

Clarifying risk-factor associations with quantitative breast tumor features: a pooled analysis of 24 studies.

Daniel Adams¹, Amber N. Hurson², Thomas U. Ahearn², Irene L. Andrulis^{3, 4}, Antonis C. Antoniou⁵, Päivi Auvinen⁶⁻⁸, Anita Bane⁹, Sabine Behrens¹⁰, Amy Berrington de Gonzalez¹, Katarzyna Białkowska¹¹, Clara Bodelon¹², Manjeet K. Bolla⁵, Maria Borrero¹³, Hermann Brenner¹⁴, Ian W. Brock¹⁵, Annegien Broeks¹⁶, Nicola J. Camp¹⁷, Melissa H. Cessna¹³, Stephen J. Chanock², Georgia Chenevix-Trench¹⁸, Ellen Copson¹⁹, Fergus J. Couch²⁰, Angela Cox¹⁵, Simon S. Cross¹⁵, Peter Devilee^{21, 22}, Alison M. Dunning²³, Douglas F. Easton^{5, 23}, Diana M. Eccles²⁴, Jonine D. Figueroa^{2, 25, 26}, Gord Glendon³, Anna González-Neira^{27, 28}, Ute Hamann²⁹, Jaana M. Hartikainen^{6, 30}, Mikael Hartman³¹⁻³³, Bernadette A.M. Heemskerk-Gerritsen³⁴, Peh Joo Ho³², James M. Hodge¹², Bernd Holleczech³⁵, Antoinette Hollestelle³⁴, Maartje J. Hoening^{34, 36}, kConFab Investigators^{37, 38}, SGBCC Investigators^{31, 32, 39-49}, Anna Jakubowska^{11, 50}, Michael E. Jones¹, Louise Jones⁵¹, Brandt Jones¹⁷, Audrey Jung¹⁰, Rudolf Kaaks¹⁰, Renske Keeman¹⁶, Scott Lawrence², Jingmei Li^{33, 39, 52}, Jolanta Lissowska⁵³, Jan Lubiński¹¹, Arto Mannermaa^{6, 30, 54}, Mehdi Manoochehri²⁹, John W.M. Martens³⁴, Wilma E. Mesker⁵⁵, Noor Muhammad⁵⁶, Anna Marie Mulligan^{9, 57}, Nadia Obi^{58, 59}, Janet E. Olson⁶⁰, Ana Osorio^{28, 61}, Penelope D. Ottewell¹⁵, Alpa V. Patel¹², Guillermo Pita²⁷, Karolina Prajzendanc¹¹, Muhammad U. Rashid^{29, 56}, Kathryn J. Ruddy⁶², Fatemeh Safizadeh¹⁴, Ben Schöttker^{63, 64}, Petra Seibold¹⁰, William J. Tapper²⁴, Maria Tengström⁶⁻⁸, Lauren R. Teras¹², Heather Thorne^{37, 38}, Mieke A.M. Timmermans³⁴, Rob A.E.M. Tollenaar⁵⁵, Celine M. Vachon⁶⁰, Qin Wang⁵, Jelle Wesseling^{16, 65}, Siddhartha Yadav⁶², Xiaohong R. Yang², M. Pilar Zamora⁶⁶, Jenny Chang-Claude^{10, 67}, Marjanka K. Schmidt^{16, 68, 69,*}, Paul D.P. Pharoah^{70,*}, Mustapha Abubakar^{2,*}, Montserrat García-Closas^{1, 71,*,**}

¹ Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK

² Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

³ Fred A. Litwin Center for Cancer Genetics, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada

⁴ Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada

⁵ Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

⁶ Translational Cancer Research Area, University of Eastern Finland, Kuopio, Finland

⁷ Institute of Clinical Medicine, Oncology, University of Eastern Finland, Kuopio, Finland

⁸ Department of Oncology, Cancer Center, Kuopio University Hospital, Kuopio, Finland

⁹ Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada

¹⁰ Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany

¹¹ Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland

¹² Department of Population Science, American Cancer Society, Atlanta, GA, USA

¹³ Intermountain Healthcare, Salt Lake City, UT, USA

¹⁴ Cancer Prevention Graduate School, German Cancer Research Center (DKFZ), Heidelberg, Germany

- ¹⁵ Sheffield Institute for Nucleic Acids (SInFoNiA), Division of Clinical Medicine, School of Medicine and Population Health, University of Sheffield, Sheffield, UK
- ¹⁶ Division of Molecular Pathology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Amsterdam, the Netherlands
- ¹⁷ Department of Internal Medicine and Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA
- ¹⁸ Cancer Research Program, QIMR Berghofer, Brisbane, Queensland, Australia
- ¹⁹ Department of Medical Oncology, University of Southampton, Southampton, UK
- ²⁰ Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA
- ²¹ Department of Pathology, Leiden University Medical Center, Leiden, the Netherlands
- ²² Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands
- ²³ Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK
- ²⁴ Faculty of Medicine, University of Southampton, Southampton, UK
- ²⁵ Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh, Edinburgh, UK
- ²⁶ Cancer Research UK Edinburgh Centre, The University of Edinburgh, Edinburgh, UK
- ²⁷ Human Genotyping Unit, Spanish National Cancer Research Centre (CNIO), Madrid, Spain
- ²⁸ Centre for Biomedical Network Research on Rare Diseases (CIBERER), Instituto de Salud Carlos III, Madrid, Spain

²⁹ Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany

³⁰ Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Finland

³¹ Department of Surgery, National University Health System, Singapore City, Singapore

³² Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore City, Singapore

³³ Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore, Singapore City, Singapore

³⁴ Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, the Netherlands

³⁵ Saarland Cancer Registry, Saarbrücken, Germany

³⁶ Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, the Netherlands

³⁷ Research Department, Peter MacCallum Cancer Center, Melbourne, Victoria, Australia

³⁸ The Sir Peter MacCallum Department of Oncology, The University of Melbourne, Parkville, Victoria, Australia

³⁹ Genome Institute of Singapore, Agency for Science, Technology and Research (A*STAR), Singapore City, Singapore

⁴⁰ Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore City, Singapore

⁴¹ Cancer Genetics Service, National Cancer Centre, Singapore City, Singapore

- ⁴² Breast Department, KK Women's and Children's Hospital, Singapore City, Singapore
- ⁴³ SingHealth Duke-NUS Breast Centre, Singapore City, Singapore
- ⁴⁴ Department of General Surgery, Tan Tock Seng Hospital, Singapore City, Singapore
- ⁴⁵ Division of Surgery and Surgical Oncology, National Cancer Centre, Singapore City, Singapore
- ⁴⁶ Department of General Surgery, Singapore General Hospital, Singapore City, Singapore
- ⁴⁷ Division of Breast Surgery, Department of General Surgery, Changi General Hospital, Singapore City, Singapore
- ⁴⁸ Division of Radiation Oncology, National Cancer Centre, Singapore City, Singapore
- ⁴⁹ Division of Medical Oncology, National Cancer Centre, Singapore City, Singapore
- ⁵⁰ Independent Laboratory of Molecular Biology and Genetic Diagnostics, Pomeranian Medical University, Szczecin, Poland
- ⁵¹ Barts Cancer Institute, Queen Mary University of London, London, UK
- ⁵² National Cancer Centre Singapore, Singapore Health Services (SingHealth), Singapore City, Singapore
- ⁵³ Department of Cancer Epidemiology and Prevention, M Sklodowska-Curie National Research Oncology Institute, Warsaw, Poland
- ⁵⁴ Biobank of Eastern Finland, Kuopio University Hospital, Kuopio, Finland
- ⁵⁵ Department of Surgery, Leiden University Medical Center, Leiden, the Netherlands
- ⁵⁶ Department of Basic Sciences, Shaukat Khanum Memorial Cancer Hospital and Research Centre (SKMCH & RC), Lahore, Pakistan

⁵⁷ Laboratory Medicine Program, University Health Network, Toronto, Ontario, Canada

⁵⁸ Institute for Occupational and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

⁵⁹ Institute for Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

⁶⁰ Department of Quantitative Health Sciences, Division of Epidemiology, Mayo Clinic, Rochester, MN, USA

⁶¹ Health Research Institute, Fundación Jiménez Díaz University Hospital, Universidad Autónoma de Madrid (IIS-FJD, UAM), Madrid, Spain

⁶² Department of Oncology, Mayo Clinic, Rochester, MN, USA

⁶³ Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany

⁶⁴ Network Aging Research, University of Heidelberg, Heidelberg, Germany

⁶⁵ Department of Pathology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Amsterdam, the Netherlands

⁶⁶ Servicio de Oncología Médica, Hospital Universitario La Paz, Madrid, Spain

⁶⁷ Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany

⁶⁸ Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Amsterdam, the Netherlands

⁶⁹ Department of Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands

⁷⁰ Department of Computational Biomedicine, Cedars-Sinai Medical Center, West Hollywood, CA, USA

⁷¹ Cancer Epidemiology and Prevention Research Unity, The Institute of Cancer Research and Imperial College London, London, UK

* These authors jointly supervised the work.

**** Corresponding Author:**

Montserrat García-Closas

32 Oakleaf Avenue, Sutton, SM2 5GP, UK

montse.garcia-closas01@icr.ac.uk

Word count: 2965

Abstract

Background. Breast cancer is etiologically heterogeneous; however, it remains unclear which tumor features are linked to specific risk factors. We conducted a large, pooled analysis to evaluate independent, dose-response associations between established risk factors and quantitative tumor features.

Methods. Case-case analyses of 15,731 invasive breast cancers from 24 studies assessed reproductive and hormonal factors, body mass index (BMI), alcohol, smoking, and family history in relation to quantitative immunohistochemistry measures on tissue microarrays (ER, PR, HER2, KI67, TP53) and histological grade, including dose-response trends of heterogeneity. Case-control analyses in 10 population-based studies estimated odds ratios (ORs) for subtype-specific risk. Only noteworthy findings are reported, defined as Bayesian False Discovery Probability (BFDP) <0.2 of heterogeneity after adjustment for tumor features.

Results. Nulliparity and later age at menopause were associated with higher ER-positivity (p -trend=0.021 and 0.001, respectively), with corresponding OR[ER+](95% CI)=1.49(1.16-1.90) for nulliparous vs. parous and 1.07(1.03-1.11) per 5 years. Current combined menopausal hormone therapy (MHT) use was associated with lower grade (p -trend <0.001), with OR[grade1]=3.37(2.69-4.21) for current vs. never users. Higher BMI was associated with lower ER-positivity and higher grade in premenopausal women (p -trend <0.001 and <0.001), with OR[ER+]=0.80(0.74-0.87) and OR[grade1]=0.75(0.63-0.88) per 5 units, and with

higher PR-positivity and higher grade in postmenopausal women ($p\text{-trend} < 0.001$ and < 0.001), with $OR[PR+] = 1.08(1.03-1.14)$ and $OR[grade3] = 1.10(1.04-1.17)$ per 5 units.

Conclusion. This large-scale study showed that nulliparity, age at menopause, MHT, and BMI have independent, dose-response associations with ER, PR, and grade, clarifying patterns of etiologic heterogeneity. Associations with HER2, KI67 and TP53, or other risk factors were not noteworthy.

Introduction

Breast cancer is a heterogeneous disease encompassing multiple subtypes distinguished by their histological, molecular, and genomic profiles.^{1–3} These biological differences, together with distinct age-specific incidence patterns observed for different subtypes,⁴ suggest that breast cancers may arise through distinct causal pathways, a concept known as etiologic heterogeneity.

Mounting evidence supports etiologic heterogeneity for certain risk factors and breast cancer subtypes.^{5–12} The strongest evidence comes from studies of reproductive and hormonal risk factors in relation to tumor subtypes defined by estrogen receptor (ER) status.⁶ Further refinement of subtypes into clinical groupings, such as 'intrinsic-like' subtypes based on specific marker combinations of ER, progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and tumor grade/KI67, has provided additional insights into etiologic heterogeneity beyond ER status alone.^{7,8,13–15} A recent scoping review focusing on ER and triple-negative status concluded that there is convincing evidence for heterogeneity by ER status for parity, and suggestive evidence for age at first birth, oral contraceptive (OC) use, menopausal hormone therapy (MHT), and both pre- and postmenopausal body mass index (BMI), with little or no evidence for lifestyle risk factors such as alcohol and smoking.⁶ There is also limited evidence to suggest that grade, KI67, and TP53 status may be important markers of etiologic heterogeneity for reproductive, hormonal, or lifestyle risk factors.^{9,16}

The lack of clear evidence of etiologic heterogeneity for many breast cancer risk factors may reflect differences among studies in the definition of tumor subtypes and the limited evaluation of the independent contributions of individual tumor

features in cancer etiology while accounting for their correlations.^{16–18} Clarifying how distinct tumor features relate to specific risk factor profiles could provide insights into the biological mechanisms by which exposures influence carcinogenesis and tumor heterogeneity. In addition, most studies of etiologic heterogeneity use dichotomized definitions of immunohistochemical (IHC) markers indicated for clinical decision-making (e.g., ER, PR, and HER2 status), without evaluating dose-response relationships with risk factors that can strengthen the evidence for associations.

To address these potential limitations in the study of etiologic heterogeneity, a pooled analysis was conducted within the Breast CAncer STRatification (B-CAST) project of the Breast Cancer Association Consortium (BCAC), encompassing 24 case-control studies. Associations between established reproductive, hormonal, and lifestyle breast cancer risk factors and quantitative, standardized measures of five IHC markers (ER, PR, HER2, KI67, and TP53) from tumor microarrays (TMAs), and tumor grade, were assessed. By modeling multiple tumor features simultaneously, we aimed to identify independent and dose-response relationships in etiologic heterogeneity. These analyses can improve our understanding of how risk factors relate to specific breast tumor features and potentially inform prevention strategies tailored to their etiological pathways.

Methods

Study population. This study included 19,140 invasive breast cancer cases and 56,267 controls from 24 studies conducted within BCAC. Cases were eligible for inclusion only if TMA cores were available. Following quality control procedures (**Supplementary Methods**), 18,553 cases were retained for analysis. Case-control analyses were restricted to population-based studies to mitigate selection and

information bias; this subset included 10 of the 24 studies, comprising 6,050 cases and 27,437 controls. The relevant institutional review boards granted approval for all included studies, and written informed consent was obtained from all participants. Contributing studies are detailed in **Table S1**.

Tumor marker quantification. TMAs including tissue cores from breast tumors of cases were stained with IHC for ER, PR, HER2, KI67, and TP53. A summary of available cores by study and marker is provided in **Table S1**. Digital image analysis (HALO®, Indica Labs, Albuquerque, NM) quantified IHC-positive and IHC-negative nuclei per core. For each participant, the proportion of marker-positive cells (henceforth, marker-positivity) was calculated as the ratio of positive to total nuclei summed across all available cores. For dichotomous analyses, ER, PR, KI67, and TP53 were classified positive using a threshold of $\geq 10\%$ positive nuclei, and HER2 using a threshold of $\geq 2.3\%$ that best discriminated clinical HER2 status.

Statistical analyses. The primary outcome measures were tumor marker expression derived from TMA scoring of IHC markers and tumor grade (1, 2, 3) obtained from clinical records. Marker expression was analyzed both as a continuous and a dichotomous measure (positive vs. negative).

The following risk factors were evaluated at age at reference date (interview for controls and diagnosis for cases): age at menarche (per 2 year increase), parity status (parous [ref.], nulliparous), age at first birth (per 5 year increase), lifetime breastfeeding duration (per 6 month decrease), menopause status (postmenopausal [ref.], premenopausal), age at menopause (per 5 year increase), BMI (per 5 unit increase) with an interaction with menopausal status, alcohol use (never [ref.], ever), age at smoking initiation (never [ref.], ≤ 16 years of age, 17 - 19 years of age, ≥ 20

years of age), family history of breast cancer in a first-degree relative (absent [ref.], present), estrogen-only MHT use, combined MHT use, and OC use (never [ref.], former, current). Summary statistics and prevalence of missing data for each risk factor are presented in **Table S2**, with further modeling considerations detailed in the **Supplementary Methods**.

In case-case analyses to assess heterogeneity, beta-binomial regression estimated ratios ($\exp[\beta]$) of marker-positivity and dose-response trends (p-trend of heterogeneity). For dichotomous markers and tumor grade, dichotomous and ordinal logistic regression models, respectively, estimated case-case odds ratios (OR) of heterogeneity. For each tumor feature, two models were fitted: (i) an age- and study-adjusted model, and (ii) a model additionally adjusted for the other categorical tumor features.

In case-control analyses, multinomial logistic regression estimated ORs for subtype-specific risks, with outcomes defined by dichotomous markers and grade with controls as the reference. These models were not adjusted for other tumor features.

All models included all risk factors and were adjusted for study and age at reference date. Estimates were reported with 95% confidence intervals (CI), and tests of significance were two-sided ($p < 0.05$). To account for false discoveries arising from multiple comparisons in case-case analyses, the Bayesian False Discovery Probability (BFDP) was calculated.¹⁹ Associations with BFDP < 0.2 were considered noteworthy. Analyses were performed using R version 4.4.0 (R Core Team, Vienna, Austria). Details are provided in the **Supplementary Methods**.

Results

Each case contributed approximately two cores per tumor marker. The number of cases and available tissue cores for each marker are summarized in **Table 1**, with further details by study in **Table S1**.

This section highlights noteworthy findings (BFDP < 0.2) from case-case analyses of heterogeneity adjusted for all tumor features, unless otherwise specified, with corresponding case-control analyses for subtype-specific risks. Plots illustrating the relationship between select risk factors and continuous marker expression are provided in **Figure S1**. A complete summary of all results is provided in **Tables S3-S5**.

Reproductive history. In case-case analyses, nulliparous compared with parous women had higher ER-positivity ($\exp(\beta)$ [95% CI] = 1.08 [1.01-1.16]; p-trend of heterogeneity = 0.02; **Figures 1, 2; Table S3**) and were more likely to have ER-positive than ER-negative tumors (OR of heterogeneity = 1.80 [1.23-2.64]; **Figures 3, 4; Table S4**). Corresponding case-control analyses showed that nulliparous women were at higher risk of developing ER-positive (OR = 1.49 [1.16-1.90]), while suggesting lower risk of ER-negative tumors (OR = 0.60 [0.36-0.99]; **Figure 5; Table S5**).

After adjustment for tumor features, women experiencing a later menopause had higher ER-positivity ($\exp(\beta)$ = 1.02 [1.01-1.03] per 5 years; p-trend = 0.001; **Figures 1, 2; Table S4**) and were more likely to have ER-positive tumors (OR = 1.12 [1.05-1.20] per 5 years; **Figures 3, 4; Table S4**). Case-control analyses showed that women with later menopause were at higher risk of ER-positive (OR = 1.07 [1.03-

1.11] per 5 years), while suggesting lower risk of ER-negative tumors (OR = 0.93 [0.87-1.00] per 5 years; **Figure 5; Table S5**).

Exogenous hormone use. In case-case analyses, women using combined MHT were less likely to have higher-grade tumors compared with never users (OR: 0.60 [0.51-0.72]; p-trend < 0.001; **Figures 3, 4; Table S4**). Case-control analyses showed that women using combined MHT were at a 3-fold higher risk of grade 1 tumors (OR = 3.37 [2.69-4.21]) and a nearly 2-fold higher risk of grade 2 tumors (OR = 1.78 [1.53-2.07]), but not grade 3 tumors (OR = 1.15 [0.92-1.45]; **Figure 5; Table S5**). Combined MHT use was also associated with lower KI67-positivity in unadjusted analyses, but this association was not noteworthy after adjustment for other tumor features (**Figures 1, 2; Table S3**).

Further analyses were performed to clarify the association of combined MHT use in relation to tumors cross-classified by ER and grade. The results confirmed that the association varied across grade, particularly among ER-positive tumors, with little evidence of heterogeneity by ER status alone (**Figure S2**).

Body mass index. In case-case analyses, higher BMI was associated with higher-grade tumors both among premenopausal (OR = 1.10 [1.04-1.17] per 5 units; p-trend < 0.001; **Figures 3, 4; Table S4**) and postmenopausal women (OR = 1.11 [1.07-1.16] per 5 units; p-trend < 0.001; **Figures 3, 4; Table S4**). Among premenopausal women, case-control analyses showed that higher BMI was inversely associated with the risk of grade 1 tumors (OR = 0.75 [0.63-0.88] per 5 units) and grade 2 tumors (OR = 0.83 [0.76-0.91] per 5 units), but not grade 3 tumors (OR = 0.94 [0.85-1.04] per 5 units; **Figure 5; Table S5**). In contrast, case-control analyses among postmenopausal women showed that higher BMI was associated

with a higher risk of grade 3 tumors (OR = 1.10 [1.04-1.17] per 5 units) but not grade 2 tumors (OR = 1.04 [1.00-1.09]), while suggesting a decreased risk of grade 1 tumors (OR = 0.91 [0.84-0.98]; **Figure 5; Table S5**).

In case-case analyses, higher BMI among premenopausal women was associated with lower ER-positivity ($\exp(\beta) = 0.96$ [0.94-0.98] per 5 units; $p\text{-trend} < 0.001$; **Figures 1, 2; Table S3**) and a lower likelihood of having ER-positive tumors (OR = 0.88 [0.81-0.95] per 5 units; **Figures 3, 4; Table S4**). Case-control analyses showed this corresponded to a decreased risk of ER-positive tumors (OR = 0.80 [0.74-0.87] per 5 units), but not ER-negative tumors (OR = 1.05 [0.92-1.19] per 5 units; **Figure 5; Table S5**). Unadjusted case-case analyses suggested an association between higher BMI and PR-positive tumors in premenopausal women, but this association was attenuated and not noteworthy after adjustment for other tumor features (OR = 0.97 [0.90-1.04] per 5 units; **Figures 3, 4; Table S4**).

In case-case analyses, higher BMI among postmenopausal women was associated with higher PR-positivity ($\exp(\beta) = 1.05$ [1.03-1.06] per 5 units; $p\text{-trend} < 0.001$; **Figures 1, 2; Table S3**) and a higher likelihood of having PR-positive tumors (OR = 1.12 [1.06-1.19] per 5 units; **Figures 3, 4; Table S4**). Case-control analyses showed this corresponded to a higher risk of PR-positive tumors (OR = 1.08 [1.03-1.14] per 5 units), but not PR-negative tumors (OR = 0.99 [0.94-1.05] per 5 units; **Figure 5; Table S5**).

Alcohol use and cigarette smoking. In case-case analyses, neither alcohol consumption nor cigarette smoking demonstrated noteworthy associations with any examined tumor features (**Figures 1-4; Tables S3, S4**).

Family history of breast cancer in first-degree relatives. In unadjusted case-case analyses, women with positive family history had higher ER-positivity (unadjusted $\exp(\beta) = 1.05$ [1.02-1.08]; $p\text{-trend} < 0.001$), but the association was not noteworthy after adjustment ($\exp(\beta) = 1.03$ [1.00-1.05]; $p\text{-trend} = 0.03$; **Figures 1, 2; Table S3**).

In analyses adjusted for other tumor features, no noteworthy associations with risk factors were observed for TP53, KI67, or HER2.

Discussion

This large, pooled analysis of 24 studies, including up to 15,731 cases of invasive breast cancer and 27,437 controls, provides robust evidence clarifying which risk factors are independently associated with specific tumor features. The strongest evidence for heterogeneity was observed for nulliparity, age at menopause, combined MHT use, and BMI in relation to subtypes defined by ER, PR, and grade. No additional evidence for heterogeneity meeting the stringent BFDP < 0.2 threshold was found for HER2, KI67, or TP53 after adjustment for other tumor features, or for other risk factors. Associations not meeting this threshold require further confirmation.

While previous BCAC studies^{8,9,15,17} applied standardized IHC staining and scoring on TMAs, they included fewer tumor markers and smaller sample sizes. Here, we expanded both the marker panel and the sample size, improving consistency, comparability, and statistical power. This enabled simultaneous evaluation of multiple tumor features and dose-response relationships using quantitative measures. Assessing TP53, KI67, and tumor grade, alongside the more

commonly studied ER, PR, and HER2, allowed investigation of biologically relevant but relatively understudied features in relation to breast cancer etiology. By focusing on noteworthy associations (BFDP < 0.2), we provide clear and robust evidence for specific associations, with the full results presented in **Tables S3-S5**.

A recent systematic review by Hurson et al. of multiple risk factors and ER status concluded, after expert evaluation, that there was convincing evidence of etiologic heterogeneity by ER status only for parity, and suggestive evidence for age at first birth, BMI in pre- and postmenopausal women, OC, and MHT use.⁶ Our analyses showed an association between parity and ER status consistent in direction and magnitude with this review. Compared with parous women, nulliparous women had a 1.5-fold higher risk of developing ER-positive tumors and a suggestion for a lower risk of ER-negative tumors, with a dose-response relationship between nulliparity and higher ER-positivity in case-case analyses. While associations of ER with age at first birth were significant at $p < 0.05$, they did not meet our stringent noteworthiness threshold.

Higher BMI in premenopausal women has been linked to lower breast cancer risk predominantly for ER and/or PR-positive tumors and luminal intrinsic subtypes.^{6,10,18,20} Our results confirm this observation and further demonstrate a dose-response relationship with higher ER-positivity. Beyond ER, BMI was independently associated with lower tumor grade in a dose-response manner: a 5-unit increase in BMI was associated with a 25% and 17% lower risk of grade 1 and 2 tumors, respectively, and showed no association with grade 3 tumors. This pattern is consistent with BMI-associated reductions in circulating sex hormones in premenopausal women,²¹ which may preferentially influence ER-positive tumors,²² and suggests an additional role for grade.

Among postmenopausal women, higher BMI is linked to higher breast cancer risk, predominantly of hormone receptor-positive tumors.^{6,18} After adjustment for multiple tumor features, our analyses show that PR plays a more prominent role than ER. A 5-unit increase in BMI was associated with an 8% higher risk of PR-positive tumors, with no further contribution by ER. This finding is consistent with previous meta-analyses,^{17,18} although they included some studies used in our analyses. A 5-unit increase in BMI was also associated with a 10% higher risk of developing grade 3 tumors with the suggestion of an inverse association with grade 1 tumors, which was consistent with previous BCAC analyses including fewer participants.⁹ While underlying mechanisms remain unclear, our findings are consistent with leptin signaling and chronic inflammation pathways, which have been associated with higher grade,^{23,24} and estrogen-independent promotion of PR-positive tumors.²⁵

Although Hurson et al. reported suggestive evidence for heterogeneity by ER status for OC and postmenopausal MHT use, these were not noteworthy in our analyses. In contrast, tumor grade showed clear heterogeneity for combined MHT use: current users had a 3.4-fold and 1.8-fold higher risk of grade 1 and 2 tumors, and showed no association with grade 3 tumors. These findings are consistent with previous BCAC analyses^{9,15} and the broader literature associating MHT-related breast cancers with lower grade, as well as lobular and tubular histological subtypes.^{26–28} Our data indicated that the association with grade is not fully explained by histology, as it was present across all histologic subtypes (**Figures S2, S3**).

Hurson et al reported inconclusive evidence for heterogeneity by ER status in relation to age at menopause.⁶ However, we provide noteworthy evidence for a 7% higher risk of ER-positive tumors per 5-year increase in menopausal age, and no

association with ER-negative tumors. These findings support an estrogen-related biological mechanism for this association.

Associations with TP53 IHC were not noteworthy for any of the risk factors evaluated, consistent with a previous case-case study.¹⁶ However, that study found significant associations with several risk factors and TP53 mutant-like status defined by an RNA expression signature, rather than IHC. These observations suggest that IHC-defined TP53 status may not capture functional TP53 pathway activity^{16,29} and future studies should focus on TP53 expression profiling rather than IHC.

Most study participants were of White/European ancestry, which may limit the generalizability of our findings to other populations. Hurson et al. showed little evidence of differences in ER-related etiologic heterogeneity by race and ethnic groups,⁶ suggesting that findings might be applicable across population groups. Additionally, studies were primarily drawn from the general population with unknown germline mutation status (e.g., BRCA1/2, TP53), requiring future studies to determine whether our findings apply to high-risk mutation carriers.

Despite the large sample size, missing data on tumor markers and risk factors limited statistical power for some analyses. Case-control analyses were limited to population-based studies, including those nested within prospective cohorts, as they are less prone to selection biases than other designs, which reduced sample size and power. Although our study benefits from standardized staining and automated scoring of TMAs across studies, other factors such as tissue preparation could not be standardized. This type of measurement error could have resulted in underestimation of etiologic heterogeneity since it is likely non-differential with respect to risk factors. Furthermore, the breadth of this analysis precluded detailed

assessment of specific risk factors, such as timing of pregnancies or effect modification of MHT by BMI, in spite of prior evidence suggesting that finer analyses could further clarify etiologic heterogeneity.^{15,30}

In conclusion, by jointly evaluating multiple tumor features and risk factors at unprecedented scale, we showed that ER, PR, and grade are the principal drivers of etiologic heterogeneity. These findings should be prioritized in studies evaluating specific exposures, e.g., by incorporating grade in studies of BMI or MHT in postmenopausal women, thereby refining risk estimates, guiding mechanistic studies, and informing risk-stratified prevention.

Acknowledgements

We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. ABCS thanks research pathologist Joyce Sanders for her work in the B-CAST TMA project. CNIO-BCS thanks Guillermo Pita, Charo Alonso, Nuria Álvarez, Pilar Zamora, Primitiva Menendez, the Human Genotyping-CEGEN

Unit (CNIO). Investigators from the CPS-II cohort thank the participants and Study Management Group for their invaluable contributions to this research. They also acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention National Program of Cancer Registries, as well as cancer registries supported by the National Cancer Institute Surveillance Epidemiology and End Results program. We thank the participants and the investigators of EPIC (European Prospective Investigation into Cancer and Nutrition). ESTHER thanks Hartwig Ziegler, Sonja Wolf, Volker Hermann, Christa Stegmaier, Katja Butterbach. KBCP thanks Eija Myöhänen. kConFab/AOCS wish to thank Heather Thorne, Eveline Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, and the Clinical Follow Up Study (which has received funding from the NHMRC, the National Breast Cancer Foundation, Cancer Australia, and the National Institute of Health (USA)) for their contributions to this resource, and the many families who contribute to kConFab. MARIE thanks Petra Seibold, Nadia Obi, Sabine Behrens, Ursula Eilber and Muhabbet Celik. We thank the coordinators, the research staff and especially the MMHS participants for their continued collaboration on research studies in breast cancer. The OFBCR thanks Teresa Selander, Nayana Weerasooriya and Steve Gallinger. ORIGO thanks E. Krol-Warmerdam, and J. Blom for patient accrual, administering questionnaires, and managing clinical information. ORIGO thanks E. Krol-Warmerdam, and J. Blom for patient accrual, administering questionnaires, and managing clinical information. PBCS thanks Louise Brinton, Mark Sherman, Neonila Szeszenia-Dabrowska, Beata Peplonska, Witold Zatonski, Pei Chao, Michael Stagner. The ethical approval for the POSH study is MREC /00/6/69, UKCRN ID: 1137. We thank staff in the Experimental Cancer Medicine Centre (ECMC)

supported Faculty of Medicine Tissue Bank and the Faculty of Medicine DNA Banking resource. The authors wish to acknowledge the roles of the Breast Cancer Now Tissue Bank in collecting and making available the samples and/or data, and the patients who have generously donated their tissues and shared their data to be used in the generation of this publication. The RBCS thanks Jannet Blom, Saskia Pelders, Wendy J.C. Prager – van der Smissen, and the Erasmus MC Family Cancer Clinic. SBCS thanks Sue Higham, Helen Cramp, Dan Connley, Sabapathy Balasubramanian, Esther Ludbrook, Alice Halsey and Malcolm W.R. Reed. We thank the SEARCH and EPIC teams. SGBCC thanks the participants and all research coordinators for their excellent help with recruitment, data and sample collection. SKKDKFZS thanks all study participants, clinicians, family doctors, researchers and technicians for their contributions and commitment to this study. UBCS thanks all study participants as well as the ascertainment, laboratory, analytics and informatics teams at Huntsman Cancer Institute and Intermountain Healthcare for their important contributions to this study. UCIBCS thanks Irene Masunaka. UKBGS thanks Breast Cancer Now and the Institute of Cancer Research for support and funding of the Generations Study, and the study participants, study staff, and the doctors, nurses and other health care providers and health information sources who have contributed to the study. We acknowledge NHS funding to the Royal Marsden/ICR NIHR Biomedical Research Centre. The authors used artificial intelligence (Gemini) to improve grammar and readability of the text. The content was carefully verified by the authors, who take full responsibility for the work.

Funding

This work was supported by the European Union's Horizon 2020 Research and Innovation Programme grant 633784 (B-CAST). BCAC was supported by the Cancer Research UK grant: PPRPGM-Nov20\100002 and by core funding from the NIHR Cambridge Biomedical Research Centre (NIHR203312). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. Additional funding for BCAC was provided by the Confluence project which is funded with intramural funds from the National Cancer Institute Intramural Research Program, National Institutes of Health, the European Union's Horizon 2020 Research and Innovation Programme (grant numbers 634935 and 633784 for BRIDGES and B-CAST respectively), and the PERSPECTIVE I&I project, funded by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the Ministère de l'Économie et de l'Innovation du Québec through Genome Québec, the Quebec Breast Cancer Foundation. The EU Horizon 2020 Research and Innovation Programme funding source had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The ABCS study was supported by the Dutch Cancer Society [grants NKI 2007-3839; 2009 4363] and an institutional grant of the Dutch Cancer Society and of the Dutch Ministry of Health, Welfare and Sport. The CNIO-BCS was supported by the Instituto de Salud Carlos III, the Red Temática de Investigación Cooperativa en Cáncer and grants from the Asociación Española Contra el Cáncer and the Fondo de Investigación Sanitario (PI11/00923 and PI12/00070). The American Cancer Society funds the creation, maintenance, and updating of the CPS-II cohort. The coordination of EPIC is financially supported by the European Commission (DG-SANCO) and the International Agency for Research on Cancer.

The national cohorts are supported by: Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Éducation Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), Federal Ministry of Education and Research (BMBF) (Germany); the Hellenic Health Foundation, the Stavros Niarchos Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS), PI13/00061 to Granada, PI13/01162 to EPIC-Murcia, Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, ISCIII RETIC (RD06/0020) (Spain); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 and C8221/A19170 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk, MR/M012190/1 to EPIC-Oxford) (United Kingdom). The ESTHER study was supported by a grant from the Baden Württemberg Ministry of Science, Research and Arts. The GESBC was supported by the Deutsche Krebshilfe e. V. [70492] and the German Cancer Research Center (DKFZ). GLACIER was supported by Breast Cancer Now, CRUK and Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London. The KBCP was financially supported by the special Government Funding (VTR) of Kuopio University Hospital grants, Cancer Fund of North Savo, the Finnish Cancer Organizations, and by the strategic funding of the University of Eastern Finland. kConFab is supported by a grant from the National Breast Cancer Foundation, and previously by the National Health and Medical Research Council (NHMRC), the Queensland Cancer Fund, the Cancer Councils of

New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia. The MARIE study was supported by the Deutsche Krebshilfe e.V. [70-2892-BR I, 106332, 108253, 108419, 110826, 110828], the Hamburg Cancer Society, the German Cancer Research Center (DKFZ) and the Federal Ministry of Education and Research (BMBF) Germany [01KH0402]. The MCBCS was supported by the NIH grants R35CA253187, R01CA192393, R01CA116167, R01CA176785 a NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer [P50CA116201], and the Breast Cancer Research Foundation. The MMHS study was supported by NIH grants CA97396, CA128931, CA116201, CA140286 and CA177150. MSKCC is supported by grants from the Breast Cancer Research Foundation and Robert and Kate Niehaus Clinical Cancer Genetics Initiative. The Northern California Breast Cancer Family Registry (NC-BCFR) and Ontario Familial Breast Cancer Registry (OFBCR) were supported by grant U01CA164920 from the USA National Cancer Institute of the National Institutes of Health. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR. The ORIGO study was supported by the Dutch Cancer Society (RUL 1997-1505) and the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL CP16). The PBCS was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. The PLCO is supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, National Institutes of Health. The

POSH study is funded by Cancer Research UK (grants C1275/A11699, C1275/C22524, C1275/A19187, C1275/A15956 and Breast Cancer Campaign 2010PR62, 2013PR044. The RBCS was funded by the Dutch Cancer Society (DDHK 2004-3124, DDHK 2009-4318). The SBCS was supported by Sheffield Experimental Cancer Medicine Centre and Breast Cancer Now Biobank. SEARCH was funded by Cancer Research UK [C490/A10124, C490/A16561] and supported by the UK National Institute for Health Research Biomedical Research Centre at the University of Cambridge. SGBCC is funded by the National Research Foundation Singapore, NUS start-up Grant, National University Cancer Institute Singapore (NCIS) Centre Grant, Breast Cancer Prevention Programme, Asian Breast Cancer Research Fund and the NMRC Clinician Scientist Award (SI Category). SKKDKFZS is supported by the DKFZ. The SZBCS was supported by Grant PBZ_KBN_122/P05/2004 and the program of the Minister of Science and Higher Education under the name "Regional Initiative of Excellence" in 2019-2022 project number 002/RID/2018/19 amount of financing 12 000 000 PLN. UBCS was supported by funding from National Cancer Institute (NCI) grant R01 CA163353 (to N.J. Camp) and the Women's Cancer Center at the Huntsman Cancer Institute (HCI). Data collection for UBCS was supported by the Utah Population Database (UPDB) and Utah Cancer Registry (UCR). The UPDB is supported by HCI, the University of Utah, and NCI grant P30 CA042014. The UCR is funded by the NCI's SEER Program, Contract No. HHSN261201800016I, the US Centers for Disease Control and Prevention's National Program of Cancer Registries, Cooperative Agreement No. NU58DP007131, with additional support from the University of Utah and Huntsman Cancer Foundation. The UKBGS is funded by Breast Cancer Now and the Institute of Cancer Research (ICR), London. ICR acknowledges NHS

funding to the NIHR Biomedical Research Centre. This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH). The contributions of the NIH authors were made as part of their official duties as NIH federal employees, are in compliance with agency policy requirements, and are considered Works of the United States Government. However, the findings and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services.

Conflict of Interest

The authors declare no conflicts of interest.

Data Availability

The data underlying this article cannot be publicly shared in accordance with ethical guidelines regarding participant confidentiality. Access may be provided by the corresponding author upon specific request, subject to approval by the Institutional Review Board. The analytical code underlying this article has been archived in the Zenodo repository at the following URL:

<https://doi.org/10.5281/zenodo.18630325>.³¹

References

1. Dawson S, Rueda OM, Aparicio S, Caldas C. A new genome- driven integrated classification of breast cancer and its implications. *EMBO J*. 2013;32(5):617-628. doi:10.1038/emboj.2013.19
2. Koboldt DC, Fulton RS, McLellan MD, et al. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490(7418):61-70. doi:10.1038/nature11412

3. Nik-Zainal S, Davies H, Staaf J, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*. 2016;534(7605):47-54. doi:10.1038/nature17676
4. Anderson WF, Rosenberg PS, Prat A, Perou CM, Sherman ME. How Many Etiological Subtypes of Breast Cancer: Two, Three, Four, Or More? *JNCI J Natl Cancer Inst*. 2014;106(8):dju165. doi:10.1093/jnci/dju165
5. Mao X, Omeogu C, Karanth S, et al. Association of reproductive risk factors and breast cancer molecular subtypes: a systematic review and meta-analysis. *BMC Cancer*. 2023;23(1):644. doi:10.1186/s12885-023-11049-0
6. Hurson AN, Ahearn TU, Koka H, et al. Risk factors for breast cancer subtypes by race and ethnicity: a scoping review. *J Natl Cancer Inst*. 2024;116(12):1992-2002. doi:10.1093/jnci/djae172
7. Gaudet MM, Gierach GL, Carter BD, et al. Pooled Analysis of Nine Cohorts Reveals Breast Cancer Risk Factors by Tumor Molecular Subtype. *Cancer Res*. 2018;78(20):6011-6021. doi:10.1158/0008-5472.CAN-18-0502
8. Jung AY, Ahearn TU, Behrens S, et al. Distinct Reproductive Risk Profiles for Intrinsic-Like Breast Cancer Subtypes: Pooled Analysis of Population-Based Studies. *J Natl Cancer Inst*. 2022;114(12):1706-1719. doi:10.1093/jnci/djac117
9. Abubakar M, Chang-Claude J, Ali HR, et al. Etiology of hormone receptor positive breast cancer differs by levels of histologic grade and proliferation. *Int J Cancer*. 2018;143(4):746-757. doi:10.1002/ijc.31352
10. Premenopausal Breast Cancer Collaborative Group, Schoemaker MJ, Nichols HB, et al. Association of Body Mass Index and Age With Subsequent Breast Cancer Risk in Premenopausal Women. *JAMA Oncol*. 2018;4(11):e181771. doi:10.1001/jamaoncol.2018.1771
11. Gaudet MM, Carter BD, Brinton LA, et al. Pooled analysis of active cigarette smoking and invasive breast cancer risk in 14 cohort studies. *Int J Epidemiol*. 2017;46(3):881-893. doi:10.1093/ije/dyw288
12. Zhang H, Ahearn TU, Lecarpentier J, et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nat Genet*. 2020;52(6):572-581. doi:10.1038/s41588-020-0609-2
13. Klintman M, Rosendahl AH, Randeris B, et al. Postmenopausal overweight and breast cancer risk; results from the KARMA cohort. *Breast Cancer Res Treat*. 2022;196(1):185-196. doi:10.1007/s10549-022-06664-7
14. Tamimi RM, Colditz GA, Hazra A, et al. Traditional Breast Cancer Risk Factors in Relation to Molecular Subtypes of Breast Cancer. *Breast Cancer Res Treat*. 2012;131(1):159-167. doi:10.1007/s10549-011-1702-0

15. Le Cornet C, Jung A, Behrens S, Pooja M, Chang-Claude J. Exogenous hormones, tumor intrinsic subtypes, and breast cancer. *JAMA Netw Open*. Published online 2025.
16. Hurson AN, Abubakar M, Hamilton AM, et al. TP53 Pathway Function, Estrogen Receptor Status, and Breast Cancer Risk Factors in the Carolina Breast Cancer Study. *Cancer Epidemiol Biomarkers Prev*. 2022;31(1):124-131. doi:10.1158/1055-9965.EPI-21-0661
17. Yang XR, Chang-Claude J, Goode EL, et al. Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies. *J Natl Cancer Inst*. 2011;103(3):250-263. doi:10.1093/jnci/djq526
18. Suzuki R, Orsini N, Saji S, Key TJ, Wolk A. Body weight and incidence of breast cancer defined by estrogen and progesterone receptor status--a meta-analysis. *Int J Cancer*. 2009;124(3):698-712. doi:10.1002/ijc.23943
19. Wakefield J. A Bayesian Measure of the Probability of False Discovery in Genetic Epidemiology Studies. *Am J Hum Genet*. 2007;81(2):208-227.
20. Torres-de la Roche LA, Steljes I, Janni W, Friedl TWP, De Wilde RL. The Association between Obesity and Premenopausal Breast Cancer According to Intrinsic Subtypes – a Systematic Review. *Geburtshilfe Frauenheilkd*. 2020;80(6):601-610. doi:10.1055/a-1170-5004
21. Tworoger SS, Eliassen AH, Missmer SA, et al. Birthweight and Body Size throughout Life in Relation to Sex Hormones and Prolactin Concentrations in Premenopausal Women. *Cancer Epidemiol Biomarkers Prev*. 2006;15(12):2494-2501. doi:10.1158/1055-9965.EPI-06-0671
22. Eliassen AH, Missmer SA, Tworoger SS, et al. Endogenous steroid hormone concentrations and risk of breast cancer among premenopausal women. *J Natl Cancer Inst*. 2006;98(19):1406-1415. doi:10.1093/jnci/djj376
23. Garofalo C, Koda M, Cascio S, et al. Increased expression of leptin and the leptin receptor as a marker of breast cancer progression: possible role of obesity-related stimuli. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2006;12(5):1447-1453. doi:10.1158/1078-0432.CCR-05-1913
24. Danforth DN. The Role of Chronic Inflammation in the Development of Breast Cancer. *Cancers*. 2021;13(15):3918. doi:10.3390/cancers13153918
25. Giles ED, Wellberg EA, Astling DP, et al. Obesity and overfeeding affecting both tumor and systemic metabolism activates the progesterone receptor to contribute to postmenopausal breast cancer. *Cancer Res*. 2012;72(24):6490-6501. doi:10.1158/0008-5472.CAN-12-1653
26. Rosenberg LU, Granath F, Dickman PW, et al. Menopausal hormone therapy in relation to breast cancer characteristics and prognosis: a cohort study. *Breast Cancer Res*. 2008;10(5):R78. doi:10.1186/bcr2145

27. Sener SF, Winchester DJ, Winchester DP, et al. The effects of hormone replacement therapy on postmenopausal breast cancer biology and survival. *Am J Surg*. 2009;197(3):403-407. doi:10.1016/j.amjsurg.2008.11.014
28. Reeves GK, Beral V, Green J, Gathani T, Bull D, Million Women Study Collaborators. Hormonal therapy for menopause and breast-cancer risk by histological type: a cohort study and meta-analysis. *Lancet Oncol*. 2006;7(11):910-918. doi:10.1016/S1470-2045(06)70911-1
29. Shahbandi A, Nguyen HD, Jackson JG. TP53 Mutations and Outcomes in Breast Cancer: Reading beyond the Headlines. *Trends Cancer*. 2020;6(2):98-110. doi:10.1016/j.trecan.2020.01.007
30. Brouckaert O, Rudolph A, Laenen A, et al. Reproductive profiles and risk of breast cancer subtypes: a multi-center case-only study. *Breast Cancer Res BCR*. 2017;19:119. doi:10.1186/s13058-017-0909-3
31. Adams D. UK-Generations-Study/clarifying-etiological-heterogeneity: Manuscript Submission. Published online February 13, 2026. doi:10.5281/zenodo.18630325

Tables and Figure Legends

Table 1. Number of contributing studies, breast cancer cases, and stained tissue cores for each tumor marker used in analyses after quality control.

Cores denotes the total number of stained Tissue Microarray (TMA) cores linked to participants; Cases denotes the number of unique participants with available data; Cores/case denotes the mean number of cores per participant for each marker. See **Supplementary Table 1** for counts by study.

Marker	Studies	TMA cores	Cases*	Average TMA cores per case
ER	22	32867	13931	2.36
PR	21	32372	14997	2.16
HER2	22	35668	15288	2.33
TP53	23	28832	14672	1.97
KI67	24	30997	15731	1.97

* The number of cases per marker is lower than the total number of cases as no single marker was stained across all cases.

Figure 1. Heatmap illustrating the direction and strength of evidence for case-case associations between risk factors and tumor features defined by the proportion of marker-positive cells. Cell color indicates the direction of association measured by ratios of the proportion of marker-positive cells, $\exp(\beta)$: orange for $\exp(\beta) > 1$, blue for $\exp(\beta) < 1$. Shading intensity reflects the Bayesian False Discovery Probability (BFDP): white indicates $\text{BFDP} \geq 0.2$ and darker shading indicates noteworthy associations with $\text{BFDP} < 0.2$. **Panel A:** Models estimate associations between each risk factor and individual tumor features without adjustment for other tumor features; **Panel B:** Models estimate associations between each risk factor and tumor features with adjustment for all other tumor features. See **Table S3** for detailed parameter estimates.

<Figure1.pdf>

Footnote: BMI = Body Mass Index; OC = Oral Contraceptive; MHT = Menopausal Hormone Therapy

Figure 2. Case-case $\exp(\beta)$ with 95% confidence intervals for associations between risk factors and tumor features defined by the proportion of marker-positive cells. Only risk factors with at least one noteworthy (BFDP < 0.2)

association in case-case analyses (shown in **Figure 1**) are presented here.

Associations meeting this noteworthiness threshold are shown as solid shapes, while all other associations are shown as open shapes. Diamonds depict estimates from models without adjustment for other tumor features, whereas squares depict estimates from models adjusted for other tumor features. Results were estimated from beta-binomial regression models with a logarithmic link. See **Table S3** for detailed parameter estimates.

<Figure2.pdf>

Footnote: BMI = Body Mass Index; MHT = Menopausal Hormone Therapy

Figure 3. Heatmap illustrating the direction and strength of evidence for case-case associations between risk factors and tumor features defined by marker status and tumor grade. Cell color indicates the direction of association measured by odds ratios (OR): orange for $OR > 1$, blue for $OR < 1$. Shading intensity reflects the Bayesian False Discovery Probability (BFDP): white indicates $BFDP \geq 0.2$ and darker shading indicates noteworthy associations with $BFDP < 0.2$. **Panel A:** Models estimate associations between each risk factor and individual tumor features without adjustment for other tumor features; **Panel B:** Models estimate associations between each risk factor and tumor features with adjustment for all other tumor features. Results were estimated from dichotomous and ordinal logistic regression models. See **Table S4** for detailed parameter estimates.

<Figure3.pdf>

Footnote: BMI = Body Mass Index; OC = Oral Contraceptive; MHT = Menopausal Hormone Therapy

Figure 4. Case-case odds ratios (ORs) with 95% confidence intervals for associations between risk factors and tumor features defined by marker status and tumor grade. Only risk factors with at least one noteworthy (BFDP < 0.2) association in case-case analyses (shown in **Figure 3**) are presented here. Associations meeting this noteworthiness threshold are shown as solid shapes, while all other associations are shown as open shapes. Diamonds depict estimates from models without adjustment for other tumor features, whereas squares depict estimates from models adjusted for other tumor features. Results were estimated from dichotomous and ordinal logistic regression models. See **Table S4** for detailed parameter estimates.

<Figure4.pdf>

Footnote: BMI = Body Mass Index; MHT = Menopausal Hormone Therapy

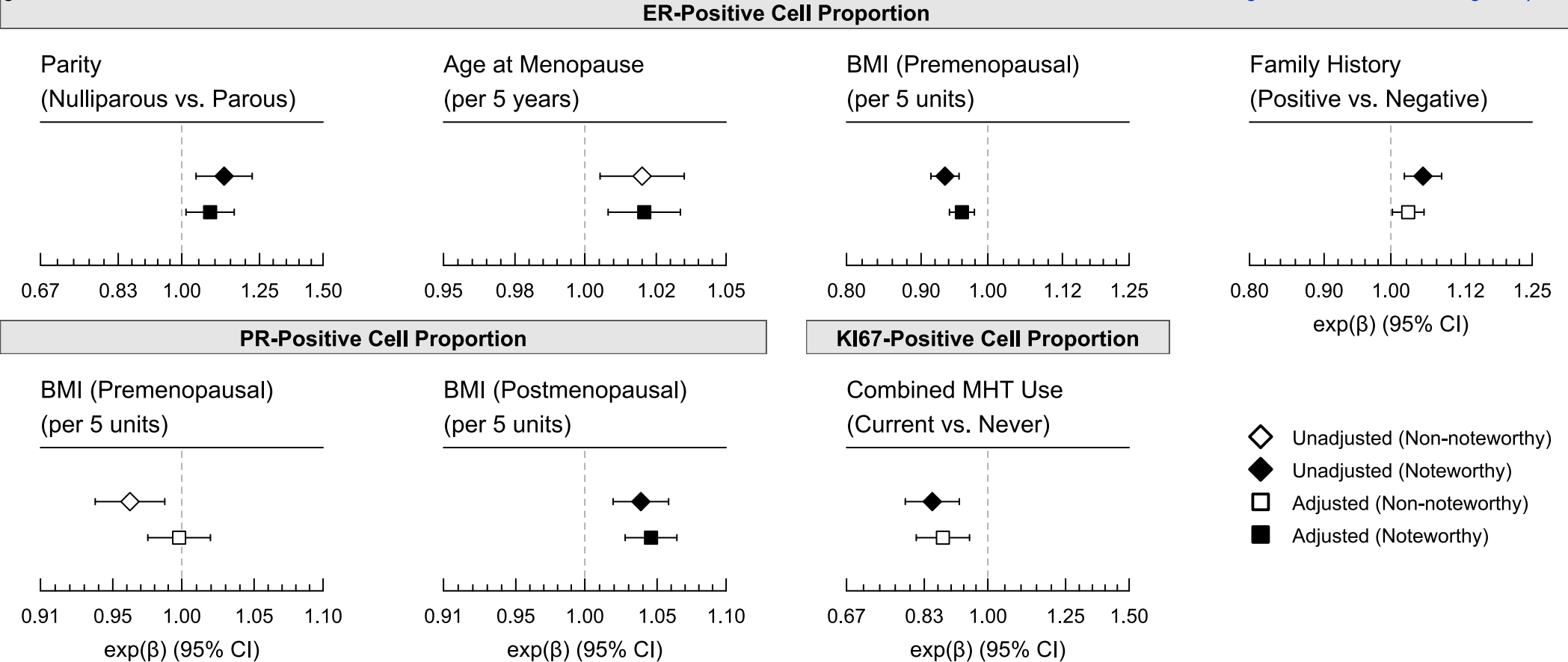
Figure 5. Case-control odds ratios (ORs) with 95% confidence intervals for associations between risk factors and tumor features defined by marker status and tumor grade. Only risk factors with at least one noteworthy (BFDP < 0.2) association in case-case analyses (shown in **Figures 1, 3**) are presented here. Results were estimated from multinomial logistic regression models. See **Table S5** for detailed parameter estimates.

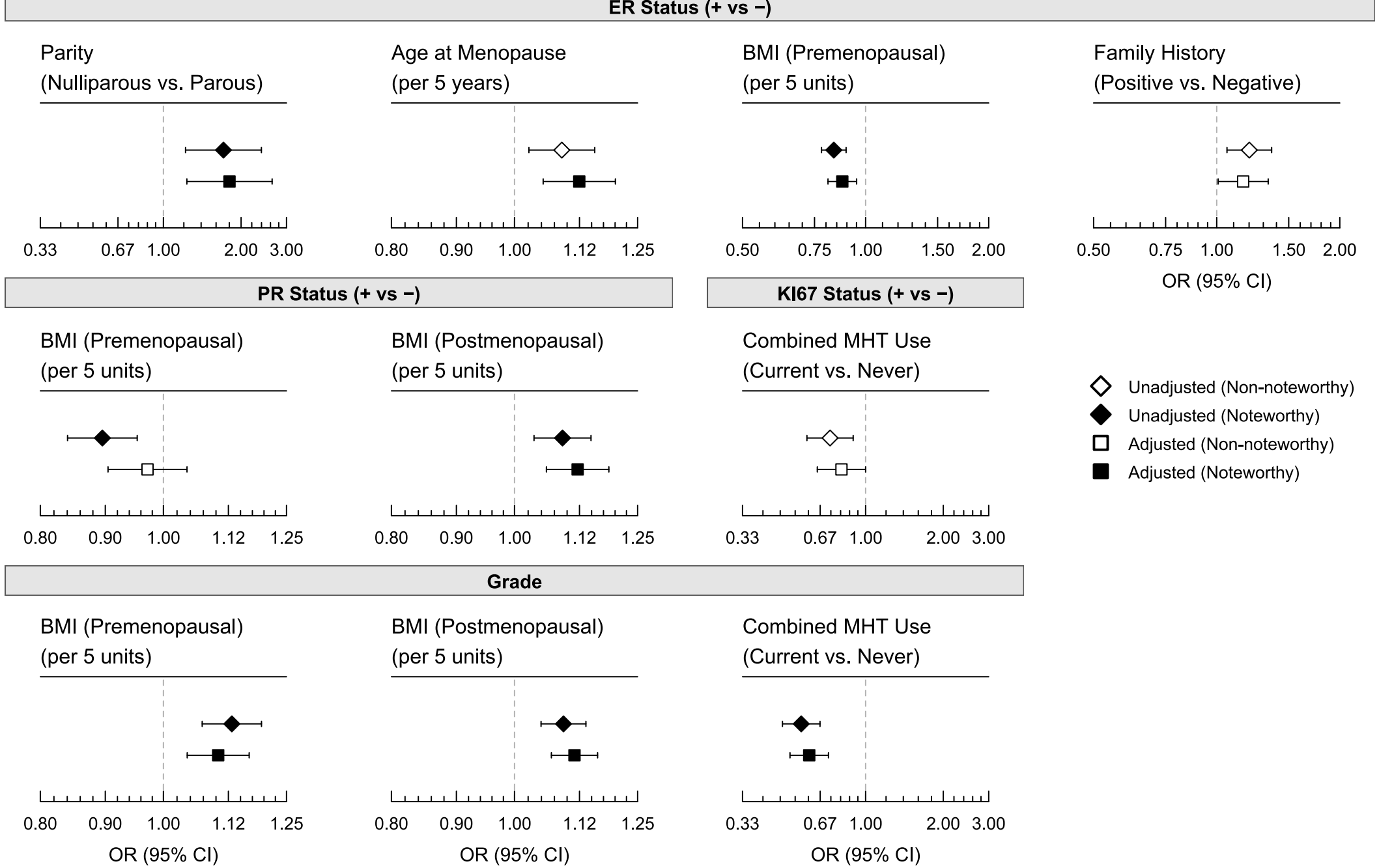
<Figure5.pdf>

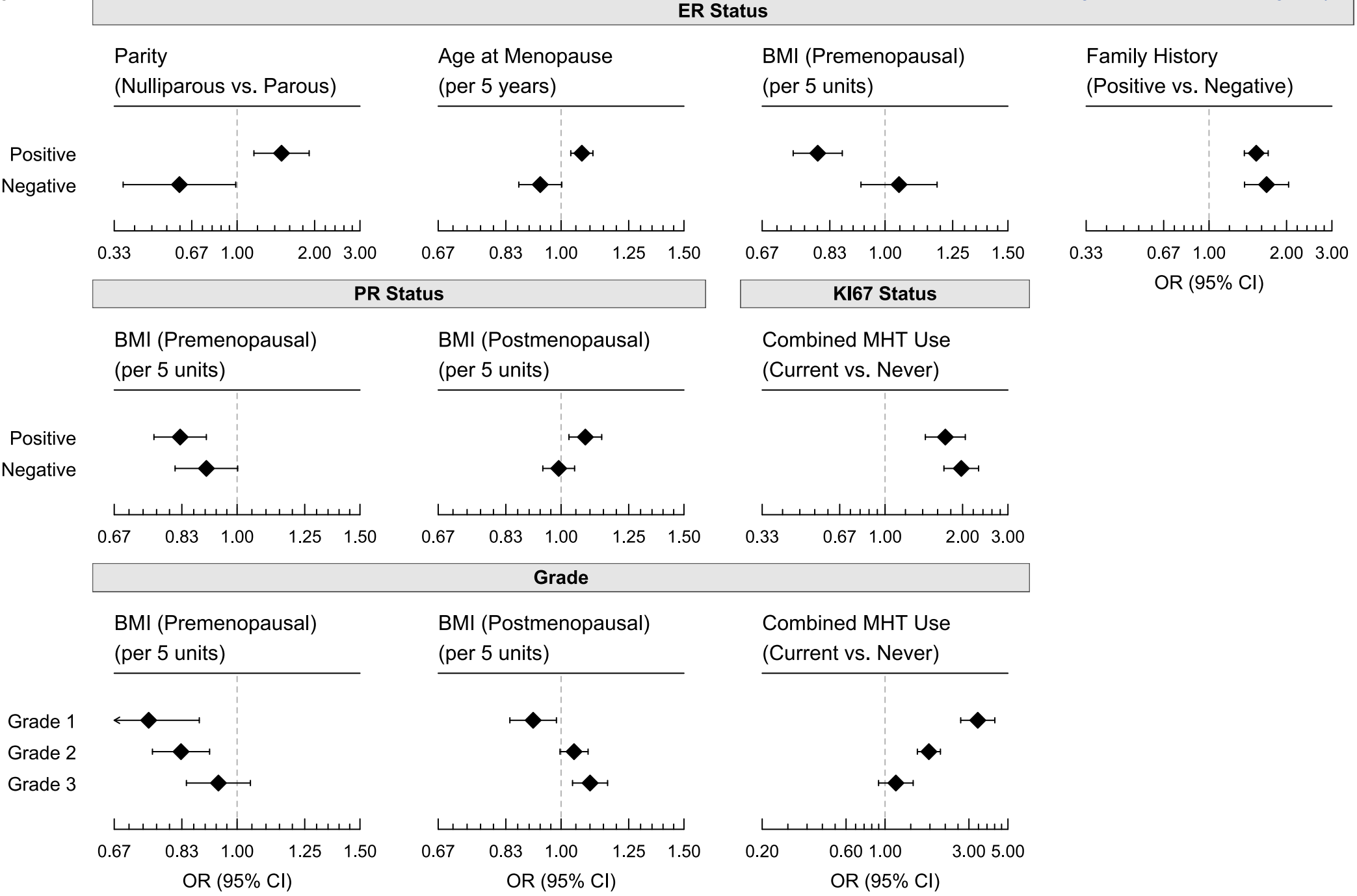
Footnote: BMI = Body Mass Index; OC = Oral Contraceptive; MHT = Menopausal Hormone Therapy

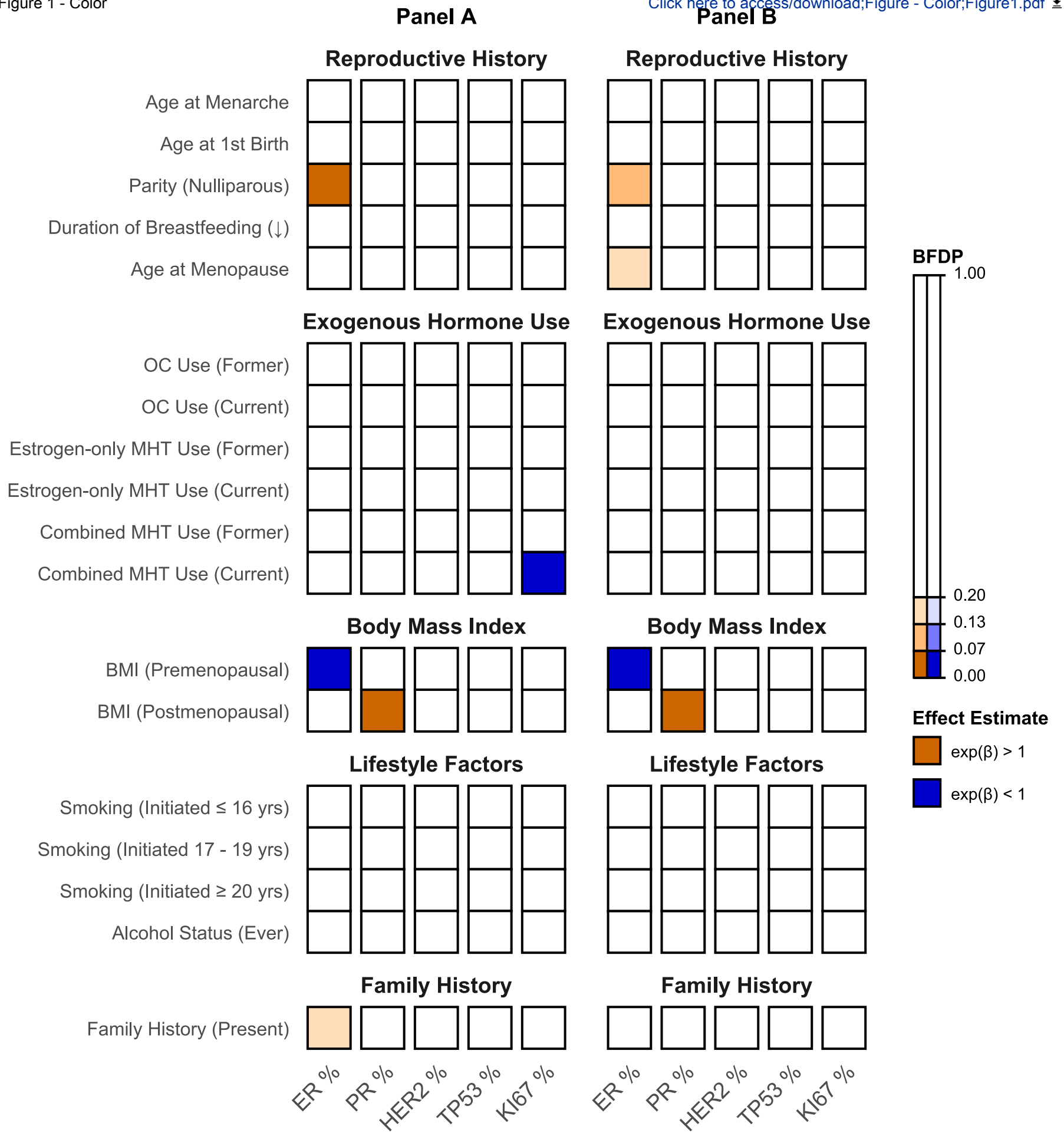
Figure 2 - Black and White

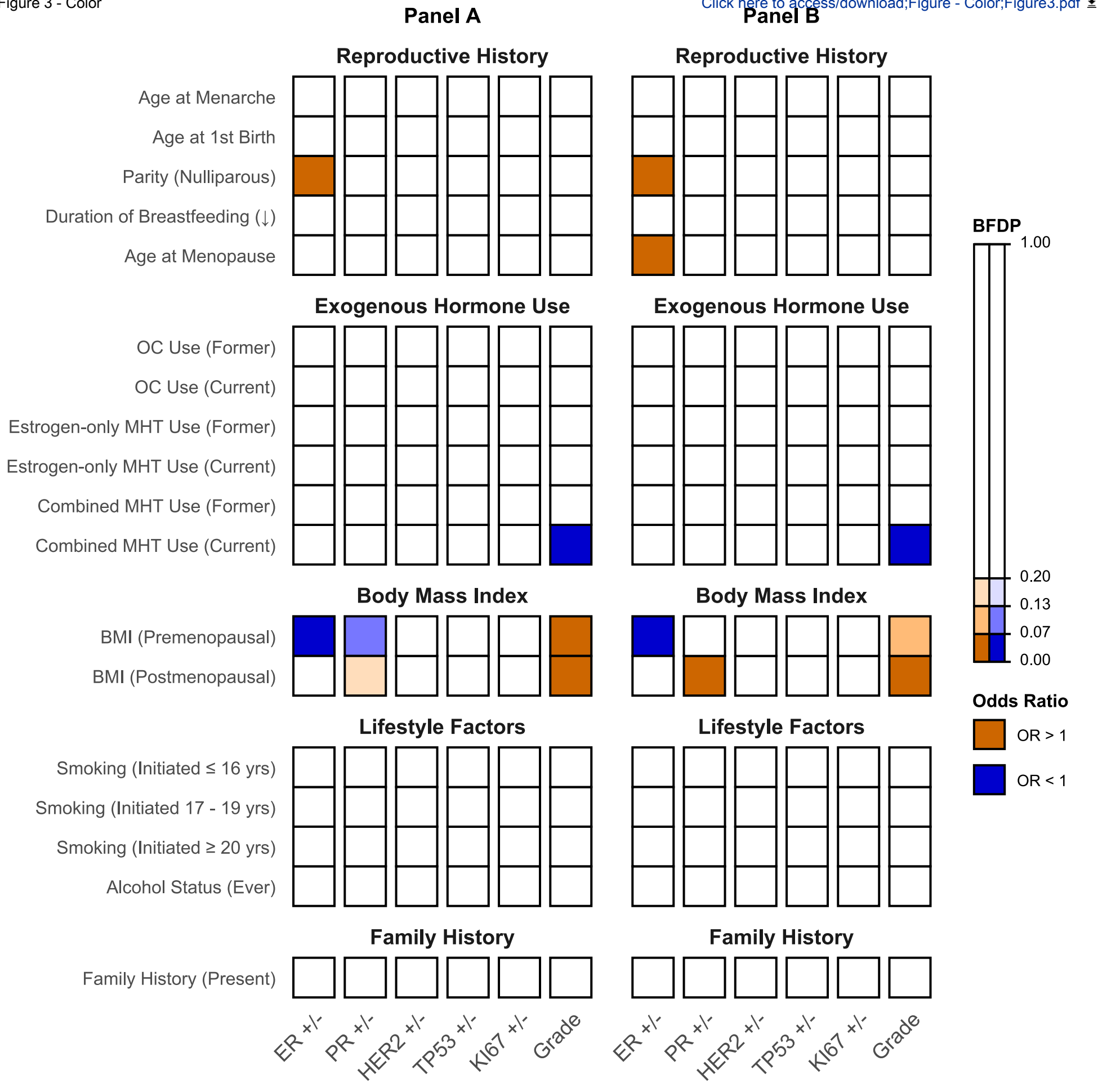
[Click here to access/download;Figure - Black and White;Figure2.pdf](#)













Click here to access/download
Supplemental Materials
Supplement.docx

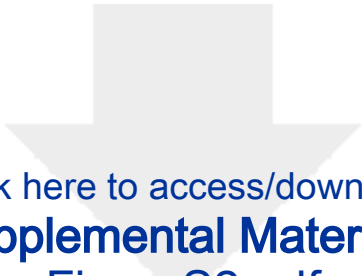




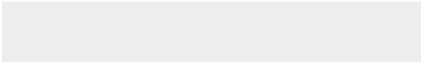


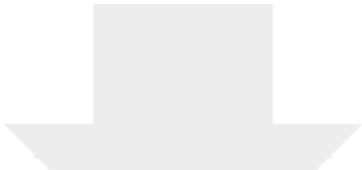


Click here to access/download
Supplemental Materials
FigureS1.pdf



Click here to access/download
Supplemental Materials
FigureS2.pdf





Click here to access/download
Supplemental Materials
FigureS3.pdf

