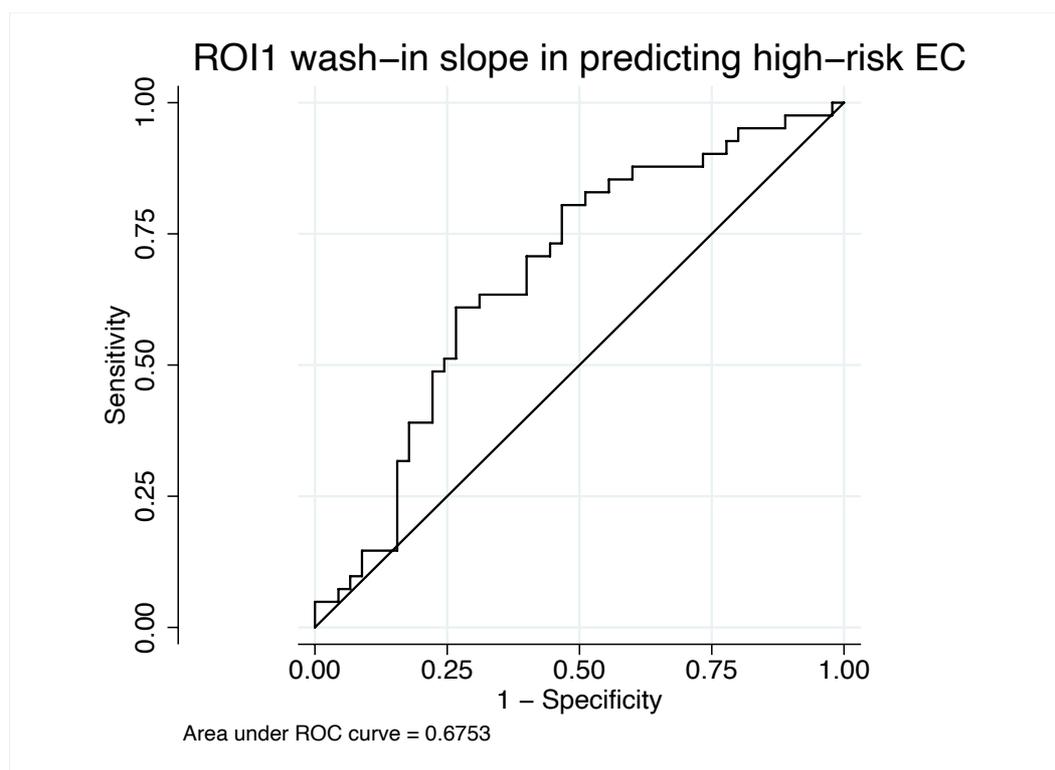


Figure 5: ROC-curve showing the ability of ROI1 wash-in slope to predict high-risk endometrial cancer. Adapted from study III, Table 4.

An area under the ROC (ROC-AUC) of 0.5 means that the sensitivity is a direct function of specificity (or *vice versa*) and that the diagnostic test thus has no discriminative ability. The best possible test has a ROC-AUC close to 1. What constitutes a satisfactory ROC-AUC depends on the situation; what is currently the best, what the consequences of misclassification are and what other information is available.

From Bayes theorem, the positive predictive value and the negative predictive value of a diagnostic test can be calculated using the prior probability (which is the prevalence of the trait in the intended population)²⁰⁷. Bayes theorem is stated as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

where $P(A|B)$ is the probability of A *given* B, $P(B|A)$ is the probability of B *given* A, $P(A)$ is the probability of A and $P(B)$ is the probability of B. In study III, we found that the sensitivity of DCE-US in diagnosing deep myometrial invasion was 74% and the specificity was 87%. Thus, the probability of a patient actually having deep myometrial invasion, given that deep myometrial invasion is diagnosed at DCE-US and that the prevalence of deep myometrial invasion is 40%¹⁵⁹ we get:

$$\begin{aligned} P(\text{deep MI} | \text{DCEUS}_{\text{deep MI}}) &= \frac{P(\text{DCEUS}_{\text{deep MI}} | \text{deep MI})P(\text{deep MI})}{P(\text{DCEUS}_{\text{deep MI}})} \\ &= \frac{P(\text{DCE}_{\text{deep MI}} | \text{deep MI})P(\text{deep MI})}{P(\text{DCEUS}_{\text{deep MI}} | \text{deep MI})P(\text{deep MI}) + P(\text{DCEUS}_{\text{deep MI}} | \text{superficial MI})P(\text{superficial MI})} \\ &= \frac{(0.74 * 0.40)}{(0.74 * 0.40) + (0.13 * 0.60)} = \frac{0.296}{0.296 + 0.078} = \frac{0.296}{0.374} = 0.79 = 79\% \end{aligned}$$

Thus, the probability that a patient actually has deep myometrial invasion, given that DCE-US diagnoses deep myometrial invasion, is 79%. This is not the same as the test accuracy, which in this case was 83%.

3.3.6 Regression analysis

Logistic regression is used to predict a dichotomous categorical variable based on one or many independent variables that can be continuous or categorical. This procedure fits a model where the independent variable is linearly related to the logarithm of the odds of the dependent variable. Thus, the main assumptions of logistic regression are that there is a linear relationship between the all continuous independent variables and the logit of the dependent variable and that the responses of the independent variable are mutually exclusive. The probability that our event of interest happens is p , and that it does not happen is $1-p$, and the odds is $p/(1-p)$. Thus, the model, for two independent variables, can be specified as;

$$\ln \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

and the probability of p happening can be recovered by:

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2)}}$$

where x_1 is the value of one variable, x_2 the value of the second variable, β_0 is the intercept were all independent variables are equal to 0, and β_1 and β_2 are the coefficients of the

independent variables. The model tries to estimate these beta-coefficients so that the fit of the model is optimal. In study III, logistic regression was used to predict FIGO stage \geq IB and high-risk group from DCE-US TIC variables. In study IV, logistic regression was used to predict high-risk group from BMI and waist circumference measurements, adjusted for age. The output of a logistic regression model with BMI and age in postmenopausal women on the odds of high-risk cancer can be seen in Table 7.

Covariate	Coefficient	95% confidence interval	<i>p</i>
BMI	-0.0412	-0.057 – -0.0254	<0.001
Age	0.0334	0.021 – 0.046	<0.001
Intercept	-0.8572	-1.850 – 0.136	0.091

Table 7: Results of logistic regression, reported as coefficients, with the independent variables treated as continuous variables.

Thus, the probability that a 52-year-old woman with a BMI of 23 will have high-risk cancer is:

$$p = \frac{1}{1 + e^{-(-0.8572 + (0.0334 * 52) + (-0.0412 * 23))}} = 0.48 = 48\%$$

while for a 78-year-old woman with a BMI of 34 it is:

$$p = \frac{1}{1 + e^{-(-0.8572 + (0.0334 * 78) + (-0.0412 * 34))}} = 0.58 = 58\%$$

3.3.7 Correlations

When two variables are related so that one can be predicted by knowledge of the other the variables are said to be correlated. This is not to say that this relation is causal, or what direction and magnitude this relation takes. The simplest relation is linear; as one changes so does the other by a certain factor. With Spearman's correlation coefficient (denoted ρ) a correlation of the actual values of the variables is not correlated, but their ranks. This allows investigating correlations that are monotonic, but not necessarily linear, such as sigmoidal or exponential. In study II, rater accuracy was correlated to their experience and we found that it was influenced by the number of cases assessed annually, but not years as a board-certified specialist in obstetrics and gynaecology. The formula for ρ is:

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

where i is the rank, averaged across ties. The table for rater accuracy and numbers assessed annually looks like this, when ranks are averaged across ties:

Rater	Accuracy	# of cases	Accuracy rank	Case rank	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
Rater 1	79.3%	40	10.5	11	2.5	3	7.5	6.25	9
Rater 2	82.8%	40	13	11	5	3	15	25	9
Rater 3	70.7%	15	4	1.5	-4	-6.5	26	16	42.25
Rater 4	84.5%	23	14.5	7	6.5	-1	-6.5	42.25	1
Rater 5	84.5%	50	14.5	13	6.5	5	32.5	42.25	25
Rater 6	75.9%	40	8.5	11	0.5	3	1.5	0.25	9
Rater 7	74.1%	60	6.5	14	-1.5	6	-9	2.25	36
Rater 8	65.5%	20	3	4.5	-5	-3.5	17.5	25	12.25
Rater 9	75.9%	80	8.5	15	0.5	7	3.5	0.25	49
Rater 10	79.3%	25	10.5	8	2.5	0	0	6.25	0
Rater 11	81.0%	36	12	9	4	1	4	16	1
Rater 12	72.4%	20	5	4.5	-3	-3.5	10.5	9	12.25
Rater 13	60.3%	20	2	4.5	-6	-3.5	21	36	12.25
Rater 14	58.6%	20	1	4.5	-7	-3.5	24.5	49	12.25
Rater 15	74.1%	15	6.5	1.5	-1.5	-6.5	9.75	2.25	42.25
							$\Sigma 157.75$	$\Sigma 278$	$\Sigma 272.5$

With these numbers we get a ρ equal to 0.573.

3.3.8 False discovery rate methods

All hypothesis testing procedures discussed above were designed to test only one hypothesis, for which the study was designed and powered. By fixing the type I error rate α at a certain level (most often 5%), we have accepted a certain risk that we falsely reject the null hypothesis. Indeed, even when working optimally and not taking various bias into account, it can be shown that most published research is likely false²⁰⁸. If we perform many hypothesis tests on the same dataset, we will thus increase the risk of having at least one false rejection of the null. To mitigate this, multiple methods have been developed, essentially setting a more conservative threshold for what we call “significant” when rejecting the null hypothesis. Since the significance level for a p -value of 0.05 was arbitrarily chosen by R. A. Fisher in 1926²⁰⁹, the most straightforward way to control the number of false rejections of the null hypothesis is simply to choose a lower value, but this will come at the expense of increasing the risk of making a type II error. Thus, the challenge is to control the false rejection rate with as little negative effect on the type II error rate (and thus statistical power) as possible. Most methods strike this balance by adjusting the significance level (or p -values) by relating it to the number of tested hypotheses. The Bonferroni method is one of the most common, where the obtained p -

values are simply multiplied by the number of hypotheses. This ensures that in a situation where all null hypotheses are actually true (and thus all rejections are false) the probability of committing even a single type I error (called *family-wise error rate: FWER*) is never more than α ²¹⁰. All methods that control the FWER are very conservative, as they maximally reduce the risk of making even a single type I error, and as such have low statistical power. In most research settings we are content with getting a few false rejections of the null since we might have planned for confirmatory experiments or validating our results in a new dataset. Thus, we can instead decide on an acceptable ratio of false to true rejections of the null, the *false discovery rate- FDR*. The Benjamini-Hochberg method²¹¹ is the most common FDR method, and the original paper describing the method is among the top 100 cited papers of all times²¹². The Benjamini-Hochberg method adjusts the estimated p -value by relating it to the total number of p -values and the rank of the p -value and has much higher power than the Bonferroni method²¹³. From these fundamental methods, both FWER and FDR methods have developed for special applications and circumstances to further increase statistical power at the same probability of a type I error. One such is the Benjamini-Yekutieli method that takes into account dependency between hypotheses²¹⁴. The Benjamini-Yekutieli method was used in study IV since many variables were compared between pre- and postmenopausal women, and we expected many of these variables to be related.

4 RESULTS

4.1 STUDY I

In survivors from endometrial cancer, the median follow-up time was 8.9 years. A total of 105 women died from endometrial cancer (9.2%). In univariate Cox proportional hazard regression, age, FIGO stage, degree of differentiation, S-phase fraction and DNA ploidy were all significantly associated with a higher risk of death in endometrial cancer. Adjuvant therapy was not. When all variables were entered in the same model, age, FIGO stage, degree of differentiation and S-phase fraction retained their prognostic information, while DNA ploidy was no longer associated with a worse outcome. A likelihood-ratio test, comparing a model with adjuvant therapy included as a predictor variable, compared to a model without adjuvant therapy included, was not significant. When running the main multivariate model with time-to-progression and overall survival as end-points, DNA ploidy lost prognostic significance while S-phase fraction retained its predictive ability. When plotting the cumulative incidence, with death from other causes as competing risks, diploid tumours with low S-phase fraction had a cumulative incidence of death of endometrial cancer of less than 5% over the study period.

4.2 STUDY II

The median intrarater overall agreement between 2D-TVU and 3D-VCI among the 15 raters was 76% (range 64-93%) in diagnosing deep myometrial invasion, and 88% (range 79-97%) in diagnosing cervical stromal involvement. The negative agreement was generally higher than the positive agreement for both deep myometrial invasion and cervical stromal involvement. The interrater reliability, measured using Fleiss' kappa, was 0.41 in diagnosing deep MI with 2D-TVU and 0.31 with 3D-VCI. In diagnosing cervical stromal involvement, the corresponding figures were 0.55 and 0.45. With 2D-TVU, the mean sensitivity in diagnosing deep myometrial invasion was 72% and the mean specificity 76%. With 3D-VCI, the corresponding figures were 70% and 66%. In diagnosing cervical stromal involvement, the mean sensitivity with 2D-TVU was 64% and the mean specificity 92%. For 3D-VCI, the corresponding figures were 56% and 90%. When rater experience measures were correlated to overall accuracy, we found that only the number of cases assessed annually, but not the number of years as a board-certified specialist in obstetrics and gynaecology, years' experience as an expert sonographer or years' experience in endometrial cancer staging, were positively correlated to overall accuracy.

4.3 STUDY III

2D-TVU, complemented with DCE-US, had higher sensitivity in diagnosing deep myometrial invasion than 2D-TVU alone (74% vs. 62%). The sensitivity in diagnosing cervical stromal involvement was also higher with DCE-US than 2D-TVU alone (75% vs. 53%). This improvement in sensitivity came without loss of specificity, which was equally good with or without DCE-US (87% and 85% in diagnosing deep myometrial invasion, 96%

and 95% in detecting cervical stromal involvement). A DCE-US focal filling pattern and a wash-in pattern 'prior' were more common in high-risk cancers compared to low-risk cancers. All quantitative TIC parameters (wash-in slope, time-to-peak, peak intensity and area-under-the-TIC) measured at ROI1 could significantly predict FIGO stage \geq IB, but no TIC parameter held up to the pre-specified cut-off value of a ROC-AUC of >0.7 . We could also confirm previous findings that a non-uniform echogenicity, higher colour score, and multiple, multifocal vessel pattern on conventional 2D-TVU was associated with high-risk cancer.

4.4 STUDY IV

Among background variables, premenopausal women had lower parity, more often a family history of colon cancer, lower waist circumference and a longer duration of abnormal bleeding, compared to postmenopausal women. Among sonographic variables, premenopausal women less often had a defined tumour, adenomyosis was more common, the endometrial-myometrial border was more often defined, and the endometrial midline was more often seen. Premenopausal women had lower FIGO surgical stage, more often endometrioid tumour and lower grade, i.e. more often low-risk tumours. Despite this low-risk presentation of endometrial cancer among premenopausal women, the vascularity measures colour score and vascular pattern did not differ between pre- and postmenopausal women. Both sensitivity and specificity in diagnosing deep myometrial invasion were higher in premenopausal women, compared to postmenopausal women. In diagnosing cervical stromal involvement there was no difference between the groups. A high BMI and high waist circumference were associated with low-risk group classification among postmenopausal women, but not among premenopausal women.

5 DISCUSSION

All studies included in this thesis used different patient material, different statistic methods and had different study design, and are thus prone to very different strengths and weaknesses.

The greatest strength of study I was the use of a large population-based material, stratified for histology and FIGO stage I. This limits selection bias and increases external validity to this large patient group, but we should be cautious to extrapolate results to more advanced stage cancer or non-endometrioid histologies. Despite including more patients than any similar study before, the statistical power of survival analysis is not driven by population size, but the number of events. We had conducted a power analysis based on the results from a smaller pilot study, but the end-point in the power analysis was overall survival, not endometrial cancer-specific survival. As more low-risk endometrial cancer patients are likely to die from other causes than endometrial cancer, this might have underpowered this study. Yet, underpowering increases the risk of making a type II error, not a type I error, why the prognostic value of S-phase fraction is likely a true finding. Increasing the cut-off value for S-phase fraction will also likely increase the effect size, making the effect easier to detect even with fewer events.

The greatest strength of study II was the large number of cases and raters, as a larger sample from any population is more likely to correctly mirror the population on which we are making inferences. Most studies on diagnostic accuracy and reliability only use a handful of raters, we included fifteen. Yet, all were considered second-opinion expert sonographers, recruited from the IETA collaboration, thus the estimated performance is likely higher than for general gynaecologists, which has also been shown previously⁹⁷. Off-line analysis was used, which is likely to have poorer diagnostic performance than live examination. Still, the measured performance was high, and for 3D-VCI the rendered volume is always examined and manipulated off-line post-acquisition, which is one of the benefits of the technique. Thus, the results of 3D-VCI are more easily generalizable than the results from 2D-TVU.

The greatest strength of study III is the use of a matched cohort design, which eliminates bias due to lack of blinding that would otherwise occur. The greatest weakness was that we only had one examiner, which limits external validity. By copying the study design from study II, it would be possible to assess if multiple raters would get the same high diagnostic accuracy with DCE-US when examining the off-line video clips and assess their interrater reliability. We could only identify one previous study assessing the diagnostic performance of DCE-US in assessing local tumour extension¹⁰⁹. That study did not compare DCE-US to 2D-TVU, so we do not know if the addition of DCE-US was beneficial from that study alone. Thus, we are the first to show an added benefit of DCE-US to the routine 2D-TVU examination, which has to be reproduced by others before it can be introduced in clinical practice.

The greatest strength of study IV is the large study population. The full IETA database was designed to include enough cases to also perform subgroup analysis on rare presentations, such as premenopausal endometrial cancer. Still, this subgroup was only 161 women, and in

study IV we tested forty-five hypotheses (excluding our primary aim of comparing diagnostic performance between pre- and postmenopausal women). Indeed, with forty-five hypotheses and an α -level of 0.05, the risk of making at least one type I error is 90% ($1-0.95^{45}$). After adjusting the significance level to 0.004 by the Benjamini-Yekutieli procedure, the risk dropped to 16.5% ($1-0.996^{45}$). Thus, this study should be considered exploratory and results need to be confirmed in other studies where our estimates can be used in power analysis when designing such studies. In previous studies, obesity has shown to increase the risk for low-risk endometrial cancer more than for high-risk endometrial cancer^{9,29,215}. In our study, we showed that obesity lowered the risk of high-risk endometrial cancer in postmenopausal women, but not in premenopausal women. Thus, this result could be a statistical artefact from unmeasured competing risks. If an obese woman has low-risk cancer early in life, and thus a hysterectomy, she is no longer under risk of later developing high-risk cancer. A non-obese woman will keep her uterus long enough to develop high-risk endometrial cancer since high age is the greatest risk factor for non-endometrioid endometrial cancer. To fully explore the association between obesity and endometrial cancer risk subtypes and outcomes, a prospective cohort study with healthy women at baseline will have to be performed. Our study design in study IV cannot circumvent this design flaw.

5.1 RESULTS IN A CLINICAL CONTEXT

The results from study I have already changed clinical management. The study is featured in the Swedish National Clinical Guidelines for Endometrial Cancer¹⁵⁴, and after its inclusion (among other studies), DNA ploidy status was removed as a preoperative qualifying factor for pelvic lymph node sampling. The results from study II shows that expert ultrasound has a high diagnostic performance and interrater reliability, well in line with what has been reported for MRI. This study also shows that 2D video clips are preferred to 3D volumes in off-line examination. This has implications when deciding on diagnostic method and designing a clinical workflow, especially in situations where the examination is performed by a technician and later interpreted by a clinician. Based in part on findings from this study, new international guidelines will recommend that this examination is centralised to high-throughput centres. Study III showed that the addition of DCE-US to the routine 2D-TVU examination can improve sensitivity in diagnosing deep myometrial invasion, without lowering specificity. A negative result is thus especially useful for ruling out deep myometrial invasion, which is especially important in cases where fertility-sparing treatment is considered. Adding DCE-US can be done without any prior laboratory testing or preparations, why it can be a useful addition in unclear or borderline cases or with contradictory, mixed features. Likewise, study IV showed that the diagnostic performance of 2D-TVU is even better in premenopausal women, which can affect the choice of diagnostic method when fertility-sparing treatment is possible. Study IV also showed that some sonographic findings considered high-risk features could be found in otherwise low-risk premenopausal women, likely due to physiological features of a cycling endometrium. Thus, this finding has an impact on how these sonographic features should be interpreted with consideration to menopausal status, further tailoring management to each patient.

5.2 METHODOLOGICAL CONSIDERATIONS

A bias is a systematic error that may or may not change the interpretation of a result. Any study is more or less prone to bias and careful methodological considerations have to be taken to reduce or mitigate the effect of such biases.

5.2.1 Selection bias

Selection bias occurs when the distribution of included variables in the study population does not represent the distribution of these variables in the target population for which we are making inferences. In study I, selection bias was reduced to a minimum by including a whole population in a geographic region during the study period. In study II the prevalence of high-risk cancer was set to match the prevalence of high-risk cancer in patients referred to tertiary centres since kappa statistics are dependent on the trait prevalence. Thus, this figure is likely lower in the general endometrial cancer population were the prevalence of high-risk cancer is lower. Although, no selection was made on image quality when designing the study population since this would likely result in an overestimation of diagnostic performance and reliability. In study III, women were recruited consecutively to reduce selection bias, but no records were kept on those that declined participation. It is unlikely that non-participation is related to the studied outcomes and close to all women accepted inclusion. In study IV, the full IETA4 database was used. This database was generated from a collaboration of multiple tertiary centres. In all studies using patients from tertiary centres, we may have introduced a selection bias at the level of referral, since more complex cases and more patients that can benefit from extended workup will be included. Although, this possible selection bias is of little importance since we are trying to improve clinical management in cases where there are scientific doubts on the best management. Low-risk cases managed at secondary centres already have an excellent prognosis, and those too sick to benefit from extended workup and subsequent treatment should be offered palliative best supportive care, thus the question of what is the best sonographic workup in those patients becomes redundant.

5.2.2 External validity

External validity relates to what extent scientific findings can be generalized to the target population. If selection bias is present, this limits the external validity, making results less generalizable. Study I was restricted to FIGO stage I endometrioid cancers only, since previous studies were plagued by a plethora of different inclusion and exclusion criteria, making generalisations of results impossible. Thus, the results from study I can be extended to this patient group, but not necessarily more advanced stages or non-endometrioid histologies. Study II used an off-line material, which is not necessary as reliable as live examination. 3D volumes are designed to be analysed off-line, so the lack of live examination might have lowered the diagnostic performance for 2D-TVU more than for 3D-VCI. Still, 2D-TVU performed better, so it is unlikely that a design with live examination would have changed the result of the study even if it ever could have been performed. Study II also included only expert sonographers and the results might thus not be generalizable to general gynaecologists. Considering how most gynaecologists are familiar with 2D video clips and

that 3D-VCI are a non-routine practice, the discrepancy between 2D-TVU and 3D-VCI is likely even larger among general gynaecologist, in favour of 2D-TVU. In study III, one examiner (E.E) performed all examinations. DCE-US does not have any routine gynaecological application, why this examination is experimental. E.E. has vast experience with gynaecologic sonography and had performed a pilot study using DCE-US on cervical cancer patients prior to study III. Thus, we do not know how well DCE-US performs in the hand of less experienced examiners and we do not know how steep the learning curve is. Therefore, results might not be generalizable to all gynaecologists. The IETA collaboration included 17 European tertiary centres specifically to have high external validity, why the results of study IV are highly generalizable to other tertiary centres, at least in high-income countries. Considering differences in age, parity and obesity distributions in low- and middle-income countries compared to high-income countries, results might not be generalizable to such countries.

5.2.3 Internal validity

Internal validity relates to what extent a dependent variable is causally related to a set of independent variables, and not just spuriously related by coincidence. Internal validity is not only a statistical issue, but knowledge of underlying biological mechanisms and familiarity with previously published literature can make a causal relationship more or less plausible. The highest internal validity is achieved in highly controlled experiments with very few unmeasured variables, allowing manipulation of one variable at a time. Although, such experiments might not result in high external validity due to the synthetic nature of the experiment. In all included studies in this thesis, some causal relationship is inferred; in study I between S-phase fraction and death of endometrial cancer, in study II between rater experience and diagnostic accuracy, in study III between higher capillary blood flow and high-risk cancer and in study IV between obesity and high-risk cancer, to list a few. All these relationships seem plausible from a mechanistic perspective but might not tell the whole truth. Neither study assesses the goodness-of-fit of the presented model nor reports on the unexplained variance of the models. Thus, the model might be misspecified or there could be unmeasured variables that confound the causal relationship.

5.2.4 Confounders, mediators, moderators and covariates

A confounder is a variable that is related to both the dependent variable and the independent variable in a, presumed, causal relationship. In study I, a poor degree of differentiation is likely related to both higher S-phase fraction and risk of death in endometrial cancer. A mediator is similar to a confounder, only that the direction of the effect between the independent variable and the mediator is reversed compared to a confounder. In study IV, the effect of BMI on the risk of high-risk cancer might have been mediated through increased (unmeasured) oestrogen levels. A moderator is a variable with an interaction with the independent variable, but not the dependent variable, that still affects the relationship between the dependent and the independent variable. A covariate is a variable with effect on the dependent variable, but not with any other variables. A *directed acyclic graph* (DAG) can

help to visualise confounders, mediators, moderators and covariates²¹⁶. The choice on what constitutes each type of variable is not only a statistical question but has to be based on knowledge of underlying biology, pathophysiology and previously published literature. Ideally, the model should be specified *a priori* from such knowledge since model building can have a huge impact on the interpretation of the causal relationship. Figure 6 and 7 illustrates how seemingly simple model building quickly becomes more complex as more variables are included and the dependency structure between variables are unclear.

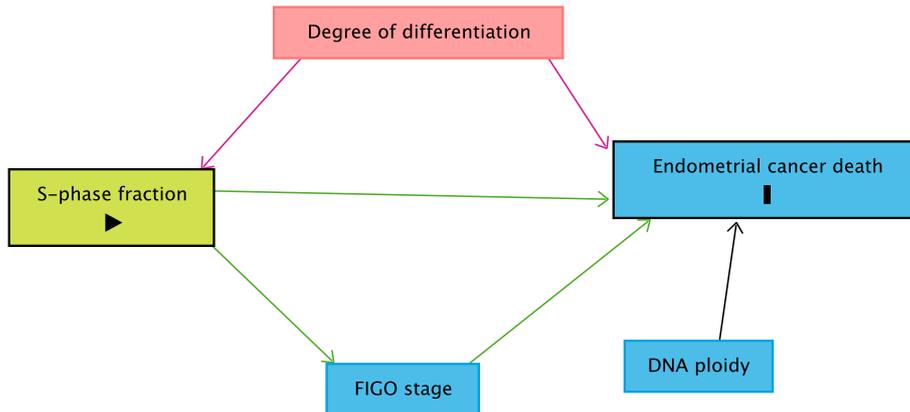


Figure 6: A simple DAG showing the causal relationship between S-phase fraction and endometrial cancer death, where degree of differentiation is a confounder, FIGO stage a mediator and DNA ploidy a covariate.

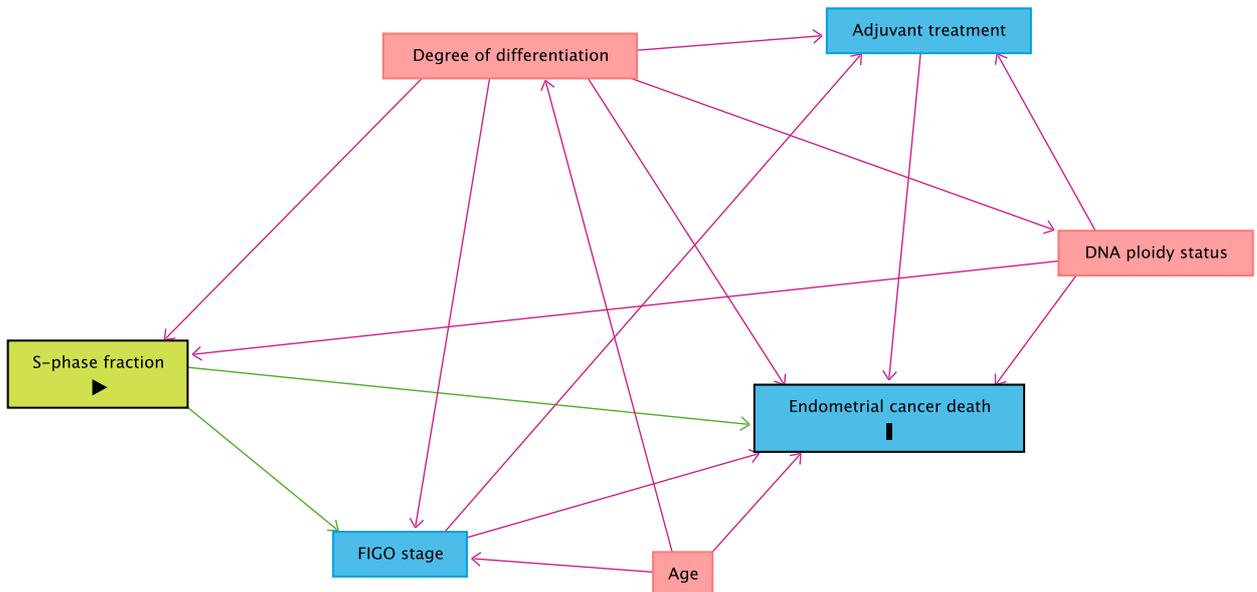


Figure 7: A more complex DAG where variables are dependent, and the direction of effects are not set in stone.

6 CONCLUSIONS

S-phase fraction is an independent prognostic factor in FIGO stage I endometrioid endometrial cancer, when adjusting for age, myometrial invasion, degree of differentiation, DNA ploidy and adjuvant therapy. DNA ploidy status did not retain prognostic value after adjusting for these covariates. A cut-off level of 5.5% for S-phase fraction gave an adjusted hazard ratio of 2.3, but the optimal cut-off value is unknown and has to be sought in a new patient material.

Interrater reliability was higher for 2D-TVU than for 3D-VCI and as high as been reported for MRI in similarly designed studies. Diagnostic accuracy is correlated to the number of endometrial cancer cases assessed annually, why this examination ideally should be centralized to high-throughput tertiary centres. For off-line assessment, 2D-TVU video clips are preferred to 3D-VCI volumes, which can be useful if second-opinion expertise is needed post-acquisition.

DCE-US improves diagnostic accuracy in subjective tumour extension assessment. This technique is already in use clinically for assessment of liver lesions, with a well-known safety profile, why DCE-US could be introduced in endometrial cancer assessment swiftly. DCE-US does not require any special preparations such as kidney function tests and could thus be used as a complement to the routine 2D-TVU examination in unclear cases, given that the ultrasound system supports this technique.

2D-TVU is even more accurate in premenopausal women, compared to postmenopausal women, which can impact the management of women wishing for fertility-sparing treatment. Tumour vascularity and endometrial thickness have to be interpreted with caution in premenopausal women, considering physiological changes in cycling endometrium, but an intact endometrial-myometrial border is a good sign in both pre- and postmenopausal women. A high BMI and a high waist circumference affected pre- and postmenopausal women differently, but this could be a statistical artefact caused by unmeasured competing risks.

7 FUTURE PERSPECTIVES

Now when the independent prognostic information for S-phase fraction has been identified, the optimal cut-off value has to be sought. First, a large dataset has to be compiled, then split into a *training set*, in which the optimal cut-off can be found, and a *validation set*, to validate the model. By using modern non-parametric modelling, no assumptions about the shape of the hazard function of S-phase fraction has to be made, greatly improving the chances of finding an optimal cut-off. This work has already been started, by engaging colleagues in Umeå, Göteborg and Lund.

To be clinically useful, the utility of DCE-US in endometrial cancer assessment has to be validated in other centres, and the interrater reliability has to be assessed. Any diagnostic tool intended to be used in everyday clinical practice must be robust enough to produce reproducible results in multiple centres and across multiple examiners. If antiangiogenic drugs are introduced in the treatment of endometrial cancer, DCE-US contrast flow quantification could possibly be used as an early marker of therapeutic response, selecting patients with proven benefit to continued treatment, and sparing patients with no therapeutic effect the side-effects of the treatment.

The effect of a high BMI and high waist circumference on the risk of developing low- and high-risk endometrial cancer in pre- and postmenopausal women is best studied using a prospective cohort where all are healthy at baseline. Competing risks can thus be measured during the follow-up period and adjusted for. We have a unique opportunity to conduct such a study in Sweden, considering our well-managed national patient records.

8 ACKNOWLEDGEMENTS

This thesis would not have been made possible if it wasn't for the contributions of some very special people.

First, the self-less patients who volunteered to participate. I hope I've made the most to honour your sacrifice of time and discomfort to better the management of patients coming after you.

I also especially want to thank;

Elisabeth Epstein, my principal supervisor. You saw talent in me and included me in your team, despite all your other obligations and have always made me feel qualified beyond my experience. You have, with your husband **Pelle Lindqvist**, whom I also want to thank for all rewarding scientific discussions (and the moose meat), shown that you can really have it all; a beautiful home and family, hobbies, travels and a flourishing clinical and scientific career.

My co-supervisor **Bo Baldetorp**, for introducing me to academia and opening the door to scientific work. At your retirement party, all witnessed to your everlasting scientific curiosity and your way of bringing the best out of all you work with, lifting them and honing their talents, and I would be the first to agree.

Maria Bjurberg, Anna Måsbäck and Linda Hartman; it has been wonderful to work with you. You've been so generous with your knowledge and have always made me feel like a peer, already as a master student. Humouring me as I've gradually figured things out, I've would not have the knowledge I do today without you.

Sanna Engblom, though thick and thin, ups and downs, we worked our way through the piles of patient records, not giving up or settling for less than our best.

My second co-supervisor **Joseph Carlson** for always being able to answer anything pathology related and for always being so available in sharing your knowledge.

My mentor **Folke Flam** for allowing me into your operating theatre and showing that it is OK to bring a little personality into your professional self.

My co-authors **Juan Luis Alcazar, Valentina Chiappa, Balazs Erdodi, Dorella Franchi, Filip Frühauf, Robert Fruscio, Stefano Guerriero, Betlem Graupera, Atilla Jacob, Alessia di Legge, Manuela Ludovisi, Floriana Mascilini, Maria Angela Pascual, Thierry van den Bosch, Daniela Fischerova, Antonia Testa, David Cibula, Lucia A. Haak, Gina Opolskiene, Ailyn Vidal Urbinati, Dirk Timmerman and Tom Bourne** for your invaluable contributions to study II and IV. Especially thanks to **Lil Valentin** for your devotion to the scientific method and its continuing perfection.

Jan Kottner, John Uebersax, Sebastian Döhler, Roger B. Newson and Ruth Heller for helpful correspondence regarding your respective fields of expertise. Academia is truly amazing when experts like yourself are so accessible and generous with your knowledge.

My PhD-siblings **Kolbrún Pálsdóttir** and **Linda Eriksson**, for paving the way and letting me learn from your experiences in navigating PhD bureaucracy and methodological pitfalls and caveats. I hope to continue following in your footsteps and learn from you.

Kristina Gemzell Danielsson, **Lalit Kumar** and **Sofie Graner** for providing my first home at Karolinska Institutet and Karolinska University Hospital, without which my life in Stockholm would have been very different. And to my former colleagues at the Gynaecology and Obstetrics Department, I hope we meet again!

My mother **Monica Walterdotter** and my father **Anders Green**, for encouraging my scientific curiosity since I could first speak and giving me the confidence to try to achieve what my heart has set on.

My friends **Christian**, **Alex**, **Jesper**, **Kristoffer**, **James** and **Hillevi**. I get by with a little help from my friends.

To **Suzanna Queckbörner** for your great support in the lab and helping me realise that a lab-heavy doctoral program was not for me.

To **Tove Fredriksson**, **Gabriella Smedfors**, **Manne Holm**, **Carl Haggård**, **Marcus Dahlquist** and **Oskar Kövamees**, for sharing mutual experiences through this PhD journey and showing that I wasn't alone.

And finally, to **Tilda**, my partner, party-planner and soon to be the mother of our first-born child. I love you!

9 ERRATA

Here I report on errors found in the included studies after acceptance and publication.

In the Statistics section of **Study III**, it is stated that “positive and negative predictive values were calculated as recommended by Altman and Bland, derived from Bayes theorem”, but this calculation is not reported in the Results section. In the study cohort (2D-TVU+DCE-US) the sensitivity in diagnosing deep myometrial invasion was 74% and the specificity 87% (Table 2). Given 40% prevalence of deep myometrial invasion¹⁵⁹, the positive predictive value is 79% and the negative predictive value is 83%. In the control cohort (2D-TVU only), the corresponding figures are 73% and 77%. In detecting cervical stromal involvement, the sensitivity in the study cohort was 75% and the specificity 96% (Table 2). Given 10% prevalence of cervical stromal involvement⁹⁶, the positive predictive value is 68% and the negative predictive value is 97%. In the control cohort, the corresponding figures was 53% and 95%. This has been amended with the editors and will be changed in the final print version of the paper.

10 REFERENCES

1. Amant F, Mirza MR, Koskas M, Creutzberg CL. Cancer of the corpus uteri. *Int J Gynecol Obstet*. 2015;131(December):S96-S104. doi:10.1016/j.ijgo.2015.06.005
2. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015;136(5):E359-E386. doi:10.1002/ijc.29210
3. Lortet-Tieulent J, Ferlay J, Bray F, Jemal A. International patterns and trends in endometrial cancer incidence, 1978-2013. *J Natl Cancer Inst*. 2018;110(4):354-361. doi:10.1093/jnci/djx214
4. Ng M, Fleming T, Robinson M, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet*. 2014;384(9945):766-781. doi:10.1016/S0140-6736(14)60460-8
5. Creasman W, Odicino F, Maisonneuve P, et al. Carcinoma of the Corpus Uteri. *Int J Gynecol Obstet*. 2006;95:S105-S143. doi:10.1016/S0020-7292(06)60031-3
6. Cancer Stat Facts: Endometrial Cancer. National Cancer Institute (NIH). <https://seer.cancer.gov/statfacts/html/corp.html>. Accessed October 30, 2019.
7. Cancer Research UK. <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/uterine-cancer/incidence#heading-One>. Accessed October 30, 2019.
8. Socialstyrelsen, Statistikdatabas för cancer. https://sdb.socialstyrelsen.se/if_can/val.aspx. Accessed October 30, 2019.
9. Nevadunsky NS, Van Arsdale A, Strickler HD, et al. Obesity and age at diagnosis of endometrial cancer. *Obstet Gynecol*. 2014;124(2 PART1):300-306. doi:10.1097/AOG.0000000000000381
10. Reeves GK, Pirie K, Beral V, Green J, Spencer E, Bull D. Cancer incidence and mortality in relation to body mass index in the Million Women Study: Cohort study. *Br Med J*. 2007;335(7630):1134-1139. doi:10.1136/bmj.39367.495995.AE
11. Hernandez A V., Pasupuleti V, Benites-Zapata VA, Thota P, Deshpande A, Perez-Lopez FR. Insulin resistance and endometrial cancer risk: A systematic review and meta-analysis. *Eur J Cancer*. 2015;51(18):2747-2758. doi:10.1016/j.ejca.2015.08.031
12. ACOG. Tamoxifen use and uterine cancer. *Obs Gynecol*. 2014;(123):1394-1397.
13. Barry JA, Azizia MM, Hardiman PJ. Risk of endometrial, ovarian and breast cancer in women with polycystic ovary syndrome: A systematic review and meta-analysis. *Hum Reprod Update*. 2014;20(5):748-758. doi:10.1093/humupd/dmu012
14. Furness S, Roberts H, Marjoribanks J, Lethaby A. Hormone therapy in postmenopausal women and risk of endometrial hyperplasia. *Cochrane Database Syst Rev*. 2012;(8). doi:10.1002/14651858.CD000402.pub4
15. Epstein E, Lindqvist PG, Olsson H. A population-based cohort study on the use of hormone treatment and endometrial cancer in southern Sweden. *Int J Cancer*. 2009;125(2):421-425. doi:10.1002/ijc.24284
16. Bonadona V, Bonaïti B, Olschwang S, et al. Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome. *JAMA*. 2011;305(22):2304-2310. doi:10.1001/jama.2011.743
17. McPherson CP, Sellers TA, Potter JD, Bostick RM, Folsom AR. Reproductive factors and risk of endometrial cancer. The Iowa Women's Health Study. *Am J Epidemiol*. 1996;143(12):1195-1202. doi:10.1093/oxfordjournals.aje.a008707
18. Dossus L, Allen N, Kaaks R, et al. Reproductive risk factors and endometrial cancer: The European prospective investigation into cancer and nutrition. *Int J Cancer*. 2010;127(2):442-451. doi:10.1002/ijc.25050
19. Ryan NAJ, Glaire MA, Blake D, Cabrera-Dandy M, Evans DG, Crosbie EJ. The proportion of

- endometrial cancers associated with Lynch syndrome: a systematic review of the literature and meta-analysis. *Genet Med*. 2019;21(10):2167-2180. doi:10.1038/s41436-019-0536-8
20. Michels KA, Pfeiffer RM, Brinton LA, Trabert B. Modification of the associations between duration of oral contraceptive use and ovarian, endometrial, breast, and colorectal cancers. *JAMA Oncol*. 2018;4(4):516-521. doi:10.1001/jamaoncol.2017.4942
 21. Jareid M, Thalabard JC, Aarflot M, Bøvelstad HM, Lund E, Braaten T. Levonorgestrel-releasing intrauterine system use is associated with a decreased risk of ovarian and endometrial cancer, without increased risk of breast cancer. Results from the NOWAC Study. *Gynecol Oncol*. 2018;149(1):127-132. doi:10.1016/j.ygyno.2018.02.006
 22. Felix AS, Gaudet MM, Vecchia C La, et al. Intrauterine devices and endometrial cancer risk: A pooled analysis of the Epidemiology of Endometrial Cancer Consortium. *Int J Cancer*. 2015;136(5):E410-E422. doi:10.1002/ijc.29229
 23. Goh JY, He S, Allen JC, Malhotra R, Tan TC. Maternal obesity is associated with a low serum progesterone level in early pregnancy. *Horm Mol Biol Clin Investig*. 2016;27(3):97-100. doi:10.1515/hmbci-2015-0030
 24. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet*. 2008;371(9612):569-578. doi:10.1016/S0140-6736(08)60269-X
 25. Calle EE, Kaaks R. Overweight, obesity and cancer: Epidemiological evidence and proposed mechanisms. *Nat Rev Cancer*. 2004;4(8):579-591. doi:10.1038/nrc1408
 26. Soliman PT, Oh JC, Schmeler KM, et al. Risk Factors for Young Premenopausal Women With Endometrial Cancer. *Obstet Gynecol*. 2005;105(3):575-580. doi:10.1097/01.AOG.0000154151.14516.f7
 27. Wise MR, Jordan V, Lagas A, et al. Obesity and endometrial hyperplasia and cancer in premenopausal women: A systematic review. *Am J Obstet Gynecol*. 2016;214(6):689.e1-689.e17. doi:10.1016/j.ajog.2016.01.175
 28. Schmeler KM, Soliman PT, Sun CC, Slomovitz BM, Gershenson DM, Lu KH. Endometrial cancer in young, normal-weight women. *Gynecol Oncol*. 2005;99(2):388-392. doi:10.1016/j.ygyno.2005.06.029
 29. Yang HP, Wentzensen N, Trabert B, et al. Endometrial cancer risk factors by 2 main histologic subtypes. *Am J Epidemiol*. 2013;177(2):142-151. doi:10.1093/aje/kws200
 30. Bokhman J V. Two pathogenetic types of endometrial carcinoma. *Gynecol Oncol*. 1983;15(1):10-17. doi:10.1016/0090-8258(83)90111-7
 31. Kurman RJ, Carcangiu ML, Herrington SC, Young RH. *WHO Classification of Tumours of Female Reproductive Organs*. 4th ed. International Agency for Research on Cancer; 2014.
 32. Colombo N, Creutzberg C, Amant F, et al. ESMO-ESGO-ESTRO Consensus Conference on Endometrial Cancer: diagnosis, treatment and follow-up. *Ann Oncol*. 2016;27(1):16-41. doi:10.1093/annonc/mdv484
 33. O'Hara A, Bell D. The genomics and genetics of endometrial cancer. *Adv Genomics Genet*. 2012;2012(2):33-47. doi:10.2147/AGG.S28953.The
 34. Cheung LWT, Hennessy BT, Li J, et al. High frequency of PIK3R1 and PIK3R2 mutations in endometrial cancer elucidates a novel mechanism for regulation of PTEN protein stability. *Cancer Discov*. 2011;1(2):170-185. doi:10.1158/2159-8290.CD-11-0039
 35. Rudd ML, Price JC, Fogoros S, et al. A unique spectrum of somatic PIK3CA (p110 α) mutations within primary endometrial carcinomas. *Clin Cancer Res*. 2011;17(6):1331-1340. doi:10.1158/1078-0432.CCR-10-0540
 36. Kandoth C, Schultz N, Cherniack AD, et al. Integrated genomic characterization of endometrial carcinoma. *Nature*. 2013;497(7447):67-73. doi:10.1038/nature12113
 37. Bell DW, Ellenson LH. Molecular Genetics of Endometrial Carcinoma. *Annu Rev Pathol Mech Dis*.

- 2019;14(1):339-367. doi:10.1146/annurev-pathol-020117-043609
38. Byron SA, Gartside M, Powell MA, et al. Fgfr2 point mutations in 466 endometrioid endometrial tumors: Relationship with msi, kras, pik3ca, cttnb1 mutations and clinicopathological features. *PLoS One*. 2012;7(2):1-9. doi:10.1371/journal.pone.0030801
 39. Jeong W-J, Ro EJ, Choi K-Y. Interaction between Wnt/ β -catenin and RAS-ERK pathways and an anti-cancer strategy via degradations of β -catenin and RAS by targeting the Wnt/ β -catenin pathway. *npj Precis Oncol*. 2018;2(1):1-10. doi:10.1038/s41698-018-0049-y
 40. Lax SF, Kendall B, Tashiro H, Slebos RJC, Ellenson LH. The frequency of p53, K-ras mutations, and microsatellite instability differs in uterine endometrioid and serous carcinoma: Evidence of distinct molecular genetic pathways. *Cancer*. 2000;88(4):814-824. doi:10.1002/(SICI)1097-0142(20000215)88:4<814::AID-CNCR12>3.0.CO;2-U
 41. Talhouk A, McAlpine JN. New classification of endometrial cancers: the development and potential applications of genomic-based classification in research and clinical care. *Gynecol Oncol Res Pract*. 2016;3(1):14. doi:10.1186/s40661-016-0035-4
 42. Umar A, Boland CR, Terdiman JP, et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst*. 2004;96(4):261-268. doi:10.1016/j.atherosclerosis.2009.05.009.Effect
 43. McAlpine J, Leon-Castillo A, Bosse T. The rise of a novel classification system for endometrial carcinoma; integration of molecular subclasses. *J Pathol*. 2018;244(5):538-549. doi:10.1002/path.5034
 44. Beroukhi R, Mermel CH, Porter D, et al. The landscape of somatic copy-number alteration across human cancers. *Nature*. 2010;463(7283):899-905. doi:10.1038/nature08822
 45. Talhouk A, McConechy MK, Leung S, et al. A clinically applicable molecular-based classification for endometrial cancers. *Br J Cancer*. 2015;113(2):299-310. doi:10.1038/bjc.2015.190
 46. Talhouk A, McConechy MK, Leung S, et al. Confirmation of ProMisE: A simple, genomics-based clinical classifier for endometrial cancer. *Cancer*. 2017;123(5):802-813. doi:10.1002/cncr.30496
 47. Kommos S, McConechy MK, Kommos F, et al. Final validation of the ProMisE molecular classifier for endometrial carcinoma in a large population-based case series. *Ann Oncol*. 2018;29(5):1180-1188. doi:10.1093/annonc/mdy058
 48. Colombo N, Preti E, Landoni F, et al. Endometrial cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2013;24(SUPPL.6). doi:10.1093/annonc/mdt353
 49. Clark TJ, Barton PM, Coomarasamy A, Gupta JK, Khan KS. Investigating Postmenopausal Bleeding for Endometrial Cancer: Cost-Effectiveness of Initial Diagnostic Strategies. *Obstet Gynecol Surv*. 2006;61(9):581-582. doi:10.1097/01.ogx.0000234641.33376.57
 50. Liu Z, Doan Q V., Blumenthal P, Dubois RW. A systematic review evaluating health-related quality of life, work impairment, and health-care costs and utilization in abnormal uterine bleeding. *Value Heal*. 2007;10(3):183-194. doi:10.1111/j.1524-4733.2007.00168.x
 51. Dreisler E, Stampe Sorensen S, Ibsen PH, Lose G. Prevalence of endometrial polyps and abnormal uterine bleeding in a Danish population aged 20-74 years. *Ultrasound Obstet Gynecol*. 2009;33(1):102-108. doi:10.1002/uog.6259
 52. Astrup K, Olivarius NDF. Frequency of spontaneously occurring postmenopausal bleeding in the general population. *Acta Obstet Gynecol Scand*. 2004;83(2):203-207. doi:10.1111/j.0001-6349.2004.00400.x
 53. Pennant ME, Mehta R, Moody P, et al. Premenopausal abnormal uterine bleeding and risk of endometrial cancer. *BJOG An Int J Obstet Gynaecol*. 2017;124(3):404-411. doi:10.1111/1471-0528.14385
 54. Clarke MA, Long BJ, Del Mar Morillo A, Arbyn M, Bakkum-Gamez JN, Wentzensen N. Association of endometrial cancer risk with postmenopausal bleeding in women a systematic review and meta-analysis. *JAMA Intern Med*. 2018;178(9):1201-1208. doi:10.1001/jamainternmed.2018.2820

55. Timmermans A, Opmeer BC, Khan KS, et al. Endometrial thickness measurement for detecting endometrial cancer in women with postmenopausal bleeding: a systematic review and meta-analysis. *Obstet Gynecol.* 2010;116(1):160-167. doi:10.1097/AOG.0b013e3181e3e7e8
56. Gupta JK, Chien PFW, Voit D, Clark TJ, Khan KS. Ultrasonographic endometrial thickness for diagnosing endometrial pathology in women with postmenopausal bleeding: a meta-analysis. *Acta Obstet Gynecol Scand.* 2002;81(9):799-816. doi:10.1034/j.1600-0412.2001.810902.x
57. Smith-Bindman R, Kerlikowske K, Feldstein V a, et al. Endovaginal ultrasound to exclude endometrial cancer and other endometrial abnormalities. *JAMA.* 1998;280(17):1510-1517. doi:10.1001/jama.280.17.1510
58. ACOG Committee Opinion No. 734: The Role of Transvaginal Ultrasonography in Evaluating the Endometrium of Women With Postmenopausal Bleeding. Presented at the: 2018. doi:10.1097/AOG.0000000000002631
59. Tsuda H, Ito YM, Todo Y, et al. Measurement of endometrial thickness in premenopausal women in office gynecology. *Reprod Med Biol.* 2018;17(1):29-35. doi:10.1002/rmb2.12062
60. Dueholm M, Jensen ML, Laursen H, Kracht P. Can the endometrial thickness as measured by transvaginal sonography be used to exclude polyps or hyperplasia in pre-menopausal patients with abnormal uterine bleeding? *Acta Obstet Gynecol Scand.* 2001;80(7):645-651. doi:10.1034/j.1600-0412.2001.800710.x
61. Valentin L. Imaging techniques in the management of abnormal vaginal bleeding in non-pregnant women before and after menopause. *Best Pract Res Clin Obstet Gynaecol.* 2014;28(5):637-654. doi:10.1016/j.bpobgyn.2014.04.001
62. Epstein E, Blomqvist L. Imaging in endometrial cancer. *Best Pract Res Clin Obstet Gynaecol.* 2014;28(5):721-739. doi:10.1016/j.bpobgyn.2014.04.007
63. Dueholm M, Hjorth IMD. Structured imaging technique in the gynecologic office for the diagnosis of abnormal uterine bleeding. *Best Pract Res Clin Obstet Gynaecol.* 2017;40:23-43. doi:10.1016/j.bpobgyn.2016.09.010
64. De Kroon CD, De Bock GH, Dieben SWM, Jansen FW. Saline contrast hysterosonography in abnormal uterine bleeding: A systematic review and meta-analysis. *BJOG An Int J Obstet Gynaecol.* 2003;110(10):938-947. doi:10.1016/S1470-0328(03)02472-8
65. Farquhar C, Ekeroma A, Furness S, Arroll B. A systematic review of transvaginal ultrasonography, sonohysterography and hysteroscopy for the investigation of abnormal uterine bleeding in premenopausal women. *Acta Obstet Gynecol Scand.* 2003;82(6):493-504. doi:10.1034/j.1600-0412.2003.00191.x
66. Epstein E, Ramirez A, Skoog L, Valentin L. Transvaginal sonography, saline contrast sonohysterography and hysteroscopy for the investigation of women with postmenopausal bleeding and endometrium > 5 mm. *Ultrasound Obstet Gynecol.* 2001;18(2):157-162. doi:10.1046/j.1469-0705.2001.00472.x
67. Grimbizis GF, Tsolakidis D, Mikos T, et al. A prospective comparison of transvaginal ultrasound, saline infusion sonohysterography, and diagnostic hysteroscopy in the evaluation of endometrial pathology. *Fertil Steril.* 2010;94(7):2720-2725. doi:10.1016/j.fertnstert.2010.03.047
68. Opolskiene G, Sladkevicius P, Valentin L. Two- and three-dimensional saline contrast sonohysterography: interobserver agreement, agreement with hysteroscopy and diagnosis of endometrial malignancy. *Ultrasound Obstet Gynecol.* 2009;33(5):574-582. doi:10.1002/uog.6350
69. Dueholm M, Hjorth IMD, Secher P, Jørgensen A, Ørtoft G. Reproducibility of Endometrial Pathologic Findings Obtained on Hysteroscopy, Transvaginal Sonography, and Gel Infusion Sonography in Women With Postmenopausal Bleeding. *J Minim Invasive Gynecol.* 2015;22(6):1036-1044. doi:10.1016/j.jmig.2015.05.019
70. Epstein E, Ramirez A, Skoog L, Valentin L. Dilatation and curettage fails to detect most focal lesions in the uterine cavity in women with postmenopausal bleeding. *Acta Obstet Gynecol Scand.* 2001;80(12):1131-1136. doi:10.1034/j.1600-0412.2001.801210.x

71. Epstein E, Skoog L, Valentin L. Comparison of Endorette® and dilatation and curettage for sampling of the endometrium in women with postmenopausal bleeding. *Acta Obstet Gynecol Scand*. 2001;80(10):959-964. doi:10.1034/j.1600-0412.2001.801015.x
72. Clark TJ, Mann CH, Shah N, Khan KS, Song F, Gupta JK. Accuracy of outpatient endometrial biopsy in the diagnosis of endometrial cancer: a systematic quantitative review. *BJOG An Int J Obstet Gynaecol*. 2002;109(3):313-321. doi:10.1111/j.1471-0528.2002.01088.x
73. Paul F, Mol BWJ, Brlmann HAM, Peter M Heintz A. The accuracy of endometrial sampling in the diagnosis of patients with endometrial carcinoma and hyperplasia: A meta-analysis. *Cancer*. 2000;89(8):1765-1772. doi:10.1002/1097-0142(20001015)89:8<1765::AID-CNCR17>3.0.CO;2-F
74. Visser NCM, Sparidaens EM, van den Brink J-W, et al. Long-term risk of endometrial cancer following postmenopausal bleeding and reassuring endometrial biopsy. *Acta Obstet Gynecol Scand*. 2016;95(12):1418-1424. doi:10.1111/aogs.13022
75. Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor Quant Methods Psychol*. 2012;8(1):23-34. doi:10.20982/tqmp.08.1.p023
76. Epstein E, Valentin L. Intraobserver and interobserver reproducibility of ultrasound measurements of endometrial thickness in postmenopausal women. *Ultrasound Obstet Gynecol*. 2002;20(5):486-491. doi:10.1046/j.1469-0705.2002.00841.x
77. Nofech-Mozes S, Ismiil N, Dubé V, et al. Interobserver Agreement for Endometrial Cancer Characteristics Evaluated on Biopsy Material. *Obstet Gynecol Int*. 2012;2012:1-6. doi:10.1155/2012/414086
78. Lax SF, Kurman RJ, Pizer ES, Wu L, Ronnett BM. A binary architectural grading system for uterine endometrial endometrioid carcinoma has superior reproducibility compared with FIGO grading and identifies subsets of advance-stage tumors with favorable and unfavorable prognosis. *Am J Surg Pathol*. 2000;24(9):1201-1208. doi:10.1097/0000478-200009000-00002
79. Sagae S, Saito T, Satoh M, et al. The reproducibility of a binary tumor grading system for uterine endometrial endometrioid carcinoma, compared with FIGO system and nuclear grading. *Oncology*. 2004;67(5-6):344-350. doi:10.1159/000082917
80. Visser NCM, Reijnen C, Massuger LFAG, Nagtegaal ID, Bulten J, Pijnenborg JMA. Accuracy of endometrial sampling in endometrial carcinoma: A systematic review and meta-analysis. *Obstet Gynecol*. 2017;130(4):803-813. doi:10.1097/AOG.0000000000002261
81. Han G, Sidhu D, Duggan MA, et al. Reproducibility of histological cell type in high-grade endometrial carcinoma. *Mod Pathol*. 2013;26(12):1594-1604. doi:10.1038/modpathol.2013.102
82. Gilks CB, Oliva E, Soslow RA. Poor interobserver reproducibility in the diagnosis of high-grade endometrial carcinoma. *Am J Surg Pathol*. 2013;37(6):874-881. doi:10.1097/PAS.0b013e31827f576a
83. Talhouk A, Hoang LN, McConechy MK, et al. Molecular classification of endometrial carcinoma on diagnostic specimens is highly concordant with final hysterectomy: Earlier prognostic information to guide treatment. *Gynecol Oncol*. 2016;143(1):46-53. doi:10.1016/j.ygyno.2016.07.090
84. van der Putten LJM, van de Vijver K, Bartosch C, et al. Reproducibility of measurement of myometrial invasion in endometrial carcinoma. *Virchows Arch*. 2017;470(1):63-68. doi:10.1007/s00428-016-2035-5
85. Van Holsbeke C, Ameye L, Testa AC, et al. Development and external validation of (new) ultrasound based mathematical models for preoperative prediction of high-risk endometrial cancer. *Ultrasound Obstet Gynecol*. 2013;(September 2013):586-595. doi:10.1002/uog.13216
86. Lin MY, Dobrotwir A, McNally O, Abu-Rustum NR, Narayan K. Role of imaging in the routine management of endometrial cancer. *Int J Gynecol Obstet*. 2018;143:109-117. doi:10.1002/ijgo.12618
87. Comprehensive Cervical Cancer Control: A Guide to Essential Practice. In: 2nd ed. Geneva: World Health Organization; 2014. <https://www.ncbi.nlm.nih.gov/books/NBK269617/>.
88. Leone FPG, Timmerman D, Bourne T, et al. Terms, definitions and measurements to describe the sonographic features of the endometrium and intrauterine lesions: A consensus opinion from the

- International Endometrial Tumor Analysis (IETA) group. *Ultrasound Obstet Gynecol.* 2010;35(1):103-112. doi:10.1002/uog.7487
89. Epstein E, Fischerova D, Valentin L, et al. Ultrasound characteristics of endometrial cancer as defined by International Endometrial Tumor Analysis (IETA) consensus nomenclature: prospective multicenter study. *Ultrasound Obstet Gynecol.* 2018;51(6):818-828. doi:10.1002/uog.18909
 90. Sladkevicius P, Installé A, Van Den Bosch T, et al. International Endometrial Tumor Analysis (IETA) terminology in women with postmenopausal bleeding and sonographic endometrial thickness ≥ 4.5 mm: agreement and reliability study. *Ultrasound Obstet Gynecol.* 2018;51(2):259-268. doi:10.1002/uog.18813
 91. Van Calster B, Wynants L, Verbeek JFM, et al. Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. *Eur Urol.* 2018;74(6):796-804. doi:10.1016/j.eururo.2018.08.038
 92. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagnostic Progn Res.* 2019;3(1):1-8. doi:10.1186/s41512-019-0064-7
 93. Eriksson LSE, Epstein E, Testa AC, et al. An ultrasound-based risk model to predict lymph node metastases before surgery in women with endometrial cancer: a model development study. *Ultrasound Obstet Gynecol.* 2019. doi:10.1002/uog.21950
 94. Alcazar JL, Orozco R, Martinez-astorquiza Corral T, et al. Transvaginal ultrasound for preoperative assessment of myometrial invasion in patients with endometrial cancer: a systematic review and meta-analysis. *Ultrasound Obs Gynecol.* 2015;46(May):405-413. doi:10.1002/uog.14905
 95. Alcázar JL, Pérez L, Güell O, et al. Diagnostic Performance of Transvaginal Ultrasound for Detecting Cervical Invasion In Women With Endometrial Carcinoma: A Systematic Review and Meta-analysis. *J Ultrasound Med.* 2018. doi:10.1002/jum.14682
 96. Verbakel JY, Mascilini F, Wynants L, et al. Validation of ultrasound strategies to assess tumor extension and to predict high-risk endometrial cancer in women from the prospective IETA (International Endometrial Tumor Analysis)-4 cohort. *Ultrasound Obstet Gynecol.* 2020;55(1):115-124. doi:10.1002/uog.20374
 97. Eriksson LSE, Lindqvist PG, Floter Radestad A, et al. Transvaginal ultrasound assessment of myometrial and cervical stromal invasion in women with endometrial cancer: interobserver reproducibility among ultrasound experts and gynecologists. *Ultrasound Obstet Gynecol.* 2015;45(4):476-482. doi:10.1002/uog.14645
 98. Christensen JW, Dueholm M, Hansen ES, Marinovskij E, Lundorf E, Ørtoft G. Assessment of myometrial invasion in endometrial cancer using three-dimensional ultrasound and magnetic resonance imaging. *Acta Obstet Gynecol Scand.* 2016;95(1):55-64. doi:10.1111/aogs.12806
 99. Rodríguez-Trujillo A, Martínez-Serrano MJ, Martínez-Román S, et al. Preoperative Assessment of Myometrial Invasion in Endometrial Cancer by 3D Ultrasound and Diffusion-Weighted Magnetic Resonance Imaging: A Comparative Study. *Int J Gynecol Cancer.* 2016;26(6):1105-1110. doi:10.1097/IGC.0000000000000724
 100. Dietrich CF, Averkiou MA, Correas JM, Lassau N, Leen E, Piscaglia F. An EFSUMB introduction into dynamic contrast-enhanced ultrasound (DCE-US) for quantification of tumour perfusion. *Ultraschall der Medizin.* 2012;33(4):344-351. doi:10.1055/s-0032-1313026
 101. Paul Sidhu AS, Cantisani V, Dietrich CF, et al. The EFSUMB Guidelines and Recommendations for the Clinical Practice of Contrast-Enhanced Ultrasound (CEUS) in Non-Hepatic Applications: Update 2017 (Long Version) Die EFSUMB-Leitlinien und Empfehlungen für den klinischen Einsatz des kontrastverstärkten Ul. 2018;2017:154-180. doi:10.1055/s-0044-101254
 102. Strobel D, Seitz K, Blank W, et al. Contrast-enhanced ultrasound for the characterization of focal liver lesions - Diagnostic accuracy in clinical practice (DEGUM multicenter trial). *Ultraschall der Medizin.* 2008;29(5):499-505. doi:10.1055/s-2008-1027806
 103. Claudon M, Dietrich CF, Choi BI, et al. Guidelines and good clinical practice recommendations for contrast enhanced ultrasound (CEUS) in the liver - Update 2012. *Ultraschall der Medizin.* 2013;34(1):11-29. doi:10.1055/s-0032-1325499

104. Marret H, Sauget S, Giraudeau B, et al. Contrast-enhanced sonography helps in discrimination of benign from malignant adnexal masses. *J Ultrasound Med.* 2004;23(12):1629-1639; quiz 1641-1642.
105. Testa AC, Ferrandina G, Fruscella E, et al. The use of contrasted transvaginal sonography in the diagnosis of gynecologic diseases: a preliminary study. *J Ultrasound Med.* 2005;24(9):1267-1278.
106. Pálsdóttir K, Epstein E. A Pilot Study on Diagnostic Performance of Contrast-Enhanced Ultrasonography for Detection of Early Cervical Cancer. *Ultrasound Med Biol.* 2018;44(8):1664-1671. doi:10.1016/j.ultrasmedbio.2018.04.018
107. Liu Y, Tian JW, Xu Y, Cheng W. Role of transvaginal contrast-enhanced ultrasound in the early diagnosis of endometrial carcinoma. *Chin Med J (Engl).* 2012;125(3):416-421. doi:10.3760/cma.j.issn.0366-6999.2012.03.003
108. Liu Y, Xu Y, Cheng W, Liu X. Quantitative contrast-enhanced ultrasonography for the differential diagnosis of endometrial hyperplasia and endometrial neoplasms. *Oncol Lett.* September 2016:3763-3770. doi:10.3892/ol.2016.5206
109. Song Y, Yang J, Liu Z, Shen K. Preoperative evaluation of endometrial carcinoma by contrast-enhanced ultrasonography. *BJOG An Int J Obstet Gynaecol.* 2009;116(2):294-298. doi:10.1111/j.1471-0528.2008.01981.x
110. Lieng M, Qvigstad E, Dahl GF, Istre O. Flow differences between endometrial polyps and cancer: A prospective study using intravenous contrast-enhanced transvaginal color flow Doppler and three-dimensional power Doppler ultrasound. *Ultrasound Obstet Gynecol.* 2008;32(7):935-940. doi:10.1002/uog.6267
111. Zhou H, Xiang H, Duan L, et al. Application of Combined Two-Dimensional and Three-Dimensional Transvaginal Contrast Enhanced Ultrasound in the Diagnosis of Endometrial Carcinoma. *Biomed Res Int.* 2015;2015:1-10. doi:10.1155/2015/292743
112. Nougaret S, Horta M, Sala E, et al. Endometrial Cancer MRI staging: Updated Guidelines of the European Society of Urogenital Radiology. *Eur Radiol.* 2019;29(2):792-805. doi:10.1007/s00330-018-5515-y
113. Deng L, Wang Q, Chen X, Duan X, Wang W, Guo Y. The Combination of Diffusion- and T2-Weighted Imaging in Predicting Deep Myometrial Invasion of Endometrial Cancer. *J Comput Assist Tomogr.* 2015;39(5):661-673. doi:10.1097/RCT.0000000000000280
114. Luomaranta A, Leminen A, Loukovaara M. Magnetic resonance imaging in the assessment of high-risk features of endometrial carcinoma: a meta-analysis. *Int J Gynecol Cancer.* 2015;25(5):837-842. doi:10.1097/IGC.0000000000000194
115. Lalwani N, Dubinsky T, Javitt MC, et al. ACR Appropriateness Criteria® pretreatment evaluation and follow-up of endometrial cancer. *Ultrasound Q.* 2014;30(1):21-28. doi:10.1097/RUQ.0000000000000068
116. Alcázar JL, Gastón B, Navarro B, Salas R, Aranda J, Guerriero S. Transvaginal ultrasound versus magnetic resonance imaging for preoperative assessment of myometrial infiltration in patients with endometrial cancer: a systematic review and meta-analysis. *J Gynecol Oncol.* 2017;28(4):405-413. doi:10.3802/jgo.2017.28.e86
117. Haldorsen IS, Husby JA, Werner HMJ, et al. Standard 1.5-T MRI of endometrial carcinomas: Modest agreement between radiologists. *Eur Radiol.* 2012;22(7):1601-1611. doi:10.1007/s00330-012-2400-y
118. Frost JA, Webster KE, Bryant A, Morrison J. Lymphadenectomy for the management of endometrial cancer. *Cochrane Database Syst Rev.* 2017;(10). doi:10.1002/14651858.CD007585.pub4
119. Todo Y, Kato H, Kaneuchi M, Watari H, Takeda M, Sakuragi N. Survival effect of para-aortic lymphadenectomy in endometrial cancer (SEPAL study): a retrospective cohort analysis. *Lancet.* 2010;375(9721):1165-1172. doi:10.1016/S0140-6736(09)62002-X
120. Kong A, Johnson N, Kitchener HC, Lawrie TA. Adjuvant radiotherapy for stage i endometrial cancer: An updated cochrane systematic review and meta-analysis. *J Natl Cancer Inst.* 2012;104(21):1625-1634. doi:10.1093/jnci/djs374

121. Yi L, Zhang H, Zou J, Luo P, Zhang J. Adjuvant chemoradiotherapy versus radiotherapy alone in high-risk endometrial cancer: A systematic review and meta-analysis. *Gynecol Oncol.* 2018;149(3):612-619. doi:10.1016/j.ygyno.2018.03.004
122. de Boer SM, Powell ME, Mileskin L, et al. Adjuvant chemoradiotherapy versus radiotherapy alone in women with high-risk endometrial cancer (PORTEC-3): patterns of recurrence and post-hoc survival analysis of a randomised phase 3 trial. *Lancet Oncol.* 2019;20(9):1273-1285. doi:10.1016/S1470-2045(19)30395-X
123. Koh WJ, Abu-Rustum NR, Bean S, et al. Uterine Neoplasms, Version 1.2018: Clinical practice guidelines in oncology. *JNCCN J Natl Compr Cancer Netw.* 2018;16(2):170-199. doi:10.6004/jnccn.2018.0006
124. Schuenke M, Schulte E, Schumacher U. *Thieme Atlas of Anatomy- Neck and Internal Organs.* (Ross LM, Lamperti ED, eds.). Stuttgart: Thieme; 2006.
125. Geppert B, Lönnerfors C, Bollino M, Arechvo A, Persson J. A study on uterine lymphatic anatomy for standardization of pelvic sentinel lymph node detection in endometrial cancer. *Gynecol Oncol.* 2017;145(2):256-261. doi:10.1016/j.ygyno.2017.02.018
126. Abu-Rustum NR, Gomez JD, Alektiar KM, et al. The incidence of isolated paraaortic nodal metastasis in surgically staged endometrial cancer patients with negative pelvic lymph nodes. *Gynecol Oncol.* 2009;115(2):236-238. doi:10.1016/j.ygyno.2009.07.016
127. Chiang AJ, Yu KJ, Chao KC, Teng NNH. The incidence of isolated para-aortic nodal metastasis in completely staged endometrial cancer patients. *Gynecol Oncol.* 2011;121(1):122-125. doi:10.1016/j.ygyno.2010.11.026
128. Kumar S, Podratz KC, Bakkum-Gamez JN, et al. Prospective assessment of the prevalence of pelvic, paraaortic and high paraaortic lymph node metastasis in endometrial cancer. *Gynecol Oncol.* 2014;132(1):38-43. doi:10.1016/j.ygyno.2013.10.002
129. Nasioudis D, Holcomb K. Incidence of isolated para-aortic lymph node metastasis in early stage endometrial cancer. *Eur J Obstet Gynecol Reprod Biol.* 2019;242:43-46. doi:10.1016/j.ejogrb.2019.09.003
130. Bodurtha Smith AJ, Fader AN, Tanner EJ. Sentinel lymph node assessment in endometrial cancer: a systematic review and meta-analysis. *Am J Obstet Gynecol.* 2017;216(5):459-476.e10. doi:10.1016/j.ajog.2016.11.1033
131. Holloway RW, Abu-Rustum NR, Backes FJ, et al. Sentinel lymph node mapping and staging in endometrial cancer: A Society of Gynecologic Oncology literature review with consensus recommendations. *Gynecol Oncol.* 2017;146(2):405-415. doi:10.1016/j.ygyno.2017.05.027
132. Barlin JN, Khoury-Collado F, Kim CH, et al. The importance of applying a sentinel lymph node mapping algorithm in endometrial cancer staging: Beyond removal of blue nodes. *Gynecol Oncol.* 2012;125(3):531-535. doi:10.1016/j.ygyno.2012.02.021
133. Ehrisman J, Secord AA, Berchuck A, et al. Performance of sentinel lymph node biopsy in high-risk endometrial cancer. *Gynecol Oncol Reports.* 2016;17:69-71. doi:10.1016/j.gore.2016.04.002
134. Bogani G, Murgia F, Ditto A, Raspagliesi F. Sentinel node mapping vs. lymphadenectomy in endometrial cancer: A systematic review and meta-analysis. *Gynecol Oncol.* 2019;153(3):676-683. doi:10.1016/j.ygyno.2019.03.254
135. Ballester M, Dubernard G, Lécure F, et al. Detection rate and diagnostic accuracy of sentinel-node biopsy in early stage endometrial cancer: A prospective multicentre study (SENTI-ENDO). *Lancet Oncol.* 2011;12(5):469-476. doi:10.1016/S1470-2045(11)70070-5
136. Rossi EC, Kowalski LD, Scalici J, et al. A comparison of sentinel lymph node biopsy to lymphadenectomy for endometrial cancer staging (FIRES trial): a multicentre, prospective, cohort study. *Lancet Oncol.* 2017;18(3):384-392. doi:10.1016/S1470-2045(17)30068-2
137. Persson J, Salehi S, Bollino M, Lönnerfors C, Falconer H, Geppert B. Pelvic Sentinel lymph node detection in High-Risk Endometrial Cancer (SHREC-trial)—the final step towards a paradigm shift in

- surgical staging. *Eur J Cancer*. 2019;116:77-85. doi:10.1016/j.ejca.2019.04.025
138. Plante M, Stanleigh J, Renaud MC, Sebastianelli A, Grondin K, Grégoire J. Isolated tumor cells identified by sentinel lymph node mapping in endometrial cancer: Does adjuvant treatment matter? *Gynecol Oncol*. 2017;146(2):240-246. doi:10.1016/j.ygyno.2017.05.024
 139. Bakkum-Gamez JN, Mariani A, Dowdy SC, et al. The impact of surgical guidelines and periodic quality assessment on the staging of endometrial cancer. *Gynecol Oncol*. 2011;123(1):58-64. doi:10.1016/j.ygyno.2011.06.018
 140. How J, Lau S, Press J, et al. Accuracy of sentinel lymph node detection following intra-operative cervical injection for endometrial cancer: A prospective study. *Gynecol Oncol*. 2012;127(2):332-337. doi:10.1016/j.ygyno.2012.08.018
 141. Holloway RW, Gupta S, Stavitzski NM, et al. Sentinel lymph node mapping with staging lymphadenectomy for patients with endometrial cancer increases the detection of metastasis. *Gynecol Oncol*. 2016;141(2):206-210. doi:10.1016/j.ygyno.2016.02.018
 142. Eriksson AGZ, Ducie J, Ali N, et al. Comparison of a sentinel lymph node and a selective lymphadenectomy algorithm in patients with endometrioid endometrial carcinoma and limited myometrial invasion. *Gynecol Oncol*. 2016;140(3):394-399. doi:10.1016/j.ygyno.2015.12.028
 143. Buda A, Di Martino G, Restaino S, et al. The impact on survival of two different staging strategies in apparent early stage endometrial cancer comparing sentinel lymph nodes mapping algorithm and selective lymphadenectomy: An Italian retrospective analysis of two reference centers. *Gynecol Oncol*. 2017;147(3):528-534. doi:10.1016/j.ygyno.2017.09.033
 144. How JA, O'Farrell P, Amajoud Z, et al. Sentinel lymph node mapping in endometrial cancer: a systematic review and meta-analysis. *Minerva Ginecol*. 2018;70(2):194-214. doi:10.23736/S0026-4784.17.04179-X
 145. Backes FJ, Cohen D, Salani R, et al. Prospective clinical trial of robotic sentinel lymph node assessment with isosulfane blue (ISB) and indocyanine green (ICG) in endometrial cancer and the impact of ultrastaging (NCT01818739). *Gynecol Oncol*. 2019;153(3):496-499. doi:10.1016/j.ygyno.2019.03.252
 146. Piedimonte S, Richer L, Souhami L, et al. Clinical significance of isolated tumor cells and micrometastasis in low-grade, stage I endometrial cancer. *J Surg Oncol*. 2018;118(7):1194-1198. doi:10.1002/jso.25259
 147. Goebel EA, St Laurent JD, Nucci MR, Feltmate CM. Retrospective detection of isolated tumor cells by immunohistochemistry in sentinel lymph node biopsy performed for endometrial carcinoma: is there clinical significance? *Int J Gynecol Cancer*. 2020;30(3):291-298. doi:10.1136/ijgc-2019-000934
 148. Wei J, Zhang W, Feng L, Gao W. Comparison of fertility-sparing treatments in patients with early endometrial cancer and atypical complex hyperplasia: A meta-analysis and systematic review. *Med (United States)*. 2017;96(37):1-9. doi:10.1097/MD.00000000000008034
 149. Bendifallah S, Canlorbe G, Raimond E, et al. A clue towards improving the European Society of Medical Oncology risk group classification in apparent early stage endometrial cancer? Impact of lymphovascular space invasion. *Br J Cancer*. 2014;110(11):2640-2646. doi:10.1038/bjc.2014.237
 150. Bendifallah S, Canlorbe G, Collinet P, et al. Just how accurate are the major risk stratification systems for early-stage endometrial cancer? *Br J Cancer*. 2015;112(5):793-801. doi:10.1038/bjc.2015.35
 151. Korkmaz V, Meydanli MM, Yalçın I, et al. Comparison of three different risk-stratification models for predicting lymph node involvement in endometrioid endometrial cancer clinically confined to the uterus. *J Gynecol Oncol*. 2017;28(6):1-11. doi:10.3802/jgo.2017.28.e78
 152. Institute NNC. Cancer Facts & Figures 2020. *CA Cancer J Clin*. 2020:1-76.
 153. Cancer Research UK. <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/uterine-cancer>. Accessed April 14, 2020.
 154. *Endometriecancer- Nationellt Vårdprogram*. 1.3. Regionala cancercentrum i samverkan; 2020.

155. Creasman W. Revised FIGO staging for carcinoma of the endometrium. *Int J Gynaecol Obstet.* 2009;105(2):109. doi:10.1016/j.ijgo.2009.02.010
156. Creasman WT, Odicino F, Maisonneuve P, et al. Carcinoma of the corpus uteri. FIGO 26th Annual Report on the Results of Treatment in Gynecological Cancer. *Int J Gynaecol Obstet.* 2006;95 Suppl 1:S105-43. doi:10.1016/S0020-7292(06)60031-3
157. Hamilton CA, Cheung MK, Osann K, et al. Uterine papillary serous and clear cell carcinomas predict for poorer survival compared to grade 3 endometrioid corpus cancers. *Br J Cancer.* 2006;94(5):642-646. doi:10.1038/sj.bjc.6603012
158. Kim HJ, Kim TJ, Lee YY, et al. A comparison of uterine papillary serous, clear cell carcinomas, and grade 3 endometrioid corpus cancers using 2009 FIGO staging system. *J Gynecol Oncol.* 2013;24(2):120-127. doi:10.3802/jgo.2013.24.2.120
159. Ståhlberg K, Kihlhed P, Bjurberg M, et al. Risk factors for lymph node metastases in women with endometrial cancer: A population-based, nation-wide register study-On behalf of the Swedish Gynecological Cancer Group. *Int J Cancer.* 2017;(140):2693-2700. doi:10.1002/ijc.30707
160. Sorbe B. Predictive and prognostic factors in definition of risk groups in endometrial carcinoma. *ISRN Obstet Gynecol.* 2012;2012:325790. doi:10.5402/2012/325790
161. Tosun G, Solmaz U, Ozdemir A, et al. Lymphovascular space invasion and cervical stromal invasion are independent risk factors for nodal metastasis in endometrioid endometrial cancer. *Aust New Zeal J Obstet Gynaecol.* 2015;55(1):81-86. doi:10.1111/ajo.12321
162. McCluggage WG, Hirschowitz L, Wilson GE, Oliva E, Soslow RA, Zaino RJ. Significant variation in the assessment of cervical involvement in endometrial carcinoma: An interobserver variation study. *Am J Surg Pathol.* 2011;35(2):289-294. doi:10.1097/PAS.0b013e3182073ac0
163. Veade AE, Foote J, Ehrisman J, et al. Associations between lymphovascular space invasion, nodal recurrence, and survival in patients with surgical stage I endometrioid endometrial adenocarcinoma. *World J Surg Oncol.* 2019;17(1):4-11. doi:10.1186/s12957-019-1620-x
164. Weinberg LE, Kunos C a, Zanotti KM. Lymphovascular space invasion (LVSI) is an isolated poor prognostic factor for recurrence and survival among women with intermediate- to high-risk early-stage endometrioid endometrial cancer. *Int J Gynecol Cancer.* 2013;23(8):1438-1445. doi:10.1097/IGC.0b013e3182a16c93
165. Cusano E, Myers V, Samant R, et al. Prognostic Significance of Lymphovascular Space Invasion in the Absence of Lymph Node Metastases in Early-Stage Endometrial Cancer. *Int J Gynecol Cancer.* 2018;28(5):890-894. doi:10.1097/IGC.0000000000001229
166. Ståhlberg K, Bjurberg M, Borgfeldt C, et al. Lymphovascular space invasion as a predictive factor for lymph node metastases and survival in endometrioid endometrial cancer—a Swedish Gynecologic Cancer Group (SweGCG) study. *Acta Oncol (Madr).* 2019;58(11):1628-1633. doi:10.1080/0284186X.2019.1643036
167. Bosse T, Peters EEM, Creutzberg CL, et al. Substantial lymph-vascular space invasion (LVSI) is a significant risk factor for recurrence in endometrial cancer – A pooled analysis of PORTEC 1 and 2 trials. *Eur J Cancer.* 2015;31:1742-1750. doi:10.1016/j.ejca.2015.05.015
168. Peters EEM, Bartosch C, McCluggage WG, et al. Reproducibility of lymphovascular space invasion (LVSI) assessment in endometrial cancer. *Histopathology.* 2019;75(1):128-136. doi:10.1111/his.13871
169. Giam M, Rancati G. Aneuploidy and chromosomal instability in cancer: A jackpot to chaos. *Cell Div.* 2015;10(1):1-12. doi:10.1186/s13008-015-0009-7
170. Lundgren C, Auer G, Frankendal B, Moberger B, Nilsson B, Nordström B. Nuclear DNA content, proliferative activity, and p53 expression related to clinical and histopathologic features in endometrial carcinoma. *Int J Gynecol Cancer.* 2002;12(1):110-118. <http://www.ncbi.nlm.nih.gov/pubmed/11860545>.
171. Susini T, Amunni G, Molino C, et al. Ten-year results of a prospective study on the prognostic role of ploidy in endometrial carcinoma: DNA aneuploidy identifies high-risk cases among the so-called “low-risk” patients with well and moderately differentiated tumors. *Cancer.* 2007;109(5):882-890.

doi:10.1002/cncr.22465

172. Pfisterer J, Kommoss F, Sauerbrei W, et al. Prognostic value of DNA ploidy and S-phase fraction in stage I endometrial carcinoma. *Gynecol Oncol.* 1995;58:149-156.
173. Pisani A, Barbuto D, Chen D, Ramos L, Lagasse L, Karlan B. HER-2/neu, p53, and DNA Analyses as Prognosticators for Survival in Endometrial Carcinoma. *Obstet Gynecol.* 1995;85(5):729-734. doi:10.1016/0029-7844(95)00037-R
174. Salvesen HB, Iversen OE, Akslen LA. Prognostic Impact of Morphometric Nuclear Grade of Endometrial Carcinoma. *Cancer.* 1998;83(5):956-964.
175. Pradhan M, Abeler VM, Danielsen HE, et al. Prognostic importance of DNA ploidy and DNA index in stage I and II endometrioid adenocarcinoma of the endometrium. *Ann Oncol.* 2012;23(5):1178-1184. doi:10.1093/annonc/mdr368
176. Proctor L, Pradhan M, Leung S, et al. Assessment of DNA Ploidy in the ProMisE molecular subgroups of endometrial cancer. *Gynecol Oncol.* 2017;146(3):596-602. doi:10.1016/j.ygyno.2017.06.020
177. Wagenius G, Bergström R, Strang P, et al. Prognostic significance of flow cytometric and clinical variables in endometrial adenocarcinoma stages I and II. *Anticancer Res.* 1992;12(3):725-732. <http://www.ncbi.nlm.nih.gov/pubmed/1622130>.
178. Susini T, Rapi S, Savino L, Boddi V, Berti P, Massi G. Prognostic value of flow cytometric deoxyribonucleic acid index in endometrial carcinoma: comparison with other clinical-pathologic parameters. *Am J Obstet Gynecol.* 1994;170(2):527-534. <http://www.ncbi.nlm.nih.gov/pubmed/8116708>.
179. Nordström B, Strang P, Bergström R, Nilsson S, Tribukait B. A Comparison of Proliferation Markers and Their Prognostic Value for Women with Endometrial Carcinoma. *Cancer.* 1996;78(9):1942-1951.
180. Lundgren C, Auer G, Frankendal B, Nilsson B, Nordström B. Prognostic factors in surgical stage I endometrial carcinoma. *Acta Oncol (Madr).* 2004;43(1):49-56. doi:10.1080/02841860310018990
181. Mangili G, Montoli S, De Marzi P, Sassi I, Aletti G, Taccagni G. The role of DNA ploidy in postoperative management of stage I endometrial cancer. *Ann Oncol.* 2008;19(7):1278-1283. doi:10.1093/annonc/mdn041
182. Song T, Lee J-W, Kim H-J, et al. Prognostic significance of DNA ploidy in stage I endometrial cancer. *Gynecol Oncol.* 2011;122(1):79-82. doi:10.1016/j.ygyno.2011.03.017
183. Green RW, Engblom S, Baldetorp B, Hartman L, Måsbäck A, Bjurberg M. Cell proliferation, measured as flow cytometric S-phase fraction, is a strong prognostic indicator in FIGO stage I endometrioid endometrial carcinoma: a population-based study. *Acta Obstet Gynecol Scand.* 2015;94(10):1064-1073. doi:10.1111/aogs.12704
184. Svanvik T, Strömberg U, Holmberg E, Marcickiewicz J, Sundfeldt K. DNA ploidy status, S-phase fraction, and p53 are not independent prognostic factors for survival in endometrioid endometrial carcinoma FIGO stage I-III. *Int J Gynecol Cancer.* 2019;29(2):305-311. doi:10.1136/ijgc-2018-000082
185. Njølstad TS, Trovik J, Hveem TS, et al. DNA ploidy in curettage specimens identifies high-risk patients and lymph node metastasis in endometrial cancer. *Br J Cancer.* 2015;112(10):1656-1664. doi:10.1038/bjc.2015.123
186. Stendahl U, Strang P, Wagenius G, Bergström R, Tribukait B. Prognostic Significance of Proliferation in Endometrial Adenocarcinomas. *Int J Gynecol Pathol.* 1991;10(3):271-284. doi:10.1097/00004347-199107000-00006
187. Sorbe B, Risberg B, Thornthwaite J. Nuclear morphometry and DNA flow cytometry as prognostic methods for endometrial carcinoma. *Int J Gynecol Cancer.* 1994;4(2):94-100. <http://www.ncbi.nlm.nih.gov/pubmed/11578391>.
188. Friberg L-G, Norén H, Delle U. Prognostic Value of DNA Ploidy and S-Phase Fraction in Endometrial Cancer Stage I and II: A Prospective 5-Year Survival Study. *Gynecol Oncol.* 1994;53(1):64-69. doi:10.1006/gyno.1994.1089

189. Von Minckwitz G, Kuhn W, Kaufmann M, et al. Prognostic importance of DNA-ploidy and S-phase fraction in endometrial cancer. *Int J Gynecol Cancer*. 1994;4(4):250-256. doi:10.1046/j.1525-1438.1994.04040250.x
190. Gudmundsson TE, Högberg T, Alm P, et al. The prognostic information of DNA ploidy and S-phase fraction may vary with histologic grade in endometrial carcinoma. *Acta Oncol*. 1995;34(6):803-812. <http://www.ncbi.nlm.nih.gov/pubmed/7576749>.
191. Kaleli S, Kösebay D, Bese T, et al. A Strong Prognostic Variable in Endometrial Carcinoma: Flow cytometric S-phase Fraction. *Cancer*. 1997;79(5):944-951.
192. Evans T, Sany O, Pearmain P, Ganesan R, Blann A, Sundar S. Differential trends in the rising incidence of endometrial cancer by type: Data from a UK population-based registry from 1994 to 2006. *Br J Cancer*. 2011;104(9):1505-1510. doi:10.1038/bjc.2011.68
193. Boll D, Karim-Kos HE, R.h.a.verhoeven, et al. Increased incidence and improved survival in endometrioid endometrial cancer diagnosed since 1989 in the Netherlands: A population based study. *Eur J Obstet Gynecol Reprod Biol*. 2013;166(2):209-214. doi:10.1016/j.ejogrb.2012.10.028
194. Barlow L, Westergren K, Holmberg L, Talbäck M. The completeness of the Swedish Cancer Register: a sample survey for year 1998. *Acta Oncol*. 2009;48(1):27-33. doi:10.1080/02841860802247664
195. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. *J Biomed Inform*. 2014;48:193-204. doi:10.1016/j.jbi.2014.02.013
196. Watson PF, Petrie A. Method agreement analysis: A review of correct methodology. *Theriogenology*. 2010;73(9):1167-1179. doi:10.1016/j.theriogenology.2010.01.003
197. Kitchen CMR. Nonparametric vs Parametric Tests of Location in Biomedical Research. *Am J Ophthalmol*. 2009;147(4):571-572. doi:10.1016/j.ajo.2008.06.031
198. Fay MP, Proschan MA. Wilcoxon-Mann-whitney or t-test? *Stat Surv*. 2010;4:1-37. doi:10.1214/09-SS051.Wilcoxon-Mann-Whitney
199. Fagerland Morten W. T- Tests, Non-Parametric Tests, and Large Studies—a Paradox of Statistical Practice? *BMC Med Res Methodol*. 2012;12(1):78.
200. Cox DR. Regression Models and Life-Tables. *J R Stat Soc Ser B*. 1972;34(2):187-220. <http://www.jstor.org/stable/2985181>.
201. Zhang MJ, Zhang X, Scheike TH. Modeling cumulative incidence function for competing risks data. *Expert Rev Clin Pharmacol*. 2008;1(3):391-400. doi:10.1586/17512433.1.3.391
202. de Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol*. 2006;59(10):1033-1039. doi:10.1016/j.jclinepi.2005.10.015
203. Kottner J, Audige L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Int J Nurs Stud*. 2011;48(6):661-671. doi:10.1016/j.ijnurstu.2011.01.016
204. Cicchetti D V, Feinstein a R. High Agreement but Low Kappa. *J Clin Epidemiol*. 1990;43(6):551-585.
205. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educ Psychol Meas*. 1960;20(1):37-46. doi:10.1177/001316446002000104
206. Bland JM, Altman DG. Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. *Lancet*. 1986;327:307-310. doi:10.1016/S0140-6736(86)90837-8
207. Altman DG, Bland JM. Diagnostic tests 2: Predictive values. *BMJ*. 1994;309(6947):102. <http://www.ncbi.nlm.nih.gov/pubmed/8038641>.
208. Ioannidis JPA. Why most published research findings are false. *Get to Good Res Integr Biomed Sci*. 2018;2(8):2-8. doi:10.1371/journal.pmed.0020124
209. Fisher RA. The Arrangement of Field Experiments. *J Minist Agric Gt Britain*. 1926;33:503-513. doi:10.1007/978-1-4612-4380-9_8

210. McDonald J. *Handbook of Biological Statistics*. 3rd ed. Sparky House Publishing; 2014. <http://www.biostathandbook.com/multiplecomparisons.html>.
211. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B*. 1995;57(1):289-300. doi:10.1111/j.2517-6161.1995.tb02031.x
212. Van Noorden R, Maher B, Nuzzo R. The top 100 papers. *Nature*. 2014;514(7524):550-553. doi:10.1038/514550a
213. Krzywinski M, Altman N. Points of significance: Comparing samples-part II. *Nat Methods*. 2014;11(4):355-356. doi:10.1038/nmeth.2900
214. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Stat*. 2001;29(4):1165-1188. doi:10.1214/aos/1013699998
215. Arem H, Irwin ML. Obesity and endometrial cancer survival: a systematic review. *Int J Obes*. 2013;37(5):634-639. doi:10.1038/ijo.2012.94
216. Textor J, van der Zander B, Gilthorpe MS, Liškiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: The R package “dagitty.” *Int J Epidemiol*. 2016;45(6):1887-1894. doi:10.1093/ije/dyw341