

Research Article

A Likelihood Ratio Approach for Utilizing Case-Control Data in the Clinical Classification of Rare Sequence Variants: Application to *BRCA1* and *BRCA2*

Maria Zanti ¹, Denise G. O'Mahony,¹ Michael T. Parsons,² Hongyan Li,³ Joe Dennis ⁴, Kristiina Aittomäkkiki,⁵ Irene L. Andrulis,^{6,7} Hoda Anton-Culver,⁸ Kristan J. Aronson ⁹, Annelie Augustinsson,¹⁰ Heiko Becher,¹¹ Stig E. Bojesen,^{12,13,14} Manjeet K. Bolla,⁴ Hermann Brenner,^{15,16,17} Melissa A. Brown,¹⁸ Saundra S. Buys,¹⁹ Federico Canzian,²⁰ Sandrine M. Caputo ^{21,22}, Jose E. Castela,²³ Jenny Chang-Claude,^{24,25} GC-HBOC study Collaborators,²⁶ Kamila Czene,²⁷ Mary B. Daly,²⁸ Arcangela De Nicolo,²⁹ Peter Devilee,^{30,31} Thilo Dörk ³², Alison M. Dunning,³³ Miriam Dwek,³⁴ Diana M. Eccles,³⁵ Christoph Engel ^{36,37}, D. Gareth Evans,^{38,39} Peter A. Fasching,⁴⁰ Manuela Gago-Dominguez,⁴¹ Montserrat García-Closas,⁴² José A. García-Sáenz,⁴³ Aleksandra Gentry-Maharaj,⁴⁴ Willemina R. R. Geurts - Giele,⁴⁵ Graham G. Giles,^{46,47,48} Gord Glendon,⁶ Mark S. Goldberg,^{49,50} Encarna B. Gómez Garcia,⁵¹ Melanie Güendert,^{52,53,54} Pascal Guénel,⁵⁵ Eric Hahnen,^{26,56} Christopher A. Haiman,⁵⁷ Per Hall,^{27,58} Ute Hamann,⁵⁹ Elaine F. Harkness,^{60,61,62} Frans B. L. Hogervorst,⁶³ Antoinette Hollestelle,⁶⁴ Reiner Hoppe,^{65,66} John L. Hopper,⁴⁷ Claude Houdayer,⁶⁷ Richard S. Houlston,⁶⁸ Anthony Howell,⁶⁹ ABCTB Investigators,⁷⁰ Milena Jakimovska ⁷¹, Anna Jakubowska,^{72,73} Helena Jernström,¹⁰ Esther M. John,^{74,75} Rudolf Kaaks,²⁴ Cari M. Kitahara,⁷⁶ Stella Koutros,⁴² Peter Kraft,^{77,78} Vessela N. Kristensen,^{79,80} James V. Lacey,^{81,82} Diether Lambrechts,^{83,84} Melanie Léoné,⁸⁵ Annika Lindblom,^{86,87} Jan Lubiński,⁷² Michael Lush,⁴ Arto Mannermaa,^{88,89,90} Mehdi Manoochehri,⁵⁹ Siranoush Manoukian,⁹¹ Sara Margolin,^{58,92} Maria Elena Martinez,^{93,94} Usha Menon ⁴⁴, Roger L. Milne,^{46,47,48} Alvaro N. Monteiro,⁹⁵ Rachel A. Murphy,^{96,97} Susan L. Neuhausen,⁹⁸ Heli Nevanlinna,⁹⁹ William G. Newman,^{38,39} Kenneth Offit,^{100,101} Sue K. Park,^{102,103,104} Paul James,^{105,106} Paolo Peterlongo,¹⁰⁷ Julian Peto,¹⁰⁸ Dijana Plaseska-Karanfilska,⁷¹ Kevin Punie,¹⁰⁹ Paolo Radice,¹¹⁰ Muhammad U. Rashid,^{59,111} Gad Rennert,¹¹² Atocha Romero ¹¹³, Efraim H. Rosenberg,¹¹⁴ Emmanouil Saloustros ¹¹⁵, Dale P. Sandler,¹¹⁶ Marjanka K. Schmidt,^{117,118,119} Rita K. Schmutzler,^{26,56,120} Xiao-Ou Shu,¹²¹ Jacques Simard ¹²², Melissa C. Southey,^{46,48,123} Jennifer Stone,^{47,124} Dominique Stoppa-Lyonnet,^{21,125,126} Rulla M. Tamimi,^{78,127} William J. Tapper,³⁵ Jack A. Taylor,^{116,128} Soo Hwang Teo,^{129,130} Lauren R. Teras,¹³¹ Mary Beth Terry,¹³² Mads Thomassen,¹³³ Melissa A. Troester,¹³⁴ Celine M. Vachon,¹³⁵ Ana Vega,^{136,137,138} Maaïke P. G. Vreeswijk,³¹ Qin Wang,⁴ Barbara Wappenschmidt,^{26,56} Clarice R. Weinberg,¹³⁹ Alicja Wolk,^{140,141} Wei Zheng,¹²¹ Bingjian Feng,¹⁴² Fergus J. Couch,¹⁴³ Amanda B. Spurdle,² Douglas F. Easton,^{4,33} David E. Goldgar,¹⁴² and Kyriaki Michailidou ^{1,4}

- ¹Biostatistics Unit, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus
- ²Population Health Program, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia
- ³Cancer Control and Population Science, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA
- ⁴Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
- ⁵Department of Clinical Genetics, Helsinki University Hospital, University of Helsinki, Helsinki, Finland
- ⁶Fred A. Litwin Center for Cancer Genetics, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada
- ⁷Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada
- ⁸Department of Medicine, Genetic Epidemiology Research Institute, University of California Irvine, Irvine, CA, USA
- ⁹Department of Public Health Sciences, and Cancer Research Institute, Queen's University, Kingston, ON, Canada
- ¹⁰Oncology, Department of Clinical Sciences in Lund, Lund University, Lund, Sweden
- ¹¹Institute for Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
- ¹²Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark
- ¹³Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark
- ¹⁴Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
- ¹⁵Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany
- ¹⁶Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany
- ¹⁷German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany
- ¹⁸School of Chemistry and Molecular Biosciences, University of Queensland, Brisbane, Queensland, Australia
- ¹⁹Department of Internal Medicine and Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA
- ²⁰Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany
- ²¹Service de Génétique, Institut Curie, Paris, France
- ²²Paris Sciences Lettres Research University, Paris, France
- ²³Oncology and Genetics Unit, Instituto de Investigación Sanitaria Galicia Sur (IISGS), Xerencia de Xestión Integrada de Vigo-SERGAS, Vigo, Spain
- ²⁴Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany
- ²⁵Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany
- ²⁶Center for Familial Breast and Ovarian Cancer, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany
- ²⁷Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
- ²⁸Department of Clinical Genetics, Fox Chase Cancer Center, Philadelphia, PA, USA
- ²⁹Center for Omics Sciences, IRCCS San Raffaele Scientific Institute, Milan, Italy
- ³⁰Department of Pathology, Leiden University Medical Center, Leiden, Netherlands
- ³¹Department of Human Genetics, Leiden University Medical Center, Leiden, Netherlands
- ³²Gynaecology Research Unit, Hannover Medical School, Hannover, Germany
- ³³Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK
- ³⁴School of Life Sciences, University of Westminster, London, UK
- ³⁵Faculty of Medicine, University of Southampton, Southampton, UK
- ³⁶Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany
- ³⁷LIFE-Leipzig Research Centre for Civilization Diseases, University of Leipzig, Leipzig, Germany
- ³⁸Division of Evolution and Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK
- ³⁹North West Genomics Laboratory Hub, Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK
- ⁴⁰Department of Gynecology and Obstetrics, Comprehensive Cancer Center Erlangen-EMN, Friedrich-Alexander University Erlangen-Nuremberg, University Hospital Erlangen, Erlangen, Germany
- ⁴¹Genomic Medicine Group, International Cancer Genetics and Epidemiology Group, Fundación Pública Galega de Medicina Xenómica, Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Spain
- ⁴²Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA
- ⁴³Medical Oncology Department, Hospital Clínico San Carlos, Instituto de Investigación Sanitaria San Carlos (IdISSC), Centro Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain
- ⁴⁴MRC Clinical Trials Unit, Institute of Clinical Trials & Methodology, University College London, London, UK
- ⁴⁵Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, Netherlands
- ⁴⁶Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, Australia

- ⁴⁷Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia
- ⁴⁸Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia
- ⁴⁹Department of Medicine, McGill University, Montréal, QC, Canada
- ⁵⁰Division of Clinical Epidemiology, Royal Victoria Hospital, McGill University, Montréal, QC, Canada
- ⁵¹Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, Netherlands
- ⁵²Molecular Epidemiology Group, C080, German Cancer Research Center (DKFZ), Heidelberg, Germany
- ⁵³Molecular Biology of Breast Cancer, University Womens Clinic Heidelberg, University of Heidelberg, Heidelberg, Germany
- ⁵⁴Institute of Diabetes Research, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
- ⁵⁵Team 'Exposome and Heredity', CESP, Gustave Roussy, INSERM, University Paris-Saclay, UVSQ, Villejuif, France
- ⁵⁶Center for Integrated Oncology (CIO), Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany
- ⁵⁷Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
- ⁵⁸Department of Oncology, Södersjukhuset, Stockholm, Sweden
- ⁵⁹Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany
- ⁶⁰Division of Informatics, Imaging and Data Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK
- ⁶¹Nightingale & Genesis Prevention Centre, Wythenshawe Hospital, Manchester University NHS Foundation Trust, Manchester, UK
- ⁶²NIHR Manchester Biomedical Research Unit, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK
- ⁶³Family Cancer Clinic, The Netherlands Cancer Institute-Antoni van Leeuwenhoek hospital, Amsterdam, Netherlands
- ⁶⁴Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, Netherlands
- ⁶⁵Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany
- ⁶⁶University of Tübingen, Tübingen, Germany
- ⁶⁷Department of Genetics, F76000 and Normandy University, UNIROUEN, Inserm U1245, Normandy Centre for Genomic and Personalized Medicine, Rouen University Hospital, Rouen, France
- ⁶⁸Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK
- ⁶⁹Division of Cancer Sciences, University of Manchester, Manchester, UK
- ⁷⁰Australian Breast Cancer Tissue Bank, Westmead Institute for Medical Research, University of Sydney, Sydney, New South Wales, Australia
- ⁷¹Research Centre for Genetic Engineering and Biotechnology 'Georgi D. Efremov', MASA, Skopje, North Macedonia
- ⁷²Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland
- ⁷³Independent Laboratory of Molecular Biology and Genetic Diagnostics, Pomeranian Medical University, Szczecin, Poland
- ⁷⁴Department of Epidemiology and Population Health, Stanford University School of Medicine, Stanford, CA, USA
- ⁷⁵Department of Medicine, Division of Oncology, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA
- ⁷⁶Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA
- ⁷⁷Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA, USA
- ⁷⁸Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
- ⁷⁹Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway
- ⁸⁰Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway
- ⁸¹Department of Computational and Quantitative Medicine, City of Hope, Duarte, CA, USA
- ⁸²City of Hope Comprehensive Cancer Center, City of Hope, Duarte, CA, USA
- ⁸³Laboratory for Translational Genetics, Department of Human Genetics, KU Leuven, Leuven, Belgium
- ⁸⁴VIB Center for Cancer Biology, VIB, Leuven, Belgium
- ⁸⁵Genetic and Cancer Medical Laboratory HCL-CLB, Hospices Civils de Lyon, Bron, France
- ⁸⁶Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden
- ⁸⁷Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden
- ⁸⁸Translational Cancer Research Area, University of Eastern Finland, Kuopio, Finland
- ⁸⁹Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Finland
- ⁹⁰Biobank of Eastern Finland, Kuopio University Hospital, Kuopio, Finland
- ⁹¹Unit of Medical Genetics, Department of Medical Oncology and Hematology, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan, Italy
- ⁹²Department of Clinical Science and Education, Södersjukhuset, Karolinska Institutet, Stockholm, Sweden
- ⁹³Moore's Cancer Center, University of California San Diego, La Jolla, CA, USA
- ⁹⁴Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, La Jolla, CA, USA

- ⁹⁵Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, FL, USA
- ⁹⁶School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada
- ⁹⁷Cancer Control Research, BC Cancer, Vancouver, BC, Canada
- ⁹⁸Department of Population Sciences, Beckman Research Institute of City of Hope, Duarte, CA, USA
- ⁹⁹Department of Obstetrics and Gynecology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland
- ¹⁰⁰Clinical Genetics Research Lab, Department of Cancer Biology and Genetics, Memorial Sloan Kettering Cancer Center, New York, NY, USA
- ¹⁰¹Clinical Genetics Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA
- ¹⁰²Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea
- ¹⁰³Integrated Major in Innovative Medical Science, Seoul National University College of Medicine, Seoul, Republic of Korea
- ¹⁰⁴Cancer Research Institute, Seoul National University, Seoul, Republic of Korea
- ¹⁰⁵Sir Peter MacCallum Department of Oncology, The University of Melbourne, Melbourne, Victoria, Australia
- ¹⁰⁶Parkville Familial Cancer Centre, Peter MacCallum Cancer Centre and Royal Melbourne Hospital, Melbourne, Victoria, Australia
- ¹⁰⁷Genome Diagnostics Program, IFOM-ETS the AIRC Institute of Molecular Oncology, Milan, Italy
- ¹⁰⁸Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK
- ¹⁰⁹Department of General Medical Oncology and Multidisciplinary Breast Centre, Leuven Cancer Institute and University Hospitals Leuven, Leuven, Belgium
- ¹¹⁰Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Research, Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Milan, Italy
- ¹¹¹Department of Basic Sciences, Shaukat Khanum Memorial Cancer Hospital and Research Centre (SKMCH & RC), Lahore, Pakistan
- ¹¹²Clalit National Cancer Control Center, Carmel Medical Center and Technion Faculty of Medicine, Haifa, Israel
- ¹¹³Medical Oncology Department, Hospital Universitario Puerta de Hierro, Madrid, Spain
- ¹¹⁴Department of Pathology, The Netherlands Cancer Institute-Antoni van Leeuwenhoek hospital, Amsterdam, Netherlands
- ¹¹⁵Department of Oncology, University Hospital of Larissa, Larissa, Greece
- ¹¹⁶Epidemiology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA
- ¹¹⁷Division of Molecular Pathology, The Netherlands Cancer Institute, Amsterdam, Netherlands
- ¹¹⁸Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute-Antoni van Leeuwenhoek hospital, Amsterdam, Netherlands
- ¹¹⁹Department of Clinical Genetics, Leiden University Medical Center, Leiden, Netherlands
- ¹²⁰Center for Molecular Medicine Cologne (CMMC), Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany
- ¹²¹Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA
- ¹²²Genomics Center, Centre Hospitalier Universitaire de Québec-Université Laval Research Center, Québec City, QC, Canada
- ¹²³Department of Clinical Pathology, The University of Melbourne, Melbourne, Victoria, Australia
- ¹²⁴Genetic Epidemiology Group, School of Population and Global Health, University of Western Australia, Perth, Western Australia, Australia
- ¹²⁵Department of Tumour Biology, INSERM U830, Paris, France
- ¹²⁶Université Paris Descartes, Paris, France
- ¹²⁷Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA
- ¹²⁸Epigenetic and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA
- ¹²⁹Breast Cancer Research Programme, Cancer Research Malaysia, Subang Jaya, Selangor, Malaysia
- ¹³⁰Department of Surgery, Faculty of Medicine, University of Malaya, UM Cancer Research Institute, Kuala Lumpur, Malaysia
- ¹³¹Department of Population Science, American Cancer Society, Atlanta, GA, USA
- ¹³²Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA
- ¹³³Department of Clinical Genetics, Odense University Hospital, Odense C, Denmark
- ¹³⁴Department of Epidemiology, Gillings School of Global Public Health and UNC Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
- ¹³⁵Department of Quantitative Health Sciences, Division of Epidemiology, Mayo Clinic, Rochester, MN, USA
- ¹³⁶Centro de Investigación en Red de Enfermedades Raras (CIBERER), Madrid, Spain
- ¹³⁷Fundación Pública Galega de Medicina Xenómica, Santiago de Compostela, Spain
- ¹³⁸Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Spain
- ¹³⁹Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA

¹⁴⁰*Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden*

¹⁴¹*Department of Surgical Sciences, Uppsala University, Uppsala, Sweden*

¹⁴²*Department of Dermatology, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT, USA*

¹⁴³*Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA*

Correspondence should be addressed to Kyriaki Michailidou; kyriakimi@cing.ac.cy

Received 20 December 2022; Revised 6 March 2023; Accepted 16 June 2023; Published 14 September 2023

Academic Editor: Garry Cutting

Copyright © 2023 Maria Zanti et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A large number of variants identified through clinical genetic testing in disease susceptibility genes are of uncertain significance (VUS). Following the recommendations of the American College of Medical Genetics and Genomics (ACMG) and Association for Molecular Pathology (AMP), the frequency in case-control datasets (PS4 criterion) can inform their interpretation. We present a novel case-control likelihood ratio-based method that incorporates gene-specific age-related penetrance. We demonstrate the utility of this method in the analysis of simulated and real datasets. In the analysis of simulated data, the likelihood ratio method was more powerful compared to other methods. Likelihood ratios were calculated for a case-control dataset of *BRCA1* and *BRCA2* variants from the Breast Cancer Association Consortium (BCAC) and compared with logistic regression results. A larger number of variants reached evidence in favor of pathogenicity, and a substantial number of variants had evidence against pathogenicity—findings that would not have been reached using other case-control analysis methods. Our novel method provides greater power to classify rare variants compared with classical case-control methods. As an initiative from the ENIGMA Analytical Working Group, we provide user-friendly scripts and preformatted Excel calculators for implementation of the method for rare variants in *BRCA1*, *BRCA2*, and other high-risk genes with known penetrance.

1. Introduction

Clinical genetic testing of disease susceptibility genes often identifies variants of uncertain significance (VUS), complicating the clinical management of carriers and their families [1]. The assessment of the clinical significance of these rare sequence variants, including missense substitutions, in-frame deletions and insertions, and intronic variants, is essential to directing the clinical management of carriers and their relatives towards appropriate prevention, early detection, and personalized treatments.

The most widely used method for the interpretation of germline variants is via the application of the standards and guidelines recommended by the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) [2]. Strength levels (very strong, strong, moderate, and supporting) are assigned to independent lines of evidence for or against variant pathogenicity. These strength levels are then combined and used in a scoring system to provide a clinical class, expressed as pathogenic, likely pathogenic, likely benign, benign, or VUS. These guidelines integrate various sources of information including the variant's nature and position (e.g., nonsense, frameshift, and missense) and clinical data (e.g., prevalence in affected individuals and controls), and the combination of this information is interpreted to establish the significance of the variant under investigation with respect to risk. These criteria were recently reinterpreted in a quantitative Bayesian framework, which derived ranges of likelihood ratios (LRs) consistent with each of the evidence strength levels [3]. For case-control data, the specific criterion (PS4) states that a relative risk (RR) or odds ratio (OR) > 5.0 with nominal statistical significance (i.e., the

confidence interval of the RR or OR does not include 1) provides strong evidence in favor of pathogenicity [2].

A significant advance in the classification of variants in cancer and other disease genes was the development of the multifactorial integrated likelihood ratio model [4]; this model combines multiple features under the assumption that each of them is an independent predictor of variant pathogenicity in a Bayesian framework, thus providing a quantitative estimate of the pathogenicity of a variant [5]. The ENIGMA consortium [6] has been applying and extending this multifactorial likelihood model. To date, application of this model has included clinically calibrated prior probabilities of pathogenicity derived from bioinformatic prediction of variant effect and location, along with a combined LR derived from clinical data [5], such as family history of cancer [7], breast cancer tumor pathology [8], variant cosegregation with disease [9, 10], and variant cooccurrence in trans with a pathogenic variant (PV) in the same gene [7]. This model can also incorporate LRs derived from variant frequency in cases and controls. Recently, case-control information derived from genotype data for 20 variants was incorporated into a comprehensive multifactorial likelihood analysis of *BRCA1* and *BRCA2* variants by ENIGMA [11], using a method incorporating gene- and age-specific penetrance of PV carriers only. Such case-control LR calculations take into consideration gene- and age-specific penetrance values, and hence they might be expected to outperform the statistical measures currently recommended by ACMG/AMP for the analysis of case-control data (i.e., OR or RR estimates).

In this paper, we present a novel case-control LR method, based on the same principle as used in Parsons et al.'s [11], that incorporates age information in both

carriers and noncarriers in the dataset. The method can be used to obtain evidence in favor or against pathogenicity for rare variants in any gene for which there exist known age-specific penetrance estimates based on data obtained from case-control studies. We illustrate the use of this method to calculate LRs for 24 *BRCA1* and 68 *BRCA2* variants from breast cancer case-control genotype data generated by the Breast Cancer Association Consortium (BCAC) as part of the large-scale OncoArray project [12]. We further demonstrate the utility of this case-control LR approach to aid in the interpretation of the clinical significance of variants using evidence aligned to ACMG/AMP code strengths or other classification methods.

2. Methods

2.1. Case-Control Datasets

2.1.1. Simulated Case-Control Dataset. Genotype data simulations were performed using the R (v3.6.1) (<https://www.r-project.org/>) statistical computing language. To create case-control datasets, genotypes for cases and controls were simulated using a Poisson distribution with λ equal to the mean number of events (variant carriers) in the given interval, expressed as

$$\begin{aligned}\lambda_{\text{Cases}} &= N \times \text{RR} \times \text{MAF}, \\ \lambda_{\text{Controls}} &= N \times \text{MAF},\end{aligned}\quad (1)$$

where N denotes the sample size, RR denotes the relative breast cancer risk of the causal variant and MAF denotes the minor allele frequency of the variant in the general population. Ages were simulated using a normal distribution, with the mean and standard deviation following the gene-specific age distribution in the CARRIER population-based study [13].

Genotype data simulations were carried out for variants conferring a RR of 1 (indicating no increased risk), 2, 3, 4, 5, 6, 7, 8, 9, or 10, minor allele frequency in controls of 0.0001, 0.00005, or 0.00003, and sample size of $N = 20,000$ (20,000 breast cancer cases and 20,000 controls), 30,000 (30,000 breast cancer cases and 30,000 controls), or 50,000 (50,000 breast cancer cases and 50,000 controls). For each of these 90 scenarios, we simulated 10,000 replicates.

Additionally, in order to account for the possibility that age information is not available, we repeated the analysis using same age for all individuals.

2.1.2. BCAC OncoArray Dataset. Genotype data were generated as part of the BCAC component of the OncoArray project [12] (studies included in the analysis are listed in Supplementary Table S1) and were available for 75,657 breast cancer cases and 52,987 controls of European ancestry. The majority of studies were population-based case-control studies or case-control studies nested within population-based cohorts. However, a subset of studies oversampled cases with a family history of breast cancer. Of these, 464 breast cancer cases and 1,347 controls had missing information regarding their age at diagnosis or

interview, respectively and were excluded from the analyses. Another 1,445 cases and 858 controls were removed because their ages fell outside the interval of 21–80 years (the age range for which penetrance estimates were available). Cluster plots of 56 *BRCA1* and 127 *BRCA2* variants, nominated by ENIGMA researchers for inclusion in the OncoArray project were manually checked to review the automated calls. This was performed since automated genotype calling for rare variants from GWAS chips has been shown to be suboptimal [14]. Genotypes were adjusted for 41 *BRCA1* and 91 *BRCA2* variants, while 3 *BRCA1* and 2 *BRCA2* variant genotypes were determined to have been called correctly by automated clustering. Genotype recalling was not performed for 12 *BRCA1* and 34 *BRCA2* variants due to the low quality of the genotype data; these variants were not considered further.

After genotype cluster review and recalling, 16 *BRCA1* and 19 *BRCA2* variants were excluded from further analysis due to their high frequency ($>0.1\%$). Additionally, case-control LR calculations were not possible for four *BRCA1* and six *BRCA2* variants due to the absence of variant carriers in the postfiltering dataset. After these exclusions, case-control LR and logistic regression analyses were performed for 24 *BRCA1* and 68 *BRCA2* variants. It should be noted that some of the variants selected for the array have subsequently been classified or were those whose pathogenicity status were known and were included as positive or negative controls.

2.2. Statistical Analyses

2.2.1. Case-Control Likelihood Ratio Method. This method (detailed in Supplementary File 1) compares the likelihood of the distribution of the variant of interest among cases and controls under the hypothesis that the variant is associated with similar risks of the disease in question as the “average” pathogenic variant (H_p), compared to the likelihood under the hypothesis that it is a benign variant not associated with increased risk (H_b). These risks may be age-, sex-, and/or country-specific. Thus

$$\text{LR} = \frac{\Pr(\text{Data}|H_p)}{\Pr(\text{Data}|H_b)},\quad (2)$$

where Data denotes observed data on carrier status of a variant of interest, case-control status, and age at diagnosis or interview, combined over all individuals in the dataset.

In order to calculate the above LR, we follow a survival analysis framework. We first determine the probability that an individual with genotype k remains unaffected at age t , $S_k(t)$, and the corresponding probability that an individual with genotype k is affected at age t , $f_k(t)$ (where $k = 0$ or 1 for non-carriers and carriers, respectively). These probabilities can be computed from the age-specific baseline incidence, $\lambda_0(t)$, and the age-specific log-relative risk of an assumed pathogenic variant in the gene of interest, $\beta(t)$. These probabilities are given by

$$S_k(t) = \exp\left(-\int_0^t \lambda_0(t) e^{\beta(t)k} dt\right), \quad (3)$$

$$f_k(t) = S_k(t) e^{\beta(t)k}.$$

As detailed in Supplementary File 1, the likelihood ratio is to close approximation, given by

$$\text{ccLR} = \frac{\left(\prod_{v_j=1} S_1(t_j) e^{\beta(t_j)d_j/S_0(t_j)}\right) / \left(\sum_j S_1(t_j) e^{\beta(t_j)d_j/S_0(t_j)}\right)^K}{1/N^K}, \quad (4)$$

where N is the total number of individuals, K is the number of variant carriers, v_j is the variant status (0 for noncarriers and 1 for variant carriers), and d_j is the disease status (0 for controls and 1 for cases) for individual j .

The baseline incidence rates $\lambda_0(t)$ were taken from the age-specific background rates for England and Wales (1998-2002) (<https://ci5.iarc.fr/CI5I-X/Default.aspx>), and the age-specific breast cancer relative risks for pathogenic variant carriers $\beta(t)$ were taken from the recent large-scale BRIDGES (Breast Cancer Risk after Diagnostic Gene Sequencing) project [15]. To allow for possible carrier frequency differences by country, stratified LR calculations were performed within each country and then multiplied to provide a final LR.

Likelihood ratios are further translated into ACMG/AMP code strength categories according to published recommendations [3]. Likelihood ratio estimates in favor of variant pathogenicity are scored as very strong, $\text{LR} \geq 350$; strong, $350 > \text{LR} \geq 18.7$; moderate, $18.70 > \text{LR} \geq 4.33$; and supporting, $4.33 > \text{LR} \geq 2.08$. Likelihood ratio evidence for benign variant status is scored as very strong, $\text{LR} \leq 0.0029$; strong, $0.0029 < \text{LR} \leq 0.053$; moderate, $0.053 < \text{LR} \leq 0.231$; and supporting, $0.231 < \text{LR} \leq 0.48$. No evidence strength corresponded to estimates of $0.48 \leq \text{LR} < 2.08$.

In a series of sensitivity analyses, the method was applied using three other published RR estimates: from case series unselected for family history of breast cancer [16], cohort series of *BRCA1* and *BRCA2* carriers [17], and breast cancer hazard ratio estimates for missense *BRCA1* and *BRCA2* variants [18]. In order to account for country-specific effects, the stratified analysis was also performed using age- and country-specific incidence rates derived from the Cancer Incidence in Five Continents, volume 9, 1998-2002, (<https://ci5.iarc.fr/CI5I-X/Default.aspx>). Age-specific breast cancer incidences for Greece and North Macedonia were retrieved from the 2020 cancer registry (European Cancer Information System (ECIS), <https://ecis.jrc.ec.europa.eu/>) since cancer incidence data were not available for the years 1998-2019. Unstratified analyses were also performed for comparison.

Detailed R scripts and preformatted Excel calculators (user can either input individual-level data or tabulated by age groups) for the calculation of case-control LRs can be found using the following GitHub link (<https://github.com/BiostatUnitCING/ccLR>). The files provided can be used to

derive estimates based on the RR from Dorling et al. [15], Kuchenbaecker et al. [17], or Antoniou et al. [16]. In addition, this method can also be used to compute case-control LRs for variants in other disease susceptibility genes by using age-specific penetrance estimates for the gene of interest (indicated by “custom” gene in the preformatted Excel calculators and R script). Furthermore, to allow for the possibility that age information is not available (or is only available for a subset of the dataset), the user can incorporate individuals with unknown age at diagnosis or interview into any of the age groups specified in the tabulated calculator.

2.2.2. Odds Ratio Analysis. Odds ratio analysis was performed using logistic regression adjusted by age and country (if applicable) and Fisher’s exact test (corrected using Haldane’s method when simulations resulted in zero variant carriers in cases or controls [19]). Logistic regression p values were estimated using the likelihood ratio test. Based on the original ACMG/AMP recommendations [2], an OR estimate greater than 5.0, with the confidence interval not including 1.0, was used to define strong evidence of pathogenicity (PS4).

2.2.3. Evaluation and Application of the Case-Control Analyses Methods. The simulated datasets were analyzed using the novel case-control LR method, logistic regression (adjusted by age), and Fisher’s exact test. The case-control LR method was applied using age-specific breast cancer ORs for *BRCA1* and *BRCA2* PVs [15]. For causal variants with a relative risk of 2 to 10, the power of the case-control LR method was estimated either as the probability of reaching at least supporting ($\text{LR} \geq 2.08$) or at least strong pathogenic ($\text{LR} \geq 18.7$) evidence. For benign variants with a relative risk of 1, the power of the case-control LR method was estimated either as the probability of reaching at least supporting ($\text{LR} \leq 0.48$) or at least strong ($\text{LR} \leq 0.053$) benign ACMG/AMP evidence. Correspondingly, type I error for pathogenicity was calculated as the probability of obtaining at least supporting or at least strong pathogenic ACMG/AMP evidence when the relative risk was set to 1. Equivalently, type I error for evidence against pathogenicity was calculated as the probability of obtaining at least supporting or at least strong benign ACMG/AMP evidence when the relative risk was greater than one. The power of the OR methods was estimated as the probability of reaching the ACMG/AMP PS4 criterion ($\text{OR} > 5.0$, CI not including 1.0, p value < 0.05). Following the analyses results of the simulated datasets, optimal LR cut-offs (to maximize power and minimize type I error) are used to define ACMG/AMP evidence strengths for the 92 variants included in the BCAC OncoArray dataset.

3. Results

3.1. Simulated Datasets. Based on the simulation results for high-risk *BRCA1* ($\text{RR} > 9$) and *BRCA2* ($\text{RR} > 5$) variants, LR of strong and very strong evidence in favor of pathogenicity ($\text{LR} \geq 18.7$) and of at least supporting evidence against

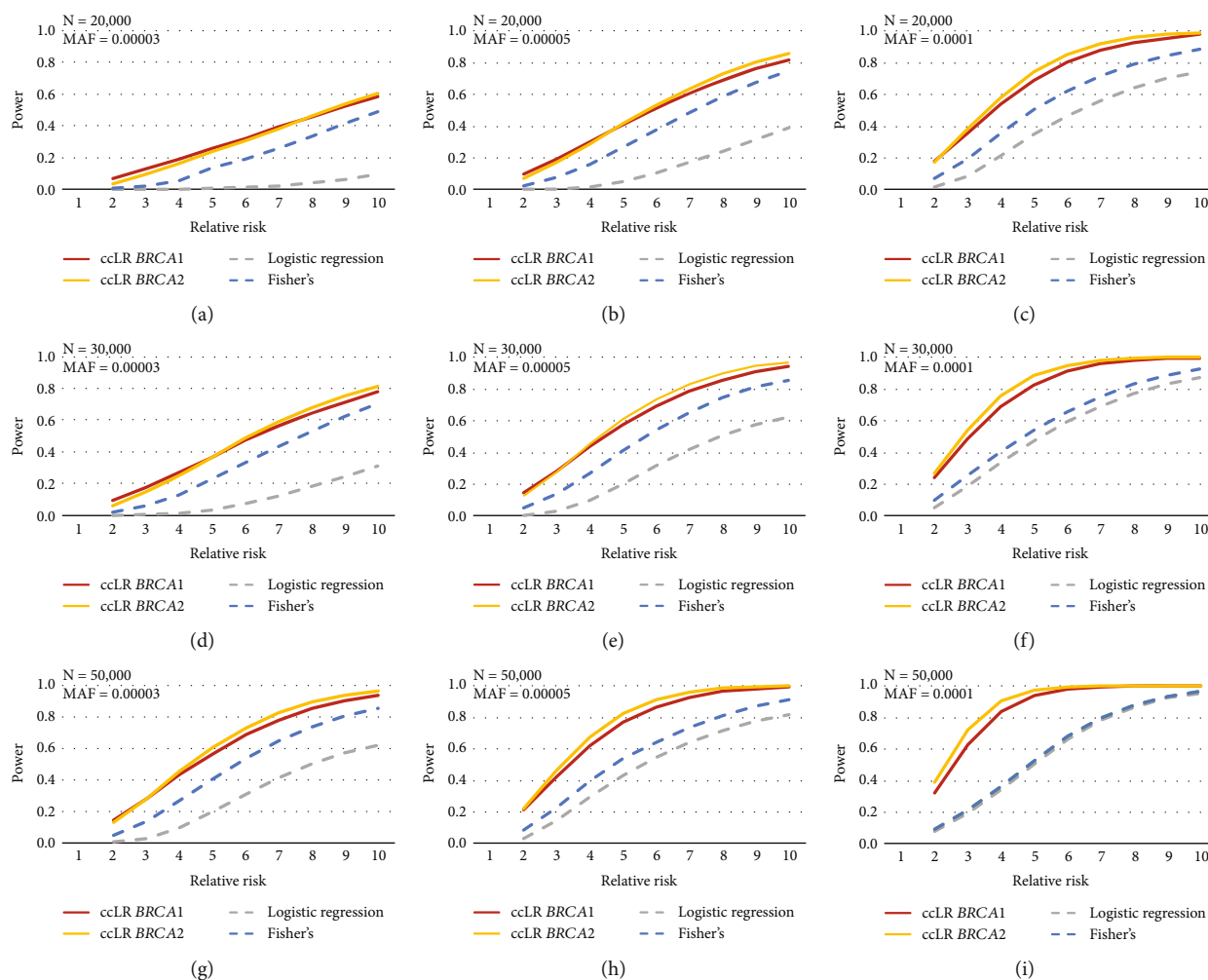


FIGURE 1: Performance of the case-control likelihood ratio method and odds ratio analysis in providing at least strong ACMG/AMP evidence in favor of pathogenicity ($LR \geq 18.7$) using simulated datasets. Power equals the probability of reaching at least strong pathogenic ACMG/AMP evidence. Genotype data simulations were carried out for causal variants conferring disease relative risk between 2 and 10. We performed 10,000 simulations for each case scenario. Results represent simulated case-control data for 20,000 (a–c), 30,000 (d–f), or 50,000 (g–i) breast cancer cases and controls and minor allele frequency of 0.00003 (a, d, g), 0.00005 (b, e, h), or 0.0001 (c, f, i). ccLR: case-control likelihood ratio; MAF: minor allele frequency; N : sample size.

pathogenicity ($LR \leq 0.48$) should be used in order to maintain a high power ($>80\%$) and low type I error (<0.05) (Supplementary Table S2).

Results for all measures in all simulated datasets show that the power to achieve strong evidence in favour of pathogenicity is consistently greater for the case-control LR method using age-specific breast cancer risks compared to standard OR analysis methods (Figure 1, Supplementary Table S2). The power to correctly categorize variants with a RR comparable to a typical *BRCA1* PV was $>80\%$ in all scenarios except for small datasets ($N \leq 30,000$) with causal variants present at a lower frequency ($MAF = 0.00003$) (Figure 1(a)).

In addition, the case-control LR method can also be used to obtain evidence against pathogenicity, something that cannot be achieved using standard OR analysis methods. Results from simulated case-control datasets of benign variants (RR of 1, Figure 2) show that the case-control LR method using the age-specific RRs of the

“average” *BRCA1* PV exhibits adequate power ($>80\%$) to identify variants with evidence against pathogenicity ($LR \leq 0.48$) for larger datasets ($N \geq 30,000$) and a MAF of 0.0001.

The implementation of the method to account for datasets with missing information, assuming the same age for all individuals, demonstrated reduced power and increased type I error in all simulations. However, the type I error was still less than 0.05 in all cases (Supplementary Figures S1 and S2, Supplementary Table S3).

3.2. BCAC OncoArray Dataset

3.2.1. Logistic Regression Results. Using logistic regression, two *BRCA2* variants (2%) (Table 1) reached strong pathogenic evidence following the ACMG/AMP classification criterion (PS4 criterion, $OR > 5$, p value <0.05 , and CI not including 1.0) [2]. Detailed logistic regression results for all variants are shown in Supplementary Table S4.

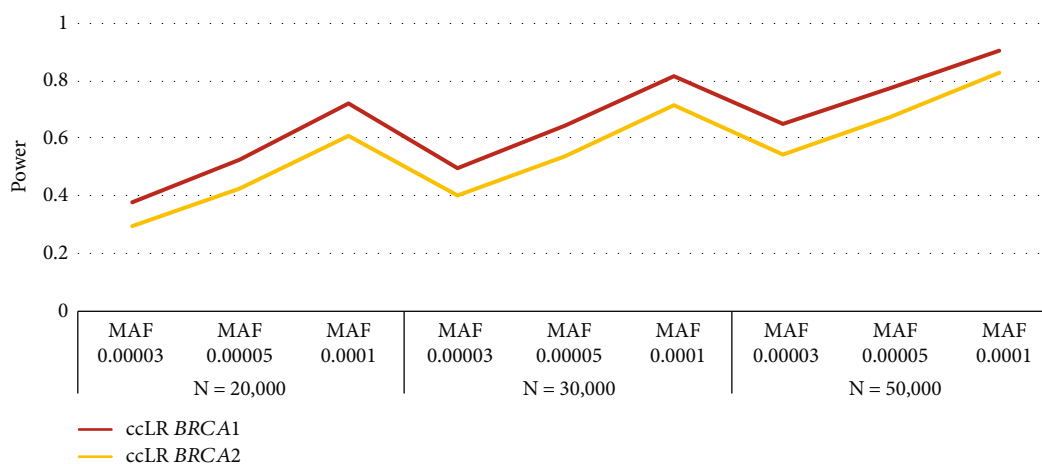


FIGURE 2: Performance of the case-control likelihood ratio method in providing ACMG/AMP evidence against pathogenicity, using simulated datasets. Power equals to the probability of reaching at least supporting benign ACMG/AMP evidence ($LR \leq 0.48$) when the relative risk was set to 1. We performed 10,000 simulations for each case scenario. Results represent simulated case-control data for 20,000, 30,000, or 50,000 breast cancer cases and controls and minor allele frequency of 0.00003, 0.00005, or 0.0001. ccLR: case-control likelihood ratio; MAF: minor allele frequency; N: sample size.

3.2.2. Case-Control LRs and ACMG/AMP Code Strengths. In the country-stratified baseline analysis (using the breast cancer ORs estimated from BRIDGES [15]), evidence in favor of pathogenicity (defined as $LR \geq 18.70$ following the simulation cut-offs) was achieved for 6 variants (6.5%) (Table 2), of which 3 variants were assigned very strong and another 3 strong strengths. Evidence against pathogenicity (defined as $LR \leq 0.48$) was observed for 59 variants (64.1%), of which 26 were assigned very strong, 14 strong, 7 moderate, and 12 supporting strengths. The results for the remaining 27 variants (29.3%) were uninformative. Case-control LRs and corresponding ACMG/AMP code strengths for all 92 *BRCA1* and *BRCA2* variants are shown in Supplementary Table S4. The different sensitivity analyses did not show any major discrepancies in the estimated LRs (Supplementary Table S5).

4. Discussion

This study provides a detailed description of the methodology to calculate case-control LRs for rare variants using case-control data based on age- and gene-specific relative risks and age information for noncarriers. The LRs are calculated by comparing the likelihood of the distribution of the variant of interest in cases and controls under the hypothesis that the variant has similar age-specific relative risks as the “average” pathogenic variant, compared to the hypothesis that it is not associated with increased (or decreased) disease risk. We evaluated the method using simulated datasets and further applied it to derive LRs for pathogenicity for individual variants from the analysis of genotype data from a large case-control study. These can now be used in combination with other evidence to inform variant classification—either according to ACMG/AMP classification standards and guidelines [2, 3] or using multifactorial likelihood modelling approaches [4, 11]. Further, we provide user-friendly scripts and preformatted Excel cal-

culators to facilitate the future implementation of this method for the calculation of case-control LRs. These resources may be readily applied for the calculation of LRs to be used in the classification of VUS in the *BRCA1* and *BRCA2* and other disease susceptibility genes with known penetrance values.

Notably, our results demonstrate the improved performance of our LR-based method for assessing variant pathogenicity as it considers gene- and age-specific penetrance for carriers and age information for noncarriers. Using simulated case-control datasets, we show that the case-control LR method using age-specific breast cancer ORs from high-penetrance genes (e.g., *BRCA1* and *BRCA2*) outperforms other OR analysis methods. These observations reflect the fact that the method presented here is more suitable for the analysis of rare variants in a case-control setting. We further provide cut-offs of LRs in favor or against pathogenicity to be used in a real setting.

Analysis of the BCAC OncoArray data using our proposed method provided informative pathogenic ACMG/AMP classification evidence for six out of the 92 variants analyzed. Furthermore, 59 variants reached evidence against pathogenicity, something that is not directly measured as a code strength through classical calculations of ORs. Given that, *a priori*, the vast majority of rare sequence variants (e.g., *BRCA1* and *BRCA2*) will be neutral with respect to risk, this is a key advantage of our approach. In contrast, using logistic regression analysis, the informative ACMG/AMP classification criterion PS4 ($OR > 5.0$, p value < 0.05 , and CI not including 1.0) was reached only for two variants.

There are possible caveats that should be recognized. The selection of cases or controls for a family history of cancer would affect the carrier probabilities. The likelihood ratios would then be inaccurate, but in principle, this could be considered by incorporating family history into the likelihoods, if known. Depletion of cases with known pathogenic variants by prior clinical sequencing could also bias the likelihood ratios; therefore, the method is best applied to

TABLE 1: Statistically significant associated variants with breast cancer risk estimated by logistic regression (based on the ACMG/AMP PS4 criterion).

Gene	Variant_ID (GRCh37/hg19)	HGVS nucleotide	HGVS protein	Variant carriers		OR (95% CI)	p value
				Cases N (frequency)	Controls N (frequency)		
<i>BRCA2</i>	chr13_32937506_C_G	c.8167G>C	p. (Asp2723His)	18/72392 (2.49 × 10 ⁻⁴)	1/50680 (1.97 × 10 ⁻⁵)	12.30 (1.66-91.23)	0.014
	chr13_32954180_C_T	c.9154C>T	p. (Arg3052Trp)	10/72563 (1.38 × 10 ⁻⁴)	1/50779 (1.97 × 10 ⁻⁵)	8.32 (1.04-66.48)	0.045

Variant nomenclature according to *BRCA2* (NM_000059.3, NP_000050.2). OR: odds ratio; CI: confidence interval.

TABLE 2: Variants with informative LRs in favor of pathogenicity, estimated by the baseline analysis.

Gene	Variant_ID (GRCh37/hg19)	HGVS nucleotide	HGVS protein	Variant carriers		LR
				Cases N (frequency)	Controls N (frequency)	
BRCA1	chr17_41234451_A_G	c.4327C>T	p.(Arg1443*)	11/72558 (1.52×10^{-4})	3/50781 (5.91×10^{-5})	526.71
	chr17_41215947_T_G	c.5096G>T	p.(Arg1699Leu)	17/72560 (2.34×10^{-4})	3/50780 (5.91×10^{-5})	307.47
BRCA2	chr13_32937506_C_G	c.8167G>C	p.(Asp2723His)	18/72392 (2.49×10^{-4})	1/50680 (1.97×10^{-5})	8193.33
	chr13_32953453_A_G	c.8755-1G>A	p.?	3/72562 (4.13×10^{-5})	—	41.18
	chr13_32954180_C_T	c.9154C>T	p.(Arg3052Trp)	10/72563 (1.38×10^{-4})	1/50779 (1.97×10^{-5})	86.82
	chr13_32968940_A_T	c.9371A>T	p.(Asn3124Ile)	16/72548 (2.21×10^{-4})	—	3530.99

Variant nomenclature according to BRCA1 (NM_007294.4, NP_009225.1), BRCA2 (NM_000059.3, NP_000050.2). LR: likelihood ratio.

population-based case-control studies. For these reasons, we highlight the ACMG/AMP recommendation to review all available evidence for/against pathogenicity for a given variant and to denote obviously conflicting findings for different evidence types, before assigning a final classification. A conservative approach may be to assign case-control weight with a cap, for example, at moderate strength for or against pathogenicity.

Our method gains power in part because it leverages data on individual-level age, but we have to acknowledge that age is not always available. The method can be implemented more approximately by assuming that individuals with unknown information are of the same age, but this reduces power because the expectation that carriers of risk variants develop the disease at a younger age is then not utilised. It may also increase type I error because the likelihood ratio may be calculated for an age that is not appropriate for the dataset (for example, if the dataset consists predominantly of older individuals), although the type I error was still low in the simulations we considered. In the tabulated, preformatted calculator, we allow the user to incorporate individuals of unknown age at diagnosis or interview into any of the age groups specified. A conservative approach would be to include individuals of unknown age in the oldest age group. In this way, case-control genotypes from both existing data and new series, with and without age data, can be incorporated. However, we would like to emphasize that pooling series, particularly from different populations with different age/ethnicity structures or with different genotyping technologies, can lead to biased results. Ideally, datasets should be analysed separately, and the overall likelihood ratio generated by multiplying the study-specific likelihood ratios.

5. Conclusions

This manuscript describes in detail a novel method used for the calculation of the case-control LR to provide evidence of variant pathogenicity. This LR method is more informative compared to logistic regression analysis (or an OR calculation based on contingency tables and Fisher's exact test). It improves power as it considers age- and gene-specific penetrance values and age information for noncarriers and can provide both evidence in favor of and against pathogenicity. In addition, this method can also be imple-

mented towards the classification of VUS in any disease susceptibility gene for which disease penetrance has been reliably estimated. Open-access scripts and preformatted Excel calculators with code and instructions on how to use the method are available at the following address: <https://github.com/BiostatUnitCING/ccLR>.

Data Availability

All scripts allowing for replication of all analyses are available in the supplementary files and public repository (<https://github.com/BiostatUnitCING/ccLR>). Requests for the genotyped BCAC raw data can be made to the Data Access Coordination Committee (DACC) of BCAC (<http://bcac.ccge.medschl.cam.ac.uk/>).

Ethical Approval

This research has been approved by the Cyprus National Bioethics Committee. All participating studies were approved by the relevant ethics committees, and informed consent was obtained from study participants [12]. For NHS and NHS2, the study protocol was approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, as well as those of participating registries as required. The ethical approval for the POSH study is MREC/00/6/69, UKCRN ID: 1137.

Disclosure

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 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. J.L.H. is a National Health and Medical Research Council (NHMRC) Senior Principal Research Fellow. M.C.S. is a NHMRC Senior Research Fellow. 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 US Government or the BCFR. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health. The opinions, findings, and conclusions expressed herein are those of the authors and do not necessarily reflect the official views of the State of California, Department of Public Health, the National Cancer Institute, the National Institutes of Health, the Centers for Disease Control and Prevention or their Contractors and Subcontractors, or the Regents of the University of California, or any of its programs. The study was performed as part of the assignment of the Ministry of Science and Higher Education of the Russian Federation (No. AAAA-A16-116020350032-1). 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. Cases and their vital status were ascertained through the Victorian Cancer Registry and the Australian Institute of Health and Welfare, including the National Death Index and the Australian Cancer Database. ABCTB Investigators are Christine Clarke, Deborah Marsh, Rodney Scott, Robert Baxter, Desmond Yip, Jane Carpenter, Alison Davis, Nirmala Pathmanathan, Peter Simpson, J. Dinny Graham, and Mythily Sachchithanathan. Samples are made available to researchers on a nonexclusive basis.

Conflicts of Interest

The following authors declare conflicts not directly relevant to this work as stated below. Usha Menon has a patent (no: EP10178345.4) for Breast Cancer Diagnostics and held personal shares in Abcodia between 1st April 2011 and 30 October 2021. She is a member of the Research Advisory Panel, Yorkshire Cancer Research, Trial Steering Committee, NOVEL, and Scientific Advisory Board of Tina's Wish. She has received grants from the Medical Research Council (MRC), Cancer Research UK, the National Institute for Health Research (NIHR), and The Eve Appeal. She is part of research collaborations with iLOF, RNG Guardian and Micronoma. All other authors declare that they have no conflict of interests.

Authors' Contributions

Maria Zanti and Denise G. O'Mahony contributed equally to this work.

Acknowledgments

This work was cofunded by the Republic of Cyprus through the Research and Innovation Foundation (Project: CULTURE/AWARD-YR/0418/0017). ABS and MTP are supported by the Australian National Health and Medical

Research Funding (APP177524). DGO is funded by the Telethon Cyprus (Telethon Cyprus: 33173233) through the Cyprus Institute of Neurology and Genetics. BCAC and individual BCAC and ENIGMA studies, funders, and grant numbers are detailed in the main text. BCAC is funded by 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 is funded by the Government of Canada through the Genome Canada and the Canadian Institutes of Health Research, the Ministère de l'Économie et de l'Innovation du Québec through Genome Québec, and the Quebec Breast Cancer Foundation. Additional funding for BCAC is provided via the Confluence project which is funded with intramural funds from the National Cancer Institute Intramural Research Program and National Institutes of Health and via the CanRisk project which is funded from the Cancer Research UK (grant PPRPGM-Nov20/100002). Genotyping of the OncoArray was funded by the NIH Grant U19 CA148065 and Cancer Research UK Grant C1287/A16563, and the PERSPECTIVE project was supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research (grant GPH-129344), and the Ministère de l'Économie, Science et Innovation du Québec through the Genome Québec and the PSRSIIRI-701 grant and the Quebec Breast Cancer Foundation. Funding for iCOGS came from the European Community's Seventh Framework Programme under grant agreement no. 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, and C8197/A16565), the National Institutes of Health (CA128978), Post-Cancer GWAS initiative (1 U19 CA148537, 1 U19 CA148065, and 1 U19 CA148112—the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), and the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund. The Australian Breast Cancer Family Study (ABCFS) was supported by grant UM1 CA164920 from the National Cancer Institute (USA). The ABCFS was also supported by the National Health and Medical Research Council of Australia, the New South Wales Cancer Council, the Victorian Health Promotion Foundation (Australia), and the Victorian Breast Cancer Research Consortium. The ABCS study was supported by the Dutch Cancer Society (grants NKI 2007-3839 and 2009 4363). The Australian Breast Cancer Tissue Bank (ABCTB) was supported by the National Health and Medical Research Council of Australia, The Cancer Institute NSW, and the National Breast Cancer Foundation. The AHS study is supported by the intramural research program of the National Institutes of Health, the National Cancer Institute (grant number Z01-CP010119), and the National Institute of Environmental Health Sciences (grant number Z01-ES049030). The work of the BBCC was partly funded by the ELAN-Fond of the University Hospital of Erlangen. The BBCC is funded by the Cancer Research UK and Breast Cancer

Now and acknowledges NHS funding to the NIHR Biomedical Research Centre and the National Cancer Research Network (NCRN). The BCEES was funded by the National Health and Medical Research Council, Australia, and the Cancer Council Western Australia and acknowledges funding from the National Breast Cancer Foundation (JS). For the BCFR-NY, BCFR-PA, and BCFR-UT, this work was supported by grant UM1 CA164920 from the National Cancer Institute. The BCINIS study is supported in part by the Breast Cancer Research Foundation (BCRF). The BREast Oncology GALician Network (BREGAN) is funded by the Acción Estratégica de Salud del Instituto de Salud Carlos III FIS PI12/02125/Cofinanciado and FEDER PI17/00918/Cofinanciado FEDER; Acción Estratégica de Salud del Instituto de Salud Carlos III FIS Intrasalud (PI13/01136); Programa Grupos Emergentes, Cancer Genetics Unit, Instituto de Investigación Biomedica Galicia Sur; Xerencia de Xestión Integrada de Vigo-SERGAS, Instituto de Salud Carlos III, Spain, Grant 10CSA012E; Consellería de Industria Programa Sectorial de Investigación Aplicada, PEME I + D e I + D Suma del Plan Gallego de Investigación, Desarrollo e Innovación Tecnológica de la Consellería de Industria de la Xunta de Galicia, Spain, Grant EC11-192; Fomento de la Investigación Clínica Independiente, Ministerio de Sanidad, Servicios Sociales e Igualdad, Spain, Grant FEDER-Innterconecta; and Ministerio de Economía y Competitividad, Xunta de Galicia, Spain. The BSUCH study was supported by the Dietmar-Hopp Foundation, the Helmholtz Society, and the German Cancer Research Center (DKFZ). CBCS is funded by the Canadian Cancer Society (grant # 313404) and the Canadian Institutes of Health Research. CCGP is supported by funding from the University of Crete. The CECILE study was supported by the Fondation de France, Institut National du Cancer (INCa), Ligue Nationale contre le Cancer, Agence Nationale de Sécurité Sanitaire, de l'Alimentation, de l'Environnement et du Travail (ANSES), and Agence Nationale de la Recherche (ANR). The CGPS was supported by the Chief Physician Johan Boserup and Lise Boserup Fund, the Danish Medical Research Council, and Herlev and Gentofte Hospital. The American Cancer Society funds the creation, maintenance, and updating of the CPS-II cohort. The California Teachers Study (CTS) and the research reported in this publication were supported by the National Cancer Institute of the National Institutes of Health under award numbers U01-CA199277; P30-CA033572; P30-CA023100; UM1-CA164917; and R01-CA077398. The collection of cancer incidence data used in the California Teachers Study was supported by the California Department of Public Health pursuant to California Health and Safety Code Section 103885; Centers for Disease Control and Prevention's National Program of Cancer Registries, under cooperative agreement 5NU58DP006344; the National Cancer Institute's Surveillance, Epidemiology and End Results Program under contract HHSN261201800032I awarded to the University of California, San Francisco; contract HHSN261201800015I awarded to the University of Southern California; and contract HHSN261201800009I awarded to the Public Health Institute. The University of Westminster curates the DietCompLyf database was funded

by Against Breast Cancer Registered Charity No. 1121258 and the NCRN. 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 the Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Éducation Nationale, and Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), and Federal Ministry of Education and Research (BMBF) (Germany); the Hellenic Health Foundation and 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 and 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. FHRISK and PROCAS are funded from NIHR grant PGfAR 0707-10031. DGE, AH, and WGN are supported by the NIHR Manchester Biomedical Research Centre (IS-BRC-1215-20007). The GC-HBOC (German Consortium of Hereditary Breast and Ovarian Cancer) is supported by the German Cancer Aid (grant nos. 110837 and 70114178, coordinator: Rita K. Schmutzler, Cologne) and the Federal Ministry of Education and Research, Germany (grant no. 01GY1901). This work was also funded by the European Regional Development Fund and Free State of Saxony, Germany (LIFE-Leipzig Research Centre for Civilization Diseases, project numbers 713-241202, 713-241202, 14505/2470, and 14575/2470). The GENICA was funded by the Federal Ministry of Education and Research (BMBF), Germany, grants 01KW9975/5, 01KW9976/8, 01KW9977/0, and 01KW0114; the Robert Bosch Foundation, Stuttgart; Deutsches Krebsforschungszentrum (DKFZ), Heidelberg; the Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, and the Department of Internal Medicine, Johanniter GmbH Bonn, Johanniter Krankenhaus, Bonn, Germany. The GEPARSIXTO study was conducted by the German Breast Group GmbH. The GESBC was supported by the Deutsche Krebshilfe e.V. (70492) and the German Cancer Research Center (DKFZ). The HABCS study was supported by the Claudia von Schilling Foundation for Breast Cancer Research, by the Lower Saxonian Cancer Society, and by the Rudolf Bartling Foundation. The HEBCS was financially supported by the Helsinki University Hospital Research Fund, the Sigrid Juselius Foundation, and the Cancer Foundation Finland. The HMBCS was supported by a grant from the Friends of

Hannover Medical School and by the Rudolf Bartling Foundation. The HUBCS was supported by a grant from the German Federal Ministry of Research and Education (RUS08/017), B.M. was supported by a grants 17-44-020498 and 17-29-06014 of the Russian Foundation for Basic Research. D.P. was supported by a grant 18-29-09129 of the Russian Foundation for Basic Research. E.K was supported by the mega grant from the Government of Russian Federation (2020-220-08-2197). Financial support for KARBAC was provided through the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet, the Swedish Cancer Society, The Gustav V Jubilee Foundation, and Bert von Kantzows Foundation. The KARMA study was supported by Märit and Hans Rausing's Initiative Against Breast Cancer. 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. LMBC is supported by the "Stichting tegen Kanker." DL is supported by the FWO. The MABCS study is funded by the Research Centre for Genetic Engineering and Biotechnology "Georgi D. Efremov," MASA. The MARIE study was supported by the Deutsche Krebshilfe e.V. (70-2892-BR I, 106332, 108253, 108419, 110826, and 110828), the Hamburg Cancer Society, the German Cancer Research Center (DKFZ), and the Federal Ministry of Education and Research (BMBF) Germany (01KH0402). MBCSG is supported by grants from the Italian Association for Cancer Research (AIRC). The MCBCS was supported by the NIH grants R35CA253187, R01CA192393, R01CA116167, and R01CA176785, a NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA116201), and the Breast Cancer Research Foundation. The Melbourne Collaborative Cohort Study (MCCS) cohort recruitment was funded by the VicHealth and Cancer Council Victoria. The MCCS was further augmented by the Australian National Health and Medical Research Council grants 209057, 396414, and 1074383 and by infrastructure provided by the Cancer Council Victoria. The MEC was supported by the NIH grants CA63464, CA54281, CA098758, CA132839, and CA164973. The MISS study is supported by funding from ERC-2011-294576 Advanced grant, Swedish Cancer Society CAN 2018/675, Swedish Research Council, Local hospital funds, Berta Kamprad Foundation FBKS 2021-19, and Gunnar Nilsson. The MMHS study was supported by NIH grants CA97396, CA128931, CA140286, CA177150, and the NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA116201). MSKCC is supported by grants from the Breast Cancer Research Foundation and Robert and Kate Niehaus Clinical Cancer Genetics Initiative. The work of MTLGEBCS was supported by the Quebec Breast Cancer Foundation, the Canadian Institutes of Health Research for the "CIHR Team in Familial Risks of Breast Cancer" program—grant # CRN-87521 and the Ministry of Economic Development, Innovation and Export Trade—grant # PSR-SIIRI-701. The NBCS has received funding from the K.G. Jebsen Centre for Breast Cancer Research; the Research Council of Norway grant

193387/V50 (to A-L Børresen-Dale and V.N. Kristensen) and grant 193387/H10 (to A-L Børresen-Dale and V.N. Kristensen); South Eastern Norway Health Authority (grant 39346 to A-L Børresen-Dale); and the Norwegian Cancer Society (to A-L Børresen-Dale and V.N. Kristensen). The NBHS was supported by NIH grant R01CA100374. The biological sample preparation was conducted by the Survey and Biospecimen Shared Resource, which is supported by P30 CA68485. The Northern California Breast Cancer Family Registry (NC-BCFR) and Ontario Familial Breast Cancer Registry (OFBCR) were supported by a grant U01CA164920 from the USA National Cancer Institute of the National Institutes of Health. The Carolina Breast Cancer Study (NCBCS) was funded by the Komen Foundation, the National Cancer Institute (P50 CA058223, U54 CA156733, and U01 CA179715), and the North Carolina University Cancer Research Fund. The NHS was supported by the NIH grants P01 CA87969, UM1 CA186107, and U19 CA148065. The NHS2 was supported by NIH grants UM1 CA176726 and U19 CA148065. 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 the Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. Genotyping for PLCO was supported by the Intramural Research Program of the National Institutes of Health, NCI, Division of Cancer Epidemiology and Genetics. 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 the Cancer Research UK (grants C1275/A11699, C1275/C22524, C1275/A19187, and C1275/A15956) and Breast Cancer Campaign 2010PR62 and 2013PR044. The RBCS was funded by the Dutch Cancer Society (DDHK 2004-3124, DDHK 2009-4318). SEARCH is funded by the 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. The University of Cambridge has received salary support for PDPP from the NHS in the East of England through the Clinical Academic Reserve. Population-based controls were from the Multi-Ethnic Cohort (MEC) funded by grants from the Ministry of Health, Singapore, National University of Singapore, and National University Health System, Singapore. The Sister Study (SISTER) is supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01-ES044005 and Z01-ES049033). The Two Sister Study (2SISTER) was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01-ES044005 and Z01-ES102245), and also by a grant from Susan G. Komen for the Cure, grant FAS0703856. SKKDKFZS is supported by the DKFZ. The SMC is funded by the Swedish Cancer Foundation and the Swedish Research Council (VR 2017-00644) grant for the Swedish Infrastructure for Medical Population-based Life-course Environmental Research

(SIMPLER). The SZBCS was supported by a 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. The TNBCC was supported by a NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA116201), a grant from the Breast Cancer Research Foundation, a generous gift from the David F. and Margaret T. Grohne Family Foundation. The UCIBCS component of this research was supported by the NIH (CA58860, CA92044) and the Lon V Smith Foundation (LVS39420). 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. The UKOPS study was funded by The Eve Appeal (The Oak Foundation) and supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre and MRC Core Funding (MC_UU_00004/01). The USRT Study was funded by the Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. Contract grant sponsor AV was supported by the Spanish Instituto de Salud Carlos III (ISCIII) funding, an initiative of the Spanish Ministry of Economy and Innovation partially supported by the European Regional Development FEDER Funds (INT20/00071; PI19/01424); the Autonomous Government of Galicia (Consolidation and structuring program: IN607B); the Fundación Mutua Madrileña (call 2018) and the AECC (PRYES211091-VEGA). We thank all the women who took part in these studies and all the researchers, clinicians, technicians, and administrative staff who have enabled this work to be carried out. ABCFS thank Maggie Angelakos, Judi Maskiell, and Gillian Dite. ABCS thanks the Blood bank Sanquin, The Netherlands. BBCS thanks Eileen Williams, Elaine Ryder-Mills, and Kara Sargus. BCEES thanks Allyson Thomson, Christobel Saunders, Terry Slevin, BreastScreen Western Australia, Elizabeth Wylie, and Rachel Lloyd. The BCINIS study would not have been possible without the contributions of Dr. K. Landsman, Dr. N. Gronich, Dr. A. Flugelman, Dr. W. Saliba, Dr. F. Lejbkowitz, Dr. E. Liani, Dr. I. Cohen, Dr. S. Kalet, Dr. V. Friedman, Dr. O. Barnet of the NICCC in Haifa, and all the contributing family medicine, surgery, pathology, and oncology teams in all medical institutes in Northern Israel. The BREGAN study would not have been possible without the contributions of the following: Manuela Gago-Dominguez, Jose Esteban Castelao, Angel Carracedo, Victor Muñoz Garzón, Alejandro Novo Domínguez, Maria Elena Martinez, Sara Miranda Ponte, Carmen Redondo Marey, Maite Peña Fernández, Manuel Enguix Castelo, Maria Torres, Manuel Calaza (BREGAN), José Antúnez, Máximo Fraga, and the staff of the Department of Pathology and Biobank of the University Hospital Complex of Santiago-CHUS, Instituto de Investigación Sanitaria de Santiago, IDIS, Xerencia de Xestión Integrada de Santiago-SERGAS; Joaquín González-Carreró, and the staff of the Department of Pathology and Biobank of University Hospital Complex of Vigo, Instituto de Investigación Biomedica Galicia Sur, SERGAS, Vigo, Spain. The BSUCH study acknowledges the Principal Inves-

tigator, Barbara Burwinkel, and, thanks to Peter Bugert, Medical Faculty of Mannheim. CBCS thanks study participants, coinvestigators, collaborators, and staff of the Canadian Breast Cancer Study, as well as project coordinators Agnes Lai and Celine Morissette. CCGP thanks Styliani Apostolaki, Anna Margiolaki, Georgios Nintos, Maria Peraki, Georgia Saloustrou, Georgia Sevastaki, and Konstantinos Pompodakis. CGPS thanks the staff and participants of the Copenhagen General Population Study. For the excellent technical assistance: Dorthe Uldall Andersen, Maria Birna Arnadottir, Anne Bank, and Dorthe Kjeldgård Hansen. The Danish Cancer Biobank is acknowledged for providing infrastructure for the collection of blood samples for the cases. 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. The authors would like to thank the California Teachers Study Steering Committee that is responsible for the formation and maintenance of the Study within which this research was conducted. A full list of California Teachers Study (CTS) team members is available at <https://www.calteachersstudy.org/team>. DIETCOMPLYF thanks the patients, nurses and clinical staff involved in the study. The DietCompLyf study was funded by the charity Against Breast Cancer (Registered Charity Number 1121258) and the NCRN. 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, and Katja Butterbach. FHRISK and PROCAS thank NIHR for funding. The GENICA Network: Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University of Tübingen, Germany (RH, Hiltrud Brauch, Wing-Yee Lo), Department of Internal Medicine, Johanniter GmbH Bonn, Johanniter Krankenhaus, Bonn, Germany (Yon-Dschun Ko, Christian Baisch), Institute of Pathology, University of Bonn, Germany (Hans-Peter Fischer), Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany (UH), Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, Germany (Thomas Brüning, Beate Pesch, Sylvia Rabstein, Anne Lotz); and Institute of Occupational Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Germany (Volker Harth). HEBCS thanks Johanna Kiiski, Carl Blomqvist, Taru A. Muranen, Kirsimari Aaltonen, Karl von Smitten, and Irja Erkkilä. HUBCS thanks Darya Prokofyeva and Shamil Gantsev. KARMA and SASBAC thank the Swedish Medical Research Counsel. KBCP thanks Eija Myöhänen. LMBC thanks Gilian Peuteman, Thomas Van Brussel, Evy-Vanderheyden and Kathleen Corthouts. MABCS thanks Milena Jakimovska (RCGEB “Georgi D. Efremov”), Snezhana Smichkoska, Emilija Lazarova, Marina Iljoska (University Clinic of Radiotherapy and Oncology), Katerina Kubelka-

Sabit, Dzengis Jasar, Mitko Karadžozov (Adzibadem-Sistina Hospital), Andrej Arsovski, and Liljana Stojanovska (Re-Medika Hospital) for their contributions and commitment to this study. MARIE thanks Petra Seibold, Nadia Obi, Sabine Behrens, Ursula Eilber, and Muhabbet Celik. MBCSG (Milan Breast Cancer Study Group): Paolo Radice, Paolo Peterlongo, Siranoush Manoukian, Bernard Peissel, Jacopo Azzollini, Claudia Monaco, Daniela Zaffaroni, Bernardo Bonanni, Irene Feroce, Mariarosaria Calvello, Aliana Guerrieri Gonzaga, Monica Marabelli, Davide Bondavalli, and the personnel of the Cogentech Cancer Genetic Test Laboratory. The MCCS was made possible by the contribution of many people, including the original investigators, the teams that recruited the participants and continue working on follow-up, and the many thousands of Melbourne residents who continue to participate in the study. The MISS study group acknowledges the former Principal Investigator, Professor Håkan Olsson. We thank the coordinators, the research staff, and especially the MMHS participants for their continued collaboration on research studies in breast cancer. MSKCC thanks Marina Corines and Lauren Jacobs. MTLGEBCS would like to thank Martine Tranchant (CHU de Québec—Université Laval Research Center), Marie-France Valois, Annie Turgeon, and Lea Heguy (McGill University Health Center, Royal Victoria Hospital; McGill University) for DNA extraction, sample management, and skilful technical assistance. J.S. is the chairholder of the Canada Research Chair in Oncogenetics. The following are NBCS collaborators: Kristine K. Sahlberg (PhD), Anne-Lise Børresen-Dale (Prof. Em.), Lars Ottestad (MD), Rolf Kåresen (Prof. Em.), Dr. Ellen Schlichting (MD), Marit Muri Holmen (MD), Toril Sauer (MD), Vilde Haakensen (MD), Olav Engebråten (MD), Bjørn Naume (MD), Alexander Fosså (MD), Cecile E. Kiserud (MD), Kristin V. Reinertsen (MD), Åslaug Helland (MD), Margit Riis (MD), Jürgen Geisler (MD), OSBREAC, and Grethe I. Grenaker Alnæs (MSc). NBHS and SBCGS thank study participants and research staff for their contributions and commitment to the studies. We would like to thank the participants and staff of the NHS and NHS2 for their valuable contributions, as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, and WY. The authors assume full responsibility for analyses and interpretation of these data. The OFBCR thanks Teresa Selander, Nayana Weerasooriya, Anna Marie Mulligan, and Steve Gallinger. 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, and Michael Stagner. We thank staff in the Experimental Cancer Medicine Centre (ECMC) for supported the 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. PREFACE thanks Sonja Oeser and Silke Landrith. The RBCS thanks Jan-

net Blom, Saskia Pelders, Wendy J.C. Prager-van der Smissen, and the Erasmus MC Family Cancer Clinic. We thank the SEARCH and EPIC teams. SKDKFZS thanks all study participants, clinicians, family doctors, researchers, and technicians for their contributions and commitment to this study. We thank the SUCCESS Study teams in Munich, Duesseldorf, Erlangen, and Ulm. UCIBCS thanks Irene Masunaka. UKBGS thanks Breast Cancer Now and the Institute of Cancer Research for their support and funding of the Generations Study, as well as the study participants, study staff, and doctors, nurses and other health care providers and health information sources who have contributed to the study. We acknowledge NHS funding for the Royal Marsden/ICR NIHR Biomedical Research Centre.

Supplementary Materials

Supplementary 1. Supplementary File 1: Case-control likelihood ratio (LR) method presented in detail.

Supplementary 2. Supplementary Table S1: BCAC studies participating in the case-control likelihood ratio analysis with the number of cases and controls. Supplementary Table S2: Power calculations of the case-control likelihood ratio method and odds ratio analysis methods using a simulated case-control datasets. Supplementary Table S2a: Power calculations using a relative risk of 1. Supplementary Table S2b: Power calculations using relative risk of 2 to 10. Supplementary Table S3: Power calculations of the case-control likelihood ratio method and odds ratio analysis methods using simulated case-control datasets of assumed same age. Supplementary Table S3a: Power calculations using relative risk of 1. Supplementary Table S3b: Power calculations using a relative risk of 2 to 10. Supplementary Table S4: Case-control evidence for the 92 *BRCA1* and *BRCA2* variants included in the BCAC OncoArray dataset.

Supplementary 3. Supplementary Table S5: Case-control likelihood ratios and evidence for assignment to ACMG/AMP code strengths for the 92 *BRCA1* and *BRCA2* variants using country-specific analyses with different penetrance models.

Supplementary 4. Supplementary Figure S1: Performance of the case-control likelihood ratio method and odds ratio analysis in providing at least strong ACMG/AMP evidence in favor of pathogenicity ($LR \geq 18.7$), using simulated datasets of assumed same age. Power equals the probability of reaching at least strong pathogenic ACMG/AMP evidence. Genotype data simulations were carried out for causal variants conferring disease relative risk between 2 and 10. We performed 10,000 simulations for each case scenario. Results represent simulated case-control data for 20,000 (A–C), or 30,000 (D–F) or 50,000 (G–I) breast cancer cases and controls, and minor allele frequency of 0.00003 (A–G), 0.00005 (B–H), or 0.0001 (C–I). ccLR: case-control likelihood ratio; MAF: minor allele frequency; *N*: sample size. Supplementary Figure S2: performance of the case-control likelihood ratio method in providing ACMG/AMP evidence against pathogenicity using simulated datasets of assumed same age. Power equals the probability of reaching at least supporting benign ACMG/AMP evidence ($LR \leq 0.48$) when

the relative risk was set to 1. We performed 10,000 simulations for each case scenario. Results represent simulated case-control data for 20,000, 30,000, or 50,000 breast cancer cases and controls and minor allele frequency of 0.00003, 0.00005, or 0.0001. cCLR: case-control likelihood ratio; MAF: minor allele frequency; N : sample size.

References

- [1] D. M. Eccles, G. Mitchell, A. N. A. Monteiro et al., “*BRCA1* and *BRCA2* genetic testing—pitfalls and recommendations for managing variants of uncertain clinical significance,” *Annals of Oncology*, vol. 26, no. 10, pp. 2057–2065, 2015.
- [2] S. Richards, N. Aziz, S. Bale et al., “Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology,” *Genetics in Medicine*, vol. 17, no. 5, pp. 405–424, 2015.
- [3] S. V. Tavtigian, M. S. Greenblatt, S. M. Harrison et al., “Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework,” *Genetics in Medicine*, vol. 20, no. 9, pp. 1054–1060, 2018.
- [4] D. E. Goldgar, D. F. Easton, A. M. Deffenbaugh, A. N. Monteiro, S. V. Tavtigian, and F. J. Couch, “Integrated evaluation of DNA sequence variants of unknown clinical significance: application to *BRCA1* and *BRCA2*,” *The American Journal of Human Genetics*, vol. 75, no. 4, pp. 535–544, 2004.
- [5] D. E. Goldgar, D. F. Easton, G. B. Byrnes et al., “Genetic evidence and integration of various data sources for classifying uncertain variants into a single model,” *Human Mutation*, vol. 29, no. 11, pp. 1265–1272, 2008.
- [6] A. B. Spurdle, S. Healey, A. Devereau et al., “ENIGMA—evidence-based network for the interpretation of germline mutant alleles: an international initiative to evaluate risk and clinical significance associated with sequence variation in *BRCA1* and *BRCA2* genes,” *Human Mutation*, vol. 33, no. 1, pp. 2–7, 2012.
- [7] D. F. Easton, A. M. Deffenbaugh, D. Pruss et al., “A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the *BRCA1* and *BRCA2* breast cancer-predisposition genes,” *The American Journal of Human Genetics*, vol. 81, no. 5, pp. 873–883, 2007.
- [8] A. B. Spurdle, EMBRACE Group, F. J. Couch et al., “Refined histopathological predictors of *BRCA1* and *BRCA2* mutation status: a large-scale analysis of breast cancer characteristics from the BCAC, CIMBA, and ENIGMA consortia,” *Breast Cancer Research*, vol. 16, no. 6, p. 3419, 2014.
- [9] S. Belman, M. T. Parsons, A. B. Spurdle, D. E. Goldgar, and B. J. Feng, “Considerations in assessing germline variant pathogenicity using cosegregation analysis,” *Genetics in Medicine*, vol. 22, no. 12, pp. 2052–2059, 2020.
- [10] D. Thompson, D. F. Easton, and D. E. Goldgar, “A full-likelihood method for the evaluation of causality of sequence variants from family data,” *American Journal of Human Genetics*, vol. 73, no. 3, pp. 652–655, 2003.
- [11] M. T. Parsons, E. Tudini, H. Li et al., “Large scale multifactorial likelihood quantitative analysis of *BRCA1* and *BRCA2* variants: an ENIGMA resource to support clinical variant classification,” *Human Mutation*, vol. 40, no. 9, pp. 1557–1578, 2019.
- [12] K. Michailidou, S. Lindström, J. Dennis et al., “Association analysis identifies 65 new breast cancer risk loci,” *Nature*, vol. 551, no. 7678, pp. 92–94, 2017.
- [13] C. Hu, S. N. Hart, R. Gnanaolivu et al., “A population-based study of genes previously implicated in breast cancer,” *New England Journal of Medicine*, vol. 384, no. 5, pp. 440–451, 2021.
- [14] J. R. Coleman, J. Euesden, H. Patel, A. A. Folarin, S. Newhouse, and G. Breen, “Quality control, imputation and analysis of genome-wide genotyping data from the Illumina HumanCoreExome microarray,” *Briefings in Functional Genomics*, vol. 15, no. 4, pp. 298–304, 2016.
- [15] L. Dorling, S. Carvalho, J. Allen et al., “Breast cancer risk genes - association analysis in more than 113,000 women,” *New England Journal of Medicine*, vol. 384, no. 5, pp. 428–439, 2021.
- [16] A. Antoniou, P. D. P. Pharoah, S. Narod et al., “Average risks of breast and ovarian cancer associated with *BRCA1* or *BRCA2* mutations detected in case series unselected for family history: a combined analysis of 22 studies,” *American Journal of Human Genetics*, vol. 72, no. 5, pp. 1117–1130, 2003.
- [17] K. B. Kuchenbaecker, J. L. Hopper, D. R. Barnes et al., “Risks of breast, ovarian, and contralateral breast cancer for *BRCA1* and *BRCA2* mutation carriers,” *JAMA*, vol. 317, no. 23, pp. 2402–2416, 2017.
- [18] H. Li, C. Engel, M. de la Hoya et al., “Risks of breast and ovarian cancer for women harboring pathogenic missense variants in *BRCA1* and *BRCA2* compared with those harboring protein truncating variants,” *Genetics in Medicine*, vol. 24, no. 10, p. 2208, 2022.
- [19] A. Agresti, *Categorical Data Analysis*, Wiley, 2nd edition, 2003.