Full title: Physical activity, sedentary time and breast cancer risk: A Mendelian randomization study

Authors

Suzanne C. Dixon-Suen [1], Sarah J. Lewis [2, 3], Richard M. Martin [3-5], Dallas R. English [1, 6], Terry Boyle [7, 8], Graham G. Giles [1, 6, 9], Kyriaki Michailidou [10-12], Manjeet K. Bolla [12], Qin Wang [12], Joe Dennis [12], Michael Lush [12], ABCTB Investigators [13], Thomas U. Ahearn [14], Christine B. Ambrosone [15], Irene L. Andrulis [16, 17], Hoda Anton-Culver [18], Volker Arndt [19], Kristan J. Aronson [20], Annelie Augustinsson [21], Päivi Auvinen [22, 23], Laura E. Beane Freeman [14], Heiko Becher [24], Matthias W. Beckmann [25], Sabine Behrens [26], Marina Bermisheva [27], Carl Blomqvist [28, 29], Natalia V. Bogdanova [30-32], Stig E. Bojesen [33-35], Bernardo Bonanni [36], Hermann Brenner [19, 37, 38], Thomas Brüning [39], Saundra S. Buys [40], Nicola J. Camp [41], Daniele Campa [26, 42], Federico Canzian [43], Jose E. Castelao [44], Melissa H. Cessna [45, 46], Jenny Chang-Claude [26, 47], Stephen J. Chanock [14], Christine L. Clarke [48], Don M. Conroy [49], Fergus J. Couch [50], Angela Cox [51], Simon S. Cross [52], Kamila Czene [53], Mary B. Daly [54], Peter Devilee [55, 56], Thilo Dörk [31], Miriam Dwek [57], Diana M. Eccles [58], A. Heather Eliassen [59, 60], Christoph Engel [61, 62], Mikael Eriksson [53], D. Gareth Evans [63, 64], Peter A. Fasching [25, 65], Olivia Fletcher [66], Henrik Flyger [67], Lin Fritschi [68], Marike Gabrielson [53], Manuela Gago-Dominguez [69, 70], Montserrat García-Closas [14], José A. García-Sáenz [71], Mark S. Goldberg [72, 73], Pascal Guénel [74], Melanie Gündert [75, 76], Eric Hahnen [77, 78], Christopher A. Haiman [79], Lothar Häberle [25], Niclas Håkansson [80], Per Hall [53, 81], Ute Hamann [82], Steven N. Hart [83], Michelle Harvie [84], Peter Hillemanns [31], Antoinette Hollestelle [85], Maartje J. Hooning [85], Reiner Hoppe [86, 87], John L. Hopper [6], Anthony Howell [88], David J. Hunter [60, 89], Anna Jakubowska [90, 91], Wolfgang Janni [92], Esther M. John [93, 94], Audrey Jung [26], Rudolf Kaaks [26], Renske Keeman [95], Cari M. Kitahara [96], Stella Koutros [14], Peter Kraft [60, 97], Vessela N. Kristensen [98, 99], Katerina Kubelka-Sabit [100], Allison W. Kurian [93, 94], James V. Lacey [101, 102], Diether Lambrechts [103, 104], Loic Le Marchand [105], Annika Lindblom [106, 107], Sibylle Loibl [108], Jan Lubiński [90], Arto Mannermaa [109-111], Mehdi Manoochehri [82], Sara Margolin [81, 112], Maria Elena Martinez [70, 113], Dimitrios Mavroudis [114], Usha Menon [115], Anna Marie Mulligan [116, 117], Rachel A. Murphy [118, 119], NBCS Collaborators [98, 99, 120-129], Heli Nevanlinna [130], Ines Nevelsteen [131], William G. Newman [63, 64], Kenneth Offit [132, 133], Andrew

F. Olshan [134], Håkan Olsson [21§], Nick Orr [135], Alpa V. Patel [136], Julian Peto [137], Dijana Plaseska-Karanfilska [138], Nadege Presneau [57], Brigitte Rack [92], Paolo Radice [139], Erika Rees-Punia [136], Gad Rennert [140], Hedy S. Rennert [140], Atocha Romero [141], Emmanouil Saloustros [142], Dale P. Sandler [143], Marjanka K. Schmidt [95, 144], Rita K. Schmutzler [77, 78, 145], Lukas Schwentner [92], Christopher Scott [83], Mitul Shah [49], Xiao-Ou Shu [146], Jacques Simard [147], Melissa C. Southey [1, 9, 148], Jennifer Stone [6, 149], Harald Surowy [75, 76], Anthony J. Swerdlow [150, 151], Rulla M. Tamimi [60, 152], William J. Tapper [58], Jack A. Taylor [143, 153], Mary Beth Terry [154], Rob A.E.M. Tollenaar [155], Melissa A. Troester [134], Thérèse Truong [74], Michael Untch [156], Celine M. Vachon [157], Vijai Joseph [132], Barbara Wappenschmidt [77, 78], Clarice R. Weinberg [158], Alicja Wolk [80, 159], Drakoulis Yannoukakos [160], Wei Zheng [146], Argyrios Ziogas [18], Alison M. Dunning [49], Paul D.P. Pharoah [12, 49], Douglas F. Easton [12, 49], Roger L. Milne [1, 6, 9 †], Brigid M. Lynch [1, 6, 161 † *], on behalf of the Breast Cancer Association Consortium

§ deceased

* **Corresponding author**: Assoc. Prof. Brigid M. Lynch, Cancer Epidemiology Division, Cancer Council Victoria, 615 St Kilda Rd, Melbourne, VIC 3004, Australia. Email: brigid.lynch@cancervic.org.au; telephone: +61 3 9514 6209.

† Joint senior authors

Affiliations

[1] Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, 615 St Kilda Road, 3004, Australia.

[2] Bristol Medical School, Department of Population Health Sciences, University of Bristol, Bristol, Oakfield House, Oakfield Grove, BS8 2BN, UK.

[3] MRC Integrative Epidemiology Unit, University of Bristol, Bristol, Oakfield House, Oakfield Grove, BS8 2BN, UK.

[4] Bristol Medical School, Department of Population Health Sciences, University of Bristol, Bristol, Canynge Hall, 39 Whatley Road, BS8 2PS, UK.

[5] NIHR Biomedical Research Centre, University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol, Bristol, Oakfield House, Oakfield Grove, BS8 2BN, UK.

[6] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, 207 Bouverie Street, 3010, Australia.

[7] Allied Health and Human Performance, University of South Australia, Adelaide, South Australia, North Terrace, 5000, Australia.

[8] Australian Centre for Precision Health, University of South Australia Cancer Research Institute, Adelaide, South Australia, North Terrace, 5000, Australia.

[9] Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, 246 Clayton Road, 3168, Australia.

[10] Biostatistics Unit, The Cyprus Institute of Neurology & Genetics, Nicosia, 6 Iroon Avenue, 2371 Ayios Dometios, Nicosia, 2371, Cyprus.

[11] Cyprus School of Molecular Medicine, The Cyprus Institute of Neurology & Genetics, Nicosia, 6 Iroon Avenue, 2371 Ayios Dometios, 2371, Cyprus.

[12] Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, 2 Worts' Causeway, CB1 8RN, UK.

[13] Australian Breast Cancer Tissue Bank, Westmead Institute for Medical Research, University of Sydney, Sydney, New South Wales, 176 Hawkesbury Road, 2145, Australia.

[14] Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, 9609 Medical Center Dr, 20850, USA.

[15] Roswell Park Comprehensive Cancer Center, Buffalo, NY, Elm & Carlton Streets, 14263, USA.

[16] Fred A. Litwin Center for Cancer Genetics, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, ON, 600 University Avenue, M5G 1X5, Canada.

[17] Department of Molecular Genetics, University of Toronto, Toronto, ON, 1 King's College Circle, M5S 1A8, Canada.

[18] Department of Medicine, Genetic Epidemiology Research Institute, University of California Irvine, Irvine, CA, 224 Irvine Hall, 92617, USA.

[19] Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Im Neuenheimer Feld 280, 69120, Germany.

[20] Department of Public Health Sciences, and Cancer Research Institute, Queen's University, Kingston, ON, 10 Stuart Street, K7L 3N6, Canada.

[21] Department of Cancer Epidemiology, Clinical Sciences, Lund University, Lund, Barngatan 4, Skånes universitetssjukhus, 222 42, Sweden.

[22] Department of Oncology, Cancer Center, Kuopio University Hospital, Kuopio, Puijonlaaksontie 2, 70210, Finland.

[23] Institute of Clinical Medicine, Oncology, University of Eastern Finland, Kuopio, Yliopistonranta 1, 70210, Finland.

[24] Institute of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Hamburg, Martinistr. 52, 20246, Germany.

[25] Department of Gynecology and Obstetrics, Comprehensive Cancer Center ER-EMN, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Universitaetsstrasse 21-23, 91054, Germany.

[26] Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Im Neuenheimer Feld 280, 69120, Germany.

[27] Institute of Biochemistry and Genetics, Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, 71 prosp. Oktyabrya, 450054, Russia.

[28] Department of Oncology, Helsinki University Hospital, University of Helsinki, Helsinki, Haartmaninkatu 4, 00290, Finland.

[29] Department of Oncology, Örebro University Hospital, Örebro, 70185, Sweden.

[30] Department of Radiation Oncology, Hannover Medical School, Hannover, Carl-Neuberg-Straße 1, 30625, Germany.

[31] Gynaecology Research Unit, Hannover Medical School, Hannover, Carl-Neuberg-Straße 1, 30625, Germany.

[32] N.N. Alexandrov Research Institute of Oncology and Medical Radiology, Minsk, Settlement of Lesnoy-2, 223040, Belarus.

[33] Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Herlev Ringvej 75, 2730, Denmark.

[34] Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Herlev Ringvej 75, 2730, Denmark.

[35] Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Blegdamsvej 3B, 2200, Denmark.

[36] Division of Cancer Prevention and Genetics, IEO, European Institute of Oncology IRCCS, Milan, via Ripamonti 435, 20141, Italy.

[37] Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Im Neuenheimer Feld 280, 69120, Germany.

[38] German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Im Neuenheimer Feld 280, 69120, Germany.

[39] Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, Bürkle-de-la-Camp-Platz 1, 44789, Germany.

[40] Department of Medicine, Huntsman Cancer Institute, Salt Lake City, UT, 2000 Circle of Hope, 84112, USA.

[41] Department of Internal Medicine and Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, 2000 Circle of Hope, 84112, USA.

[42] Department of Biology, University of Pisa, Pisa, 56126, Italy.

[43] Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Im Neuenheimer Feld 280, 69120, Germany.

[44] Oncology and Genetics Unit, Instituto de Investigacion Sanitaria Galicia Sur (IISGS), Xerencia de Xestion Integrada de Vigo-SERGAS, Vigo, Estrada Clara Campoamor nº 341, 36312, Spain.

[45] Department of Pathology, Intermountain Medical Center, Intermountain Healthcare, Murray, UT, 5121 S. Cottonwood Street, 84107, USA.

[46] Intermountain Biorepository, Intermountain Healthcare, Salt Lake City, UT, 824 West Fine Drive, Suite 400, 84119, USA.

[47] Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Martinistraße 52, 20246, Germany.

[48] Westmead Institute for Medical Research, University of Sydney, Sydney, New South Wales, 176 Hawkesbury Road, 2145, Australia.

[49] Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, 2 Worts' Causeway, CB1 8RN, UK.

[50] Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, 200 First St. SW, 55905, USA.

[51] Sheffield Institute for Nucleic Acids (SInFoNiA), Department of Oncology and Metabolism, University of Sheffield, Sheffield, Western Bank, S10 2TN, UK.

[52] Academic Unit of Pathology, Department of Neuroscience, University of Sheffield, Sheffield, Western Bank, S10 2TN, UK.

[53] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Karolinska Univ Hospital, 171 65, Sweden.

[54] Department of Clinical Genetics, Fox Chase Cancer Center, Philadelphia, PA, 333 Cottman Ave, 19111, USA.

[55] Department of Pathology, Leiden University Medical Center, Leiden, Albinusdreef 2, 2333 ZA, The Netherlands.

[56] Department of Human Genetics, Leiden University Medical Center, Leiden, Albinusdreef 2, 2333 ZA, The Netherlands.

[57] School of Life Sciences, University of Westminster, London, 309 Regent Street, W1B 2HW, UK.

[58] Faculty of Medicine, University of Southampton, Southampton, 12 University Road, SO17 1BJ, UK.

[59] Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 181 Longwood Ave, Boston, MA, 02115, USA.

[60] Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA, 02115, USA.

[61] Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Härtelstraße 16-18, 04107, Germany.

[62] LIFE - Leipzig Research Centre for Civilization Diseases, University of Leipzig, Leipzig, Philipp-Rosenthal-Straße 27, 04103, Germany.

[63] Division of Evolution and Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, Oxford Road, M13 9WL, UK.

[64] North West Genomics Laboratory Hub, Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, Oxford Road, M13 9WL, UK.

[65] David Geffen School of Medicine, Department of Medicine Division of Hematology and Oncology, University of California at Los Angeles, Los Angeles, CA, 10833 Le Conte Ave, 90095, USA.

[66] The Breast Cancer Now Toby Robins Research Centre, The Institute of Cancer Research, London, 123 Old Brompton Road, SW7 3RP, UK.

[67] Department of Breast Surgery, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Herlev Ringvej 75, 2730, Denmark.

[68] School of Public Health, Curtin University, Perth, Western Australia, Kent Street, 6102, Australia.

[69] Fundación Pública Galega de Medicina Xenómica, Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Travesía da Choupana S/N, 15706, Spain.

[70] Moores Cancer Center, University of California San Diego, La Jolla, CA, 3855 Health Sciences Drive, 92037, USA.

[71] Medical Oncology Department, Hospital Clínico San Carlos, Instituto de Investigación Sanitaria San Carlos (IdISSC), Centro Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Calle del Prof Martín Lagos, 28040, Spain.

[72] Department of Medicine, McGill University, Montréal, QC, 1001 Decarie Boulevard, H4A 3J1, Canada.

[73] Division of Clinical Epidemiology, Royal Victoria Hospital, McGill University, Montréal, QC, 1001 Decarie Boulevard, H4A 3J1, Canada.

[74] Center for Research in Epidemiology and Population Health (CESP), Team Exposome and Heredity, INSERM, University Paris-Saclay, Villejuif, 39 rue Camille Desmoulins, 94805, France.

[75] Molecular Epidemiology Group, C080, German Cancer Research Center (DKFZ), Heidelberg, Im Neuenheimer Feld 280, 69120, Germany.

[76] Molecular Biology of Breast Cancer, University Womens Clinic Heidelberg, University of Heidelberg, Heidelberg, Im Neuenheimer Feld 440, 69120, Germany.

[77] Center for Familial Breast and Ovarian Cancer, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Kerpener Str. 62, 50937, Germany.

[78] Center for Integrated Oncology (CIO), Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Kerpener Str. 62, 50937, Germany.

[79] Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, 1975 Zonal Ave, 90033, USA.

[80] Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Nobels väg 13, 171 77, Sweden.

[81] Department of Oncology, Södersjukhuset, Stockholm, 118 83, Sweden.

[82] Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), Heidelberg, Im Neuenheimer Feld 580, 69120, Germany.

[83] Department of Health Sciences Research, Mayo Clinic, Rochester, MN, 200 First St. SW, 55905, USA.

[84] Prevent Breast Cancer Research Unit, Manchester University Hospital Foundation NHS Trust, Manchester, Southmoor Road, M23 9LT, UK.

[85] Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, Dr. Molewaterplein 40, 3015 GD, The Netherlands.

[86] Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Auerbachstr. 112, 70376, Germany.

[87] University of Tübingen, Tübingen, Geschwister-Scholl-Platz, 72074, Germany.

[88] Division of Cancer Sciences, University of Manchester, Manchester, M13 9PL, UK.

[89] Nuffield Department of Population Health, University of Oxford, Oxford, OX3 7LF, UK.

[90] Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Unii Lubelskiej 1, 71-252, Poland.

[91] Independent Laboratory of Molecular Biology and Genetic Diagnostics, Pomeranian Medical University, Szczecin, ul. Powsta?ców Wlkp 72, 71-252, Poland.

[92] Department of Gynaecology and Obstetrics, University Hospital Ulm, Ulm, Prittwitzstrasse 43 89075, Germany.

[93] Department of Epidemiology & Population Health, Stanford University School of Medicine, Stanford, CA, 259 Campus Drive, 94305, USA.

[94] Department of Medicine, Division of Oncology, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, 780 Welch Road, Suite CJ250C, 94304, USA.

[95] Division of Molecular Pathology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Amsterdam, Plesmanlaan 121, 1066 CX, The Netherlands.

[96] Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, 20892, USA.

[97] Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA, 677 Huntington Avenue, 02115, USA.

[98] Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Ullernchausseen 70, 0379, Norway.

[99] Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Kirkeveien 166, 0450, Norway.

[100] Department of Histopathology and Cytology, Clinical Hospital Acibadem Sistina, Skopje, Skupi 5a, 1000, Republic of North Macedonia.

[101] Department of Computational and Quantitative Medicine, City of Hope, Duarte, CA, 1500 E. Duarte Ed, 91010, USA.

[102] City of Hope Comprehensive Cancer Center, City of Hope, Duarte, CA, 1500 E. Duarte Ed, 91010, USA.

[103] VIB Center for Cancer Biology, Leuven, Herestraat 46, Box 912, 3001, Belgium.

[104] Laboratory for Translational Genetics, Department of Human Genetics, University of Leuven, Leuven, Oude Markt 13, 3000, Belgium.

[105] Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, 701 Ilalo St, 96813, USA.

[106] Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Karolinska Univ Hospital, 171 76, Sweden.

[107] Department of Clinical Genetics, Karolinska University Hospital, Stockholm, 171 76, Sweden.

[108] German Breast Group, GmbH, Neu Isenburg, Martin-Behaim-Str. 12, 63263, Germany.

[109] Translational Cancer Research Area, University of Eastern Finland, Kuopio, Yliopistonranta 1, 70210, Finland.

[110] Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Yliopistonranta 1, 70210, Finland.

[111] Biobank of Eastern Finland, Kuopio University Hospital, Kuopio, Finland.

[112] Department of Clinical Science and Education, Södersjukhuset, Karolinska Institutet, Stockholm, 118 83, Sweden.

[113] Department of Family Medicine and Public Health, University of California San Diego, La Jolla, CA, 9500 Gilman Drive, 92093, USA.

[114] Department of Medical Oncology, University Hospital of Heraklion, Heraklion, Stavrakis Voutes 711 10, Greece.

[115] MRC Clinical Trials Unit, Institute of Clinical Trials & Methodology, University College London, London, 2nd Floor, 90 High Holborn, London, WC1V 6LJ, UK.

[116] Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, 1 King's College Circle, M5S 1A8, Canada.

[117] Laboratory Medicine Program, University Health Network, Toronto, ON, 200 Elizabeth Street, M5G 2C4, Canada.

[118] Cancer Control Research, BC Cancer, Vancouver, BC, 675 West 10th Avenue, V5Z 1L3, Canada.

[119] School of Population and Public Health, University of British Columbia, Vancouver, BC, 2206 East Mall, V6T 1Z3, Canada.

[120] Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital-Radiumhospitalet, Oslo, Ullernchausseen 70, 0379, Norway.

[121] Department of Research, Vestre Viken Hospital, Drammen, Hauges gate 89A, 3019, Norway.

[122] Section for Breast- and Endocrine Surgery, Department of Cancer, Division of Surgery, Cancer and Transplantation Medicine, Oslo University Hospital-Ullevål, Oslo, Kirkeveien 166, 0450, Norway.

[123] Department of Radiology and Nuclear Medicine, Oslo University Hospital, Oslo, Ullernchausseen 70, 0379, Norway.

[124] Department of Pathology, Akershus University Hospital, Lørenskog, Sykehusveien 25, 1478, Norway.

[125] Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital, Oslo, Ullernchausseen 70, 0379, Norway.

[126] Department of Oncology, Division of Surgery, Cancer and Transplantation Medicine, Oslo University Hospital-Radiumhospitalet, Oslo, Ullernchausseen 70, 0379, Norway.

[127] National Advisory Unit on Late Effects after Cancer Treatment, Oslo University Hospital-Radiumhospitalet, Oslo, Ullernchausseen 70, 0379, Norway.

[128] Department of Oncology, Akershus University Hospital, Lørenskog, Sykehusveien 25, 1478, Norway.

[129] Breast Cancer Research Consortium, Oslo University Hospital, Oslo, Ullernchausseen 70, 0379, Norway.

[130] Department of Obstetrics and Gynecology, Helsinki University Hospital, University of Helsinki, Helsinki, Haartmaninkatu 8, 00290, Finland.

[131] Leuven Multidisciplinary Breast Center, Department of Oncology, Leuven Cancer Institute, University Hospitals Leuven, Leuven, Oude Markt 13, 3000, Belgium.

[132] Clinical Genetics Research Lab, Department of Cancer Biology and Genetics, Memorial Sloan Kettering Cancer Center, New York, NY, 1275 York Avenue, 10065, USA.

[133] Clinical Genetics Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, 1275 York Avenue, 10065, USA.

[134] Department 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.

[135] Centre for Cancer Research and Cell Biology, Queen's University Belfast, Belfast, Ireland, BT7 1NN, UK.

[136] Department of Population Science, American Cancer Society, Atlanta, GA, 250 Williams Street NW, 30303, USA.

[137] Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, Keppel Street, WC1E 7HT, UK.

[138] Research Centre for Genetic Engineering and Biotechnology 'Georgi D. Efremov', MASA, Skopje, Boulevard Krste Petkov Misirkov, 1000, Republic of North Macedonia.

[139] Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Research, Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Milan, Via Giacomo Venezian 1, 20133, Italy.

[140] Clalit National Cancer Control Center, Carmel Medical Center and Technion Faculty of Medicine, Haifa, 7 Michal St., 35254, Israel.

[141] Medical Oncology Department, Hospital Universitario Puerta de Hierro, Madrid, Calle Manuel de Falla, 1, 28222, Spain.

[142] Department of Oncology, University Hospital of Larissa, Larissa, 411 10, Greece.

[143] Epidemiology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, 111 T.W. Alexander Drive, 27709, USA.

[144] Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Amsterdam, Plesmanlaan 121, 1066 CX, The Netherlands.

[145] Center for Molecular Medicine Cologne (CMMC), Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Robert-Koch-Str. 21, 50931, Germany.

[146] Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, 1161 21st Ave S # D3300, 37232, USA.

[147] Genomics Center, Centre Hospitalier Universitaire de Québec – Université Laval Research Center, Québec City, QC, 2705 Laurier Boulevard, G1V 4G2, Canada.

[148] Department of Clinical Pathology, The University of Melbourne, Melbourne, Victoria, Cnr Grattan Street and Royal Parade, 3010, Australia.

[149] Genetic Epidemiology Group, School of Population and Global Health, University of Western Australia, Perth, Western Australia, 35 Stirling Hwy, 6009, Australia.

[150] Division of Genetics and Epidemiology, The Institute of Cancer Research, London, SM2 5NG, UK.

[151] Division of Breast Cancer Research, The Institute of Cancer Research, London, SW7 3RP, UK.

[152] Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, 10065, USA.

[153] Epigenetic and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, 111 T.W. Alexander Drive, 27709, USA.

[154] Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, 722 West 168th Street, 10032, USA.

[155] Department of Surgery, Leiden University Medical Center, Leiden, Albinusdreef 2, 2333 ZA, The Netherlands.

[156] Department of Gynecology and Obstetrics, Helios Clinics Berlin-Buch, Berlin, Schwanebecker Chaussee 50, 13125, Germany.

[157] Department of Health Science Research, Division of Epidemiology, Mayo Clinic, Rochester, MN, 200 First Street SW, Harwick 6, 55905, USA.

[158] Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, 111 T.W. Alexander Drive, 27709, USA.

[159] Department of Surgical Sciences, Uppsala University, Uppsala, 751 05, Sweden.

[160] Molecular Diagnostics Laboratory, INRASTES, National Centre for Scientific Research 'Demokritos', Athens, Neapoleos 10, Ag. Paraskevi, 15310, Greece.

[161] Physical Activity Laboratory, Baker Heart and Diabetes Institute, Melbourne, Victoria, 75 Commercial Road, 3004, Australia.

Article type: Original Article

Wordcount: Abstract, 250 words; Manuscript, 3,009 words

References: 36

ABSTRACT

Objectives: Physical inactivity and sedentary behaviour are associated with higher breast cancer risk in observational studies, but ascribing causality is difficult. Mendelian randomization (MR) assesses causality by simulating randomized trial groups using genotype. We assessed whether lifelong physical activity or sedentary time, assessed using genotype, may be causally associated with breast cancer risk overall, pre/post-menopause, and by case-groups defined by tumour characteristics.

Methods: We performed two-sample inverse-variance-weighted MR using individual-level Breast Cancer Association Consortium case-control data from 130,957 European-ancestry women (69,838 invasive cases), and published UK Biobank data (n=91,105-377,234). Genetic instruments were single nucleotide polymorphisms (SNPs) associated in UK Biobank with wrist-worn accelerometer-measured overall physical activity (n_{snps}=5) or sedentary time (n_{snps}=6), or accelerometer-measured (n_{snps}=1) or self-reported (n_{snps}=5) vigorous physical activity.

Results: Greater genetically-predicted overall activity was associated with lower breast cancer risk, overall (OR=0.59; 95%CI 0.42-0.83 per-standard deviation [SD; ~8 milligravities acceleration]) and for most case-groups. Genetically-predicted vigorous activity was associated with lower risk of pre/perimenopausal breast cancer (OR=0.62; 95%CI 0.45-0.87, ≥3 vs. 0 self-reported days/week), with consistent estimates for most case-groups. Greater genetically-predicted sedentary time was associated with higher hormone-receptor-negative tumour risk (OR=1.77; 95%CI 1.07-2.92 per-SD [~7% time spent sedentary]), with elevated estimates for most case-groups. Results were robust to sensitivity analyses examining pleiotropy (including weighted-median-MR, MR-Egger).

Conclusion: Our study provides strong evidence that greater overall physical activity, greater vigorous activity, and lower sedentary time are likely to reduce breast cancer risk. More widespread adoption of active lifestyles may reduce the burden from the most common cancer in women.

Keywords: Breast cancer; Physical activity; Sedentary time; Mendelian randomization; Causal inference

What are the new findings?

- This study, using individual-level outcome data from the Breast Cancer Association Consortium, shows that greater levels of physical activity and less sedentary time are likely to reduce breast cancer risk.
- A systematic Mendelian randomization approach was used to minimise the effect of biases such as confounding which are likely to have affected previous studies.
- Our study provides strong evidence to suggest that physical activity and sedentary behaviour are causally associated with breast cancer.

How might it impact on clinical practice in the future?

A stronger cancer-control focus on promoting active lifestyles is likely to reduce the high burden from breast cancer.

Introduction

Greater physical activity and less sedentary time are associated with lower breast cancer risk in observational studies. International and national agencies have concluded that physical activity may reduce breast cancer risk, particularly postmenopausal disease, with associations strongest for vigorous activity.(1-3) Sedentary (sitting/reclining) time, a distinct exposure affecting 'active' and 'inactive' people, has been less well-studied, with conflicting findings.(4, 5) Physical inactivity or excess sitting may plausibly influence breast cancer initiation and/or growth. However, whether observed associations are causal or produced by biases (e.g. confounding, selection bias, reverse causation) is unclear. Mendelian randomization (MR) can simulate randomized controlled trials using observational data by substituting genotypes, which are randomly assigned at meiosis (before conception), as instruments (proxies) for exposures of interest.(6) Subject to meeting specific assumptions of instrumental variable analysis,(7) some of which can be investigated using sensitivity analyses (see Methods), MR can minimise confounding and reverse causation, potentially providing stronger evidence for causal inference.

A recent MR study assessed physical activity and breast cancer risk overall and by oestrogen-receptor (ER) status,(8) but did not examine other breast tumour types, vigorous activity, or sedentary time. We aimed to appraise the causal nature of associations between overall activity, vigorous activity, and sedentary time, and breast cancer risk, overall and by menopausal status, stage, grade, morphology, and molecular subtypes defined by hormone-receptor (ER, progesterone [PR]) and human epidermal growth factor receptor-2 (HER2) status.

Methods

Data sources

We performed two-sample MR using individual-level data from 130,957 European-ancestry women (69,838 with invasive breast cancers; 6,667 with in situ breast cancers; 54,452 controls) from 76 Breast Cancer Association Consortium (BCAC) studies (Tables 1, S1)(outcome dataset), and genetic estimates for movement-related exposures from published genome-wide association studies (GWAS) using UK Biobank data (exposure datasets; n=91,105-377,234).(9-11) Instruments were single-nucleotide polymorphisms (SNPs) associated in the UK Biobank GWAS with overall physical activity (all movement), vigorous physical activity, or sedentary time (Table S2).

Exposures

Overall physical activity

As our primary physical activity instrument we used five SNPs associated with overall activity ($p<5x10^{-8}$) in a prior GWAS of accelerometer-assessed movement in the UK Biobank (n=91,105), which explain over 0.06% of the variance in activity.(9) Doherty and colleagues assessed overall activity as average vector magnitude (milligravities) per 30-second period,(9, 12) with mean (standard deviation, SD) 29.0 (8.0) milligravities among women in UK Biobank.(13) One SD (8 milligravities) corresponds to ~50 minutes of moderate (e.g. brisk walking) activity per week.(8)

For comparability with the previous MR study on this topic,(8) we used an expanded set of ten SNPs as a secondary instrument for overall activity. These SNPs were associated at relaxed significance ($p<5x10^{-7}$) with the accelerometer-assessed overall activity phenotype in a separate UK Biobank GWAS of physical activity by Klimentidis and colleagues.(10, 11)

Commented [MSD1]: On calculating statistical power, I found an error in my reported percent variance explained (r-squared ZX) for overall activity and sedentary behaviour.

Will replace this with 0.1%, as explained below.

Sarah/Richard, please let me know if the following is correct/appropriate:

The GWAS reported 0.06% variance explained, but that was for the three SNPs associated at p<5x10-9. We know the 5 SNPs we used explain more (but how much was not reported by the GWAS).

I've now used the TwoSampleMR package function *get_r_from_pn* to estimate rsq for each of the SNPs in the instrument.

Since these SNPs are independent, is it appropriate to add the rsq values for the individual SNPs to obtain a total rsq for the instrument?

Adding the rsq for the top three SNPs indeed gives 0.06% (specifically 0.06245), as the GWAS reported.

Adding the rsq for all five SNPs yields **0.099116.** Can I report this as the instrument percent variance explained (I'd **add a sentence in the Methods citing the R package and function**)?

Vigorous physical activity

Klimentidis and colleagues identified one SNP associated ($p<5x10^{-9}$) with high-intensity movement, assessed as the fraction of 30-second intervals containing accelerations over 425 milligravities.(10) This threshold approximates expenditure output for vigorous activity (>6 metabolic equivalents of task [METs]).(14) This SNP explains approximately 0.02% of variance in high-intensity movement. They identified five SNPs associated ($p<5x10^{-9}$) with self-reported engagement in vigorous activity for at least ten minutes ≥3 vs. 0 days/week (n=377,234), (10) which explain approximately 0.06% of variance in this exposure. We examined both instruments as complementary measures for vigorous activity, each likely subject to different error (weak instrument or reporting bias).

Commented [MSD2]: Newly estimated using the TwoSampleMR package as explained above (rsq = 0.02015)

Added this because it's necessary to calculate power

Commented [MSD3]: Newly estimated using the TwoSampleMR package (sum of snp rsq = 0.06482)

Sedentary time

Doherty and colleagues applied machine-learning models, trained using body-camera and diary data, to UK Biobank accelerometry data to identify sedentary periods (sitting/reclining; MET-value typically ≤1.5).(9, 13) They identified six SNPs associated ($p<5x10^{-8}$) with the probability of engaging in sedentary behaviours, defined as the ratio of sedentary-to-total 30-second periods.(9) On average UK Biobank women spent 34.6% (SD=7.2%) of their time sedentary.(13) We used these six variants, explaining over 0.08% of variance in sedentariness,(9) as our sedentary time instrument.

Commented [MSD4]: Similar to overall activity above:

Will replace with 0.12%

The Doherty GWAS reported 0.08% variance explained, but this was for the 4 SNPs significant at p<5x10-9.

I've used the TwoSampleMR function I mentioned above to estimate rsq for the 6 SNPs. Using this method the top 4 sum to 0.08%, as the GWAS authors reported. The top 6 sum to 0.115083.

Outcomes

We estimated breast cancer risk overall, by menopausal status, and by case-groups defined by molecular/morphological subtype, stage, or grade at diagnosis, using BCAC clinical data to assign case-groups according to hypotheses arising from the literature. We defined separate case/control groups for invasive pre/peri-menopausal (n=23,999 cases;

17,686 controls) and postmenopausal (n=45,839 cases; 36,766 controls) breast cancers, using age at diagnosis/interview (</≥50 years) to assign missing menopausal status (27%). We examined subtypes separately by hormone-receptor (HR) status (ER+/- n=46,528/11,246; PR+/- n=34,891/16,432) and HER2 status (+/- n=6,945/33,214), and jointly including HER2-enriched (ER-/PR-/HER2+; n=1,974) and triple-negative (ER-/PR-/HER2-; n=4,964) cancers. We examined invasive ductal/lobular cancers (n=42,223/8,795), ductal carcinoma in situ (n=3,510), and risk by stage (stage I, n=17,583; stage II, n=15,992; stages III/IV, n=4,553) and grade (well/moderately differentiated, n=34,647; poorly/undifferentiated, n=16,432).

SNP-exposure (UK Biobank) and SNP-outcome (BCAC) associations

We extracted or derived estimates of association (beta coefficients, standard errors [SEs]) between SNPs and exposures from the UK Biobank GWAS publications,(9, 10) standardised to refer to the trait-increasing allele. Where required,(10) we converted estimates to per-SD changes in activity/sedentary time using UK Biobank activity data.

Genotypes in BCAC were determined using the OncoArray, an Illumina custom array, and imputed using IMPUTE2.(15) We harmonised UK Biobank and BCAC data so SNP-exposure and SNP-outcome estimates related to the same allele, using allele frequency information to resolve strand-ambiguous SNPs where possible (i.e., unless allele frequencies were 45%-55%). For each SNP, we derived trait-specific effect-allele dosages (range 0-2) by summing alleles predicting more activity (activity instruments) or sedentary time (sitting instrument). We assessed the association between each SNP and each outcome from individual-level BCAC data by fitting logistic regression models, adjusted for age at diagnosis (cases) or interview (controls), country, and ten principal components of genetic population structure (accounting for genetic substructure within Europeans), obtaining beta

coefficients and SEs for use in the MR analysis. Table 1 summarises the BCAC studies and participants.

Statistical analysis

We used SNP-exposure and SNP-outcome beta coefficients and SEs to estimate odds ratios (OR) and 95% confidence intervals (CI) of the effect of each trait on each outcome. For single SNPs, we divided the SNP-outcome association by the SNP-exposure association to obtain the causal estimate (Wald ratio). For multi-SNP instruments, we used inverse-variance weighted (IVW)-MR, which averages Wald ratios across SNPs, weighted by SNP-exposure beta coefficients.(16-18) IVW-MR assumes all instruments are valid or that pleiotropy is balanced.(17) We performed case-only analyses to test for differences between subtype-specific estimates.

Core assumptions of MR, which can be investigated using sensitivity analyses, are that the instrument: predicts exposure; is not associated with confounders of the exposure/outcome association; and influences the outcome only via the exposure (no horizontal pleiotropy) (6, 7, 19), summarised in Figure S1. We undertook sensitivity analyses to assess the robustness of our findings and the potential for violations of assumptions, most critically horizontal pleiotropy. We calculated Cochran's Q-statistic for between-SNP heterogeneity of effects. We applied complementary methods relaxing different MR assumptions, weighted-median MR (allows invalid instruments)(20) and MR-Egger (allows horizontal pleiotropy)(21). We inspected per-SNP causal estimates (scatter, forest plots) and leave-one-out analyses to identify SNPs distorting results. We performed MR Pleiotropy Residual Sum and Outlier (MR-PRESSO) to identify outlying SNPs with evidence of horizontal pleiotropy (global-pleiotropy and SNP-outlier tests $p<0.05$).(22) We examined the effect of excluding SNPs with imputation quality <0.9. We checked whether SNPs are

associated with other relevant traits (possible confounders, adiposity, cancer risk) or gene expression using the NHGRI-EBI GWAS Catalog(23) and PhenoScanner.(24, 25)

Data preparation and analyses were performed using R software (R Foundation for Statistical Computing, Vienna), including the 'MendelianRandomization'(18) and 'MR-PRESSO' packages.(22) Statistical power was calculated using the mRnd Mendelian randomization power calculation online tool.(26) Further details are in Supplementary Methods (Online Resource).

Results

Overall physical activity

Greater genetically-predicted physical activity was associated with lower risk of invasive breast cancer (OR=0.48;95%CI 0.30-0.78 per-SD [~8 milligravities] in overall activity), with no clearly differential effects by menopausal status, molecular subtype, morphology, stage, or grade (Table 2). We observed ORs less than 1 for all outcomes, including ER+ (OR=0.45;95%CI 0.25-0.83), PR+ (OR=0.43;95%CI 0.22-0.85), HER2+ (OR=0.48;95%CI 0.26-0.89), and HR+/HER2+ (OR=0.42;95%CI 0.20-0.88) disease. Weighted-median MR and MR-Egger results were broadly consistent (Table S3).

Heterogeneity of causal effects between SNPs was evident for some outcomes (Cochran's-Q $p_{het}<0.05$)(Table 2); this was resolved after removing outliers rs564819152 (associated previously with ovarian cancer; outlying for six outcomes) or rs6775319 (one outcome), detected by MR-PRESSO, per-SNP, and leave-one-out analyses (Figures S2-S3; Table S4). Evidence of protective associations remained strong after excluding rs564819152 (Table 2). Outlier-corrected results (OR [95%CI]) were 0.59 (0.42-0.83) for all invasive

breast cancer, 0.60 (0.43-0.85) for ER+, and 0.58 (0.37-0.91) for PR+ disease (HER2+ and HR+/HER2+ analyses had no outlying SNPs).

The protective effects were consistent across leave-one-out analyses (Table S4). SNPs were not associated in prior GWAS with confounders of the exposure/outcome relationship, but two had been identified in an ovarian cancer GWAS (Table S5). Excluding these made little difference to results (Table S4). Two SNPs have been reported to be associated ($p<5x10^{-8}$) with adiposity in UK Biobank,(24, 25, 27) consistent with reduced adiposity being a downstream effect of increased activity (Table S5).

Results were similar although slightly attenuated using the expanded ten-SNP instrument(10)(Table S6-S7). Estimates generally remained protective upon removing outlying SNPs detected by pleiotropy investigations (IVW heterogeneity tests [Table S6], MR-PRESSO, per-SNP effects [Figures S4-S5], leave-one-out analysis [Table S8]). Most estimates were similar (Table S8) upon excluding one SNP with imputation quality <0.9 (Table S8). Four of the ten SNPs were associated in prior GWAS with confounders (including height, alcohol intake, education) or cancer risk. Furthermore, rs55657917 is associated with gene expression in breast tissue, including in two genes associated with breast cancer risk (Table S5).(23-25) However, results excluding potentially confounded SNPs were relatively unchanged (Table S8). For four SNPs, the activity-increasing allele is associated with reduced adiposity in UK Biobank.(27)

Vigorous physical activity

There was little evidence that genetically-predicted acceleration over 425 milligravities (one SNP) was associated with risk of breast cancer, with wide confidence intervals crossing one, although most estimates were in the protective direction (Table 3).

The activity-increasing allele has been associated in GWAS(24, 25, 27) with greater height and decreased adiposity (Table S5).

There was weak evidence that genetically-predicted self-reported vigorous activity was associated with decreased breast cancer risk overall (OR=0.83;95%CI 0.69-1.01, ≥3 days/week vs. none), and ORs for most case-groups were less than 1 (Table 3). A protective association was seen for pre/perimenopausal breast cancer (OR=0.62;95%CI 0.45-0.87), with little evidence for an association with postmenopausal breast cancer risk (OR=0.95;95%CI 0.75-1.19) (p=0.82 for the difference in pre/peri- vs. post-menopausal estimates). A protective relationship was seen for PR+ disease (OR=0.77;95%CI 0.61-0.98). There was little evidence of pleiotropic effects (Table 3, S9-S10) except one outlier in modelling in situ cancers (Figures S6-S7), a SNP previously associated with height, age at menarche, and adiposity (Table S5).(24, 25, 27) After excluding this SNP, the in situ OR was elevated (from OR=0.94;95%CI 0.43-2.08 to OR=1.30;0.72-2.34)(Table 3); other estimates remained similar (Table S10). Excluding one SNP associated in UK Biobank GWAS with past smoking and childhood height (Table S5)(24, 25, 27) attenuated estimates slightly (Table S10). The association with pre/perimenopausal cancers remained substantially inverse (protective), with confidence intervals that did not cross the null, in all sensitivity analyses (Table S10).

Sedentary time

The estimates for genetically-predicted sedentary time were elevated (in the direction of increased risk) for almost every case-group, although CIs were wide (Table 4). Greater sedentary time was associated with higher risk of hormone-receptor-negative (HR-) tumours (OR=1.77;95%CI 1.07-2.92 per-SD [~7% time spent sedentary]), including triple-negative (ER-/PR-/HER2-) cancers (OR=2.04;95%CI 1.06-3.93) (p=0.11 for the difference in ORs by HR-status). ORs were substantially elevated for in situ cancers (OR=1.75;95%CI 1.00-3.07),

specifically ductal carcinoma in situ (OR=2.11;95%CI 0.99-4.49). The point estimate was elevated for stage I tumours (OR=1.62;95%CI 0.99-2.65), with little evidence of association with stage III/IV (OR=0.91;95%CI 0.45-1.84) (p=0.25 for the difference in estimates for risk of stage I vs stage III/IV tumours).

Heterogeneity between SNPs was not detected (all p_{het}>0.2)(Table 4), all MR methods produced broadly consistent results (Table S11), and MR-PRESSO did not identify outlying SNPs. Estimates were consistently elevated across leave-one-out analyses, including after omitting: one SNP correlated with a physical activity variant; one SNP predicting greater education and adiposity in prior GWAS(24, 25, 27, 28); or one strand-ambiguous SNP with minor allele frequency ~50%, for which effect-allele harmonisation was not definitive (Table S12). After excluding a SNP with imputation quality <0.9, which may have been an outlier for PR+ analyses (Figures S8-S9; MR-Egger $p_{pleiotropy}$=0.046 for PR+), point estimates for PR+ and most other outcomes including HR-, triple-negative, and in situ cancers, moved further from null (Table S12). Estimates for HR- and in situ cancers remained substantially elevated in all sensitivity analyses (Table S12).

Discussion

Main findings

We conducted a Mendelian randomization study using individual-level data on 130,957 women. We found that women with genetic variants predisposing them to be more active had lower breast cancer risk overall and for most case-groups defined by tumour subtypes, stage, or grade. Effect estimates for vigorous physical activity were in the protective direction for most types of breast cancer; reporting more frequent vigorous activity was associated with reduced risk of pre/perimenopausal breast cancer. Women with genetic variants predisposing them to more sedentary time had higher risk of HR- breast cancer, but

there was no strong evidence of differences in association by subtypes and weak evidence of an increased risk overall.

Strengths and limitations

A strength of our study is the use of individual-level BCAC data, which permitted examination of more outcomes than previously possible. Large sample sizes are another strength. BCAC is the largest collaboration of breast cancer studies, and we employed the most powerful available genetic instruments identified by the largest GWAS for movement-related behaviours, likely improving precision of our estimates. While statistical power was limited by the low predictive ability of the genetic instruments available (we had 52% power to detect expected effects for overall activity and overall breast cancer risk, and less power for other exposure/outcome combinations; Table S13), there is no larger datasets available to increase power. The UK Biobank studies are the only GWAS of accelerometer-assessed movement, which substantially decreases measurement error compared to self-report. Measurement error in assessing genotype is typically very low.

The UK Biobank GWAS which identified our instruments used wrist-worn accelerometers, which may not capture ambulation as well as hip-worn accelerometers;(29) while this may have slightly affected precision, no superior data are available. Gene-exposure associations were estimated from a population (UK Biobank) including men, but no strong evidence of sexual dimorphism was reported in UK Biobank,(9) so we assume that SNP-exposure estimates adequately reflect associations in women. While our instruments predict only a small fraction of variance in exposure, any weak-instrument bias would have biased estimates towards the null and cannot explain our findings.(19) Some contributing studies did not provide sufficient data on cancer diagnosis to classify cases into case groups (for example tumour subtype or stage), and therefore numbers included in these analyses were much lower.

Women without these tumour-specific outcome data may have differed from those included in analyses.

Due to the nature of the data and study design, we estimated odds ratios as the measure of effect, which in some circumstances can be prone to non-collapsibility and sparse-data bias.(30, 31) These issues are most severe when many covariates are included in models (which the present analysis did not do), and when outcomes are neither rare nor very common (many of the outcomes we investigated are rare, limiting the extent of noncollapsibility). Overall activity and sedentary time results for pre/peri- and post-menopausal breast cancer (the only sub-outcome where all participants could be classified), demonstrate a slight pattern of noncollapsibility, where the odds ratio for all invasive breast cancers does not lie between the odds ratios for each group separately. This is not a bias but a mathematical property of odds ratios.(30)

Implications

This analysis extends findings from a recent MR study of overall physical activity and breast cancer risk overall and by ER-status, using BCAC summary data.(8) Our study, using individual-level data, confirmed those findings, and showed that the risk reduction holds across multiple subtypes. Our study also examined vigorous activity and sedentary time, not previously studied in relation to breast cancer risk using MR. We assessed associations with multiple outcomes (overall and by case-group) and our results may be subject to false positives. There was no strong evidence of differences in association by case-group.

While MR may provide estimates which more closely reflect underlying causal relationships, core assumptions must be satisfied before causal conclusions can be drawn. We satisfied the first (instrument predicts exposure) by selecting genome-wide significant SNPs identified by the largest GWAS of our traits of interest. We maximised the possibility of

meeting the second (no confounding) by checking whether the SNPs were reported in prior GWAS of possible confounders (known breast cancer risk factors), and confirming that results remained consistent after excluding any SNPs that were (e.g., smoking [vigorous activity analyses], education [sedentary behaviour analyses]). We interrogated the third assumption (instrument influences outcome only through exposure) using several pleiotropy-detection approaches, acting on detected violations, and confirming consistency of results from methods relaxing this assumption. Our conclusions remained unchanged following exclusion of potentially-pleiotropic SNPs.

Several SNPs in the analyses were associated with adiposity in previous GWAS. While we cannot rule out horizontal pleiotropy (SNPs influencing adiposity independently of physical activity/sedentary time), vertical pleiotropy (same causal pathway) is more plausible; reduced adiposity is a downstream effect of increased physical activity. Vertical pleiotropy does not violate MR assumptions and excluding vertically-pleiotropic variants may distort causal estimates.(19) Nevertheless, previous MR analysis has shown evidence of a bi-directional relationship between overall activity and adiposity.(9)

Although it is possible that our findings resulted from chance, our results for physical activity are consistent with observational studies, which have suggested a 20-25% breast cancer risk reduction for the most vs. least active women, with evidence of dose-response.(3, 32) Our findings support this and furthermore suggest that these relationships are likely to be causal. The observational evidence for risk reduction, particularly for premenopausal breast cancer, is strongest for vigorous physical activity, suggesting that vigorous activity may be particularly important in preventing carcinogenesis.(3, 33) Short bouts of intense activity may be more protective than equivalent energy expenditure accumulated from light activity. We found that self-reported vigorous activity was associated with lower pre/perimenopausal

breast cancer risk and found weak evidence for a protective effect of vigorous activity overall. Future studies should continue to explore this with more powerful instruments.

For sedentary time, the observational evidence is sparse and inconsistent. Our results, which minimise likelihood of confounding (e.g. by unhealthy diet), are suggestive of a causal association with elevated risk of breast cancer, particularly for HR- and in situ cancer. While there is debate about the independence of physical activity and sedentary behaviour, they have different determinants and correlates and are often treated as separate traits. In our study the genetic instruments for sedentary behaviour and physical activity were mostly distinct; removing one SNP which predicted both traits did not change our findings, suggesting that both behaviours independently influence breast cancer risk.

Robust causal inference should triangulate findings across methods.(34) Our findings must be considered in light of biological plausibility. A reasonable body of mechanistic evidence supports numerous causal pathways between physical activity and breast cancer risk. Pathways involving adiposity, metabolic dysfunction, sex hormones, and inflammation have been most thoroughly described.(35-37) Mechanisms linking sedentary time and cancer are likely to at least partially overlap with those underpinning the physical activity relationship.(38, 39) Our findings cannot shed light on drivers of carcinogenesis. We saw suggestive differences by HR-status, but this may be a chance finding. Known adiposity-related SNPs did not seem to unduly influence our results, perhaps indicating that multiple pathways are important.

Conclusion

Increasing physical activity and reducing sedentary time are already recommended for cancer prevention. Our study adds further evidence that such behavioural changes are likely to lower future breast cancer incidence. A stronger cancer-control focus on physical activity

and sedentary time as modifiable cancer risk factors is warranted, given the heavy burden of disease attributed to the most common cancer in women.

Statements

Acknowledgements

BCAC: 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. The COGS study would not have been possible without the contributions of the following: Andrew Berchuck (OCAC), Rosalind A. Eeles, Ali Amin Al Olama, Zsofia Kote-Jarai, Sara Benlloch (PRACTICAL), Georgia Chenevix-Trench, Antonis Antoniou and Lesley McGuffog (CIMBA), Andrew Lee, and Ed Dicks, Craig Luccarini and the staff of the Centre for Genetic Epidemiology Laboratory, Javier Benitez, Anna Gonzalez-Neira and the staff of the CNIO genotyping unit, Daniel C. Tessier, Francois Bacot, Daniel Vincent, Sylvie LaBoissière and Frederic Robidoux and the staff of the McGill University and Génome Québec Innovation Centre, Sune F. Nielsen, Borge G. Nordestgaard, and the staff of the Copenhagen DNA laboratory, and Julie M. Cunningham, Sharon A. Windebank, Christopher A. Hilker, Jeffrey Meyer and the staff of Mayo Clinic Genotyping Core Facility. ABCFS thank Maggie Angelakos, Judi Maskiell, Gillian Dite. ABCS thanks the Blood bank Sanquin, The Netherlands. ABCTB Investigators: Christine Clarke, Rosemary Balleine, Robert Baxter, Stephen Braye, Jane Carpenter, Jane Dahlstrom, John Forbes, Soon Lee, Debbie Marsh, Adrienne Morey, Nirmala Pathmanathan, Rodney Scott, Allan Spigelman, Nicholas Wilcken, Desmond Yip. Samples are made available to researchers on a non-exclusive basis. BBCS thanks Eileen Williams, Elaine Ryder-Mills, Kara Sargus. BCEES thanks Allyson Thomson, Christobel Saunders, Terry Slevin, BreastScreen Western Australia, Elizabeth Wylie, 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. Lejbkowicz, 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 BREOGAN 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 (BREOGAN), 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 Xestion 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 Investigacion Biomedica Galicia Sur, SERGAS, Vigo, Spain. The BSUCH study acknowledges the Principal Investigator, Barbara Burwinkel, and thanks Peter Bugert, Medical Faculty Mannheim. CBCS thanks study participants, co-investigators, collaborators and staff of the Canadian Breast Cancer Study, and project coordinators Agnes Lai and Celine Morissette. CCGP thanks Styliani Apostolaki, Anna Margiolaki, Georgios Nintos, Maria Perraki, Georgia Saloustrou, Georgia Sevastaki, Konstantinos Pompodakis. CGPS thanks staff and participants of the Copenhagen General Population Study. For the excellent technical assistance: Dorthe Uldall Andersen, Maria Birna Arnadottir, Anne Bank, 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 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, Katja Butterbach. FHRISK thanks NIHR for funding. GC-HBOC thanks Stefanie Engert, Heide Hellebrand, Sandra Kröber and LIFE - Leipzig Research Centre for Civilization Diseases (Markus Loeffler,

Joachim Thiery, Matthias Nüchter, Ronny Baber). The GENICA Network: Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University of Tübingen, Germany [Hiltrud Brauch, Wing-Yee Lo], Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, 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 [Ute Hamann], 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, Taru A. Muranen, Kristiina Aittomäki, Kirsimari Aaltonen, Karl von Smitten, Irja Erkkilä. HMBCS thanks Peter Hillemanns, Hans Christiansen and Johann H. Karstens. HUBCS thanks Shamil Gantsev. KARMA and SASBAC thank the Swedish Medical Research Counsel. KBCP thanks Eija Myöhänen, Helena Kemiläinen. LMBC thanks Gilian Peuteman, Thomas Van Brussel, EvyVanderheyden and Kathleen Corthouts. MABCS thanks Milena Jakimovska (RCGEB “Georgi D. Efremov”), Snezhana Smichkoska, Emilija Lazarova (University Clinic of Radiotherapy and Oncology), Mitko Karadjozov (Adzibadem-Sistina Hospital), Andrej Arsovski and Liljana Stojanovska (Re-Medika Hospital) for their contributions and commitment to this study. MARIE thanks Petra Seibold, Dieter Flesch-Janys, Judith Heinz, Nadia Obi, Alina Vrieling, Sabine Behrens, Ursula Eilber, Muhabbet Celik, Til Olchers and Stefan Nickels. MBCSG (Milan Breast Cancer Study Group): Paolo Peterlongo, Siranoush Manoukian, Bernard Peissel, Jacopo Azzollini, Benedetta Beltrami, Daniela Zaffaroni, 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. 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, 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 Chair holder of the Canada Research Chair in Oncogenetics. The following are NBCS Collaborators: Kristine K. Sahlberg (PhD), Lars Ottestad (MD), Anne-Lise Børresen-Dale (PhD, Prof.Em.), 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. 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, and those of participating registries as required. 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, WY. The authors assume full responsibility for analyses and interpretation of these data. 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. The LUMC survival data were retrieved from the Leiden hospital-based cancer registry system (ONCDOC) with the help of Dr. J. Molenaar. 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. PREFACE thanks Sonja Oeser and Silke Landrith. PROCAS thanks NIHR for funding. 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, Ian Brock, Sabapathy Balasubramanian and Malcolm W.R. Reed. We thank the SEARCH and EPIC teams. SKKDKFZS 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, Duessldorf, Erlangen and Ulm. SZBCS thanks Ewa Putresza. UBCS thanks all study participants, the ascertainment, laboratory and research informatics teams at Huntsman Cancer Institute and Intermountain Healthcare, and Justin Williams, Brandt Jones, Myke Madsen, Stacey Knight and Kerry Rowe 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.

Funding

This work was supported by the following agencies. Funders had no role in study design, data collection, analysis, interpretation, writing of the report, or the decision to submit the paper for publication.

BCAC is funded by Cancer Research UK [C1287/A16563, C1287/A10118], the European Union's Horizon 2020 Research and Innovation Programme (grant numbers 634935 and 633784 for BRIDGES and B-CAST respectively), and by the European Community´s Seventh Framework Programme under grant agreement number 223175 (grant number HEALTH-F2-2009-223175) (COGS). 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.

Genotyping of the OncoArray was funded by the NIH Grant U19 CA148065, and Cancer UK Grant C1287/A16563 and the PERSPECTIVE project 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 Genome Québec and the PSRSIIRI-701 grant, and the Quebec Breast Cancer Foundation. Funding for the iCOGS infrastructure came from: the European Community's Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118,

C1287/A10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, and Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund. The DRIVE Consortium was funded by U19 CA148065.

The Australian Breast Cancer Family Study (ABCFS) was supported by grant UM1 CA164920 from the National Cancer Institute (USA). 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 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. 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 ABCS study was supported by the Dutch Cancer Society [grants NKI 2007-3839; 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 ELAN-Fond of the University Hospital of Erlangen. The BBCS is funded by 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, BCFR-UT

this work was supported by grant UM1 CA164920 from the National Cancer Institute. 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 BCINIS study is supported in part by the Breast Cancer Research Foundation (BCRF). The BREast Oncology GAlician Network (BREOGAN) is funded by Acción Estratégica de Salud del Instituto de Salud Carlos III FIS PI12/02125/Cofinanciado FEDER; Acción Estratégica de Salud del Instituto de Salud Carlos III FIS Intrasalud (PI13/01136); Programa Grupos Emergentes, Cancer Genetics Unit, Instituto de Investigacion Biomedica Galicia Sur. Xerencia de Xestion 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; and Grant FEDER-Innterconecta. Ministerio de Economia 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 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), 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 and the research reported in this publication were supported by the National Cancer Institute of the National Institutes of Health under award number U01-CA199277; P30-CA033572; P30-CA023100; UM1-CA164917; and R01-CA077398. 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 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 opinions, findings, and conclusions expressed herein are those of the author(s) 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 University of Westminster curates the DietCompLyf database 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: Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education 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. Additional cases were recruited in the context of the VERDI study, which

was supported by a grant from the German Cancer Aid (Deutsche Krebshilfe). FHRISK is funded from NIHR grant PGfAR 0707-10031. The GC-HBOC (German Consortium of Hereditary Breast and Ovarian Cancer) is supported by the German Cancer Aid (grant no 110837, coordinator: Rita K. Schmutzler, Cologne). 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, 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, as well as the Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, 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 Finnish Cancer Society, and the Sigrid Juselius Foundation. 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 grant 17-44-020498, 17-29-06014 of the Russian Foundation for Basic Research, and the study was performed as part of the assignment of the Ministry of Science and Higher Education of the Russian Federation (№AAAA-A16-116020350032-1). 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 Rausings Initiative Against Breast Cancer. The KBCP was financially supported by the special Government Funding (EVO) 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, 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 CA192393, CA116167, CA176785 an NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer [CA116201], and the Breast Cancer Research Foundation and a generous gift from the David F. and Margaret T. Grohne Family Foundation. The Melbourne Collaborative Cohort Study (MCCS) cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further augmented by Australian National Health and Medical Research Council grants 209057, 396414 and 1074383 and by infrastructure provided by Cancer Council Victoria. 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. The MEC was supported by NIH grants CA63464, CA54281, CA098758, CA132839 and CA164973. The MISS study is supported by funding from ERC-2011-294576 Advanced grant, Swedish Cancer Society, Swedish Research Council, Local hospital funds, Berta Kamprad Foundation, Gunnar Nilsson. 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 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. Biological sample preparation was conducted 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 grants 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 Carolina Breast Cancer Study (NCBCS) was funded by Komen Foundation, the National Cancer Institute (P50 CA058223, U54 CA156733, U01 CA179715), and the North Carolina University Cancer Research Fund. The NHS was supported by 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 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 Cancer Research UK (grants C1275/A11699, C1275/C22524, C1275/A19187, C1275/A15956 and Breast Cancer Campaign 2010PR62, 2013PR044. PROCAS is funded from NIHR grant PGfAR 0707-10031. 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 Tissue Bank. SEARCH is 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. The University of Cambridge has received salary support for PDPP from the NHS in the East of England through the Clinical Academic Reserve. 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 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 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. UBCS was supported by funding from National Cancer Institute grants R01 CA163353 (to N.J. Camp) and the Women's Cancer Center at the Huntsman Cancer Institute (HCI) which is funded in part by the Huntsman Cancer Foundation. Data collection is also made possible by the Utah Population Database (UPDB), Intermountain Healthcare and the Utah Cancer Registry (UCR). Support for the UPDB is provided by the University of Utah, HCI, and the Comprehensive Cancer Center Support grant NCI P30 CA42014. The UCR is funded by the NCI's SEER Program, Contract No. HHSN261201800016I, with additional support from the US Center for Disease Control and Prevention's National Program of Cancer Registries, Cooperative Agreement No. NU58DP0063200, the University of Utah and Huntsman Cancer 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) with investigators supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre. The USRT Study was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA.

RMM was supported by a Cancer Research UK (C18281/A19169) programme grant (the Integrative Cancer Epidemiology Programme). RMM is supported in part by the National Institute for Health Research Bristol Biomedical Research Centre. The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the UK National Institute for Health Research (NIHR) or the Department of Health and Social Care. DGE and WGN are supported by the all Manchester NIHR Biomedical Research Centre (IS-BRC-1215-20007). BML is funded by the Victorian Cancer Agency (MCRF-18005).

Competing interests

Matthias W. Beckmann conducts research funded by Amgen, Novartis and Pfizer. Peter A. Fasching conducts research funded by Amgen, Novartis and Pfizer. He received honoraria from Roche, Novartis and Pfizer. Allison W. Kurian declares research funding to her institution from Myriad Genetics for an unrelated project (funding dates 2017-2019). Sibylle Loibl declares grants and honoraria paid to her institution from Amgen, Novartis, Pfizer, Roche, and, outside the submitted work, grants and/or honoraria paid to her institution from AbbVie, Celgene, Seattle Genetics, PrIME/Medscape, Daiichi-Sankyo, Lilly, Samsung, BMS, Puma, Immunomedics, AstraZeneca, Pierre Fabre, Merck, GlaxoSmithKlein, EirGenix, and Bayer, and personal fees from Chugai; Dr. Loibl also has a patent EP14153692.0 pending. Usha Menon declares stock ownership in Abcodia Ltd. Rachel A. Murphy has been a consultant for Pharmavite. No other authors have conflicts to declare.

Contributorship

Project conception – BML, RLM; Project design – SCD, BML, RLM, SJL, RMM, DRE, TB; Acquisition, analysis, or interpretation of data for the work – all authors; initial drafting of manuscript – SCD, BML, RLM, SJL, RMM, DRE, TB; critical input – all authors; final approval of manuscript – all authors.

Ethics approval

This analysis and each contributing study received approval from the appropriate institutional review board or committee.

Data sharing statement

The data used in this study are de-identified patient data from 76 studies participating in the Breast Cancer Association Consortium (BCAC). Enquiries about accessing BCAC data can be directed to the BCAC coordinators at the University of Cambridge: https://bcac.ccge.medschl.cam.ac.uk/

Patient involvement

Patient co-production was not adopted for this large multi-study analysis. We thank all participants for providing their data to the contributing BCAC studies.

Tables

Table 1. Characteristics of 76 Breast Cancer Association Consortium studies, and 130,957 study participants, included in the individual-level analysis

Study acronym [a]	Country	Diagnosis years	Invasive cases (N)	In situ cases (N)	Controls (N)
ABCFS	Australia	1963-2013	1,117	-	187
ABCTB	Australia	2004-2013	920	6	375
BCEES	Australia	2009-2011	783	-	834
MCCS	Australia	1981-2012	870	180	978
HMBCS	Belarus	1994-2007	212	-	249
LMBC	Belgium	1994-2011	784	21	1,268
CBCS	Canada	2005-2009	568	108	817
MTLGEBCS	Canada	2007-2011	341	-	170
OFBCR	Canada	1967-2015	1,721	2	643
CGPS	Denmark	1981-2012	1,408	3	716
EPIC	Europe (Multiple countries)	n.r.	3,435	412	3,597
HEBCS	Finland	1997-2012	281	-	177
KBCP	Finland	1990-2012	522	34	245
CECILE	France	2005-2007	280	26	159
BBCC	Germany	1988-2013	403	8	253
BSUCH	Germany	1990-2013	252	1	168
ESTHER	Germany	2001-2004	291	3	187
GC-HBOC	Germany	1947-2014	3,378	256	1,593
GENICA	Germany	2000-2004	459	1	284
GEPARSIXTO	Germany	n.r.	386	-	-
GESBC	Germany	1992-1995	312	39	181
HABCS	Germany	1984-2010	909	19	863
MARIE	Germany	2001-2005	506	6	289
PREFACE	Germany	2001-2011	2,923	-	-
SKKDKFZS	Germany	1993-2005	1,086	9	-
SUCCESSB	Germany	2008-2011	440	-	-
SUCCESSC	Germany	2001-2011	2,836	-	-
CCGP	Greece	1983-2013	667	5	322
BCINIS	Israel	1999-2012	1,337	100	724
MBCSG	Italy	1977-2012	549	72	366
ABCS	Netherlands	2003-2011	347	-	189
ORIGO	Netherlands	1991-2005	921	113	-
RBCS	Netherlands	1975-2009	444	23	-
NBCS	Norway	1973-2011	1,163	38	-
PBCS	Poland	1998-2003	1,740	111	2,045
SZBCS	Poland	2010-2012	352	9	174
MABCS	Republic of North Macedonia	1993-2013	89	1	90
HUBCS	Russia	1977-2009	211	-	116
BREOGAN	Spain	1991-2019	1,535	129	910
HCSC	Spain	1975-2013	423	3	-
KARBAC	Sweden	1966-2013	499	3	-
KARMA	Sweden	1969-2017	2,839	339	6,983
MISS	Sweden	1983-2013	633	68	1,529
pKARMA	Sweden	1980-2015	748	86	48
SMC	Sweden	1987-2013	1,509	-	661
BBCS	UK	1985-2009	122	-	440

Study acronym [a]	Country	Diagnosis years	Invasive cases (N)	In situ cases (N)	Controls (N)
DIETCOMPLYF	UK	2004-2007	708	3	-
FHRISK	UK	1987-2015	146	31	644
POSH	UK	2000-2007	1,088	-	-
PROCAS	UK	1988-2018	380	93	1,648
SBCS	UK	2012-2015	126	2	-
SEARCH	UK	2003-2012	4,057	-	2,653
UKBGS	UK	1985-2014	1,048	584	705
UKOPS	UK	n.a.	-	-	974
2SISTER	USA	n.r.	919	151	-
AHS	USA	1994-2013	513	1	1,137
BCFR-NY	USA	1949-2011	401	53	27
BCFR-PA	USA	1969-2011	67	6	-
BCFR-UTAH	USA	1952-2009	100	1	-
CPSII	USA	1992-2009	2,393	598	3,028
CTS	USA	1998-2010	1,156	-	610
MCBCS	USA	1998-2014	749	167	212
MEC	USA	1972-2012	668	5	724
MMHS	USA	2003-2013	275	99	1,635
MSKCC	USA	1982-2012	136	2	-
NBHS	USA	2001-2009	483	112	652
NC-BCFR	USA	1967-2012	759	15	150
NCBCS	USA	1993-2012	2,074	315	1,006
NHS	USA	1976-2012	1,103	333	1,804
NHS2	USA	1989-2011	1,112	409	1,905
PLCO	USA	1994-2013	1,822	483	2,595
SISTER	USA	2003-2008	1,504	498	1,556
TNBCC	USA	2003-2013	113	-	-
UBCS	USA	1960-2015	606	60	-
UCIBCS	USA	1994-2003	427	74	258
USRT	USA	1945-2005	1,354	338	1,699
Total		**1945-2019**	**69,838**	**6,667**	**54,452**

n.a., not applicable; n.r., not recorded

a See Supplementary Table S1 (Online Resource) for study names and references.

Table 2. Association between the primary instrumental genetic variables for overall physical activity (per standard deviation) and risk of breast cancer

		Full instrument (five SNPs)		Excluding one pleiotropic SNP for outcomes with detected pleiotropy[a]	
Type of breast cancer	**N cases (vs. 54,452 controls)**	**Odds ratios (95% CI) [b]**	***P* for heterogeneity [c]**	**Odds ratios (95% CI) [b]**	***P* for heterogeneity [c]**
Invasive cancers					
All invasive	69,838	0.48 (0.30-0.78)	0.016	0.59 (0.42-0.83)	0.312
Pre/perimenopausal	[d] 23,999	0.51 (0.31-0.83)	0.419	--	
Postmenopausal	[e] 45,839	0.48 (0.28-0.80)	0.054	--	
By receptor status					
ER+	46,528	0.45 (0.25-0.83)	0.004	0.60 (0.43-0.85)	0.459
ER-	11,246	0.79 (0.37-1.66)	0.069	--	
PR+	34,891	0.43 (0.22-0.85)	0.003	0.58 (0.37-0.91)	0.223
PR-	16,432	0.65 (0.38-1.13)	0.186	--	
HER2+	6,945	0.48 (0.26-0.89)	0.479	--	
HER2-	33,214	0.58 (0.35-0.98)	0.060	--	
Combined hormone receptor- and/or HER2-defined subtypes					
ER+ or PR+; HER2+	4,816	0.42 (0.20-0.88)	0.478	--	
ER+ or PR+; HER2-	27,874	0.57 (0.28-1.18)	0.004	0.79 (0.49-1.26)	0.254
ER-; PR-; HER2+	1,974	0.53 (0.18-1.57)	0.700	--	
ER-; PR-; HER2-	4,964	0.60 (0.17-2.12)	0.015	0.95 (0.37-2.44)	0.224
ER- and PR- (all)	9,215	0.65 (0.27-1.56)	0.036	0.46 (0.22-0.96)	0.226
By morphology					
Ductal	42,223	0.52 (0.32-0.84)	0.053	--	
Lobular	8,795	0.32 (0.18-0.58)	0.500	--	
By stage at diagnosis					
Stage I	17,583	0.51 (0.32-0.82)	0.333	--	
Stage II	15,992	0.36 (0.22-0.58)	0.576	--	
Stage III/IV	4,553	0.37 (0.17-0.81)	0.499	--	
By tumour grade					
Grade 1/2	34,647	0.43 (0.23-0.81)	0.011	0.58 (0.39-0.85)	0.514
Grade 3	16,432	0.46 (0.30-0.72)	0.552	--	
In situ cancers					
All in situ	6,667	0.63 (0.34-1.18)	0.390	--	
Ductal carcinoma in situ	3,510	[f] 0.92 (0.25-3.43)	0.039	--	

Abbreviations: CI, confidence interval; ER+/-, oestrogen receptor positive/negative; GWAS, genome wide association study; HER2+/-, human epidermal growth factor receptor 2 positive/negative; PR+/-, progesterone receptor positive/negative; SNP, single nucleotide polymorphism.

a Outlying SNP rs564819152 was excluded from analyses of all invasive, ER+, PR+, HR+/HER2-, and well/moderately differentiated cancers (outlier identified by MR-PRESSO, global-pleiotropy test $p<0.05$), and HR- cancers (outlier suggested by scatter plots and leave-one-out analyses; MR-PRESSO global-pleiotropy test $p=0.053$). Outlying SNP rs6775319 was excluded from analyses of triple negative cancers (ER-/PR-/HER2-), and was identified by MR-PRESSO.

b Causal odds ratios were estimated by inverse-variance weighted Mendelian randomization, using SNPs identified in a GWAS of accelerometer-measured movement traits by Doherty et al (9)

c p-value associated with the heterogeneity test statistic (Cochran's Q statistic) measuring heterogeneity of causal effects between SNPs

d vs pre/perimenopausal controls (n=17,686), assigned using age (<50 years) if menopause status was unknown

e vs postmenopausal controls (n=36,766), assigned using age (≥50 years) if menopause status was unknown

f For analyses of ductal carcinoma in situ, likely pleiotropy was indicated by the Cochran's Q statistic (p_{het}=0.04) and the MR-Egger intercept test for horizontal pleiotropy ($p_{intercept}$=0.01). However, a clear outlying SNP could not be identified, although leave-one-out analyses suggested substantial variation in results by instrument composition.

-- No outlying SNPs were identified.

Table 3. Association between instrumental genetic variables for vigorous physical activity, assessed in two ways, and risk of breast cancer

		Accelerometer-measured activity over 425 milligravities, per fraction of time, using one SNP [a]	Self-reported vigorous physical activity (≥ 3 vs. 0 days/week)			
			Full instrument (five SNPs)		Excluding one pleiotropic SNP for outcome with detected pleiotropy [b]	
Type of breast cancer	N cases (vs. 54,452 controls)	Odds ratios (95% CI) [c]	Odds ratios (95% CI) [c]	*P* for heterogeneity [d]	Odds ratios (95% CI) [c]	*P* for heterogeneity [d]
Invasive cancers						
All invasive	69,838	0.63 (0.32-1.22)	0.83 (0.69-1.01)	0.650	--	
Pre/perimenopausal	[e] 23,999	0.80 (0.25-2.58)	0.62 (0.45-0.87)	0.788	--	
Postmenopausal	[f] 45,839	0.53 (0.24-1.21)	0.95 (0.75-1.19)	0.630	--	
By receptor status						
ER+	46,528	0.74 (0.35-1.55)	0.86 (0.70-1.07)	0.917	--	
ER-	11,246	0.58 (0.17-1.94)	0.86 (0.61-1.21)	0.418	--	
PR+	34,891	0.68 (0.30-1.54)	0.77 (0.61-0.98)	0.544	--	
PR-	16,432	0.56 (0.19-1.59)	0.95 (0.70-1.28)	0.948	--	
HER2+	6,945	0.31 (0.07-1.39)	0.83 (0.53-1.31)	0.327	--	
HER2-	33,214	1.01 (0.44-2.31)	0.86 (0.68-1.10)	0.550	--	
Combined hormone receptor- and/or HER2-defined subtypes						
ER+ or PR+; HER2+	4,816	0.41 (0.07-2.35)	1.00 (0.58-1.70)	0.321	--	
ER+ or PR+; HER2-	27,874	0.84 (0.35-2.02)	0.82 (0.64-1.06)	0.560	--	
ER-; PR-; HER2+	1,974	0.21 (0.02-2.87)	0.57 (0.27-1.20)	0.727	--	
ER-; PR-; HER2-	4,964	2.16 (0.39-12.1)	1.30 (0.79-2.12)	0.593	--	
ER- and PR- (all)	9,215	0.78 (0.21-2.91)	0.95 (0.66-1.39)	0.559	--	
By morphology						
Ductal	42,223	0.62 (0.29-1.32)	0.81 (0.65-1.00)	0.932	--	
Lobular	8,795	0.60 (0.15-2.45)	0.78 (0.53-1.17)	0.809	--	
By stage at diagnosis						
Stage I	17,583	0.47 (0.16-1.36)	0.88 (0.65-1.19)	0.598	--	
Stage II	15,992	0.66 (0.21-2.07)	0.82 (0.59-1.14)	0.788	--	
Stage III/IV	4,553	0.41 (0.06-2.63)	0.75 (0.44-1.27)	0.910	--	

		Accelerometer-measured activity over 425 milligravities, per fraction of time, using one SNP [a]	Self-reported vigorous physical activity (≥ 3 vs. 0 days/week)			
			Full instrument (five SNPs)		Excluding one pleiotropic SNP for outcome with detected pleiotropy [b]	
Type of breast cancer	**N cases (vs. 54,452 controls)**	**Odds ratios (95% CI) [c]**	**Odds ratios (95% CI) [c]**	***P* for heterogeneity [d]**	**Odds ratios (95% CI) [c]**	***P* for heterogeneity [d]**
By tumour grade						
Grade 1/2	34,647	0.54 (0.24-1.22)	0.84 (0.66-1.06)	0.640	--	
Grade 3	16,432	0.51 (0.18-1.46)	0.99 (0.73-1.33)	0.557	--	
In situ cancers						
All in situ	6,667	0.47 (0.11-2.09)	0.94 (0.43-2.08)	0.007	1.30 (0.72-2.34)	0.189
Ductal carcinoma in situ	3,510	0.65 (0.09-4.72)	0.85 (0.42-1.69)	0.204	--	

Abbreviations: CI, confidence interval; ER+/-, oestrogen receptor positive/negative; GWAS, genome wide association study; HER2+/-, human epidermal growth factor receptor 2 positive/negative; PR+/-, progesterone receptor positive/negative; SNP, single nucleotide polymorphism.

a This SNP is a missense mutation in the gene *PML*, which plays a role in tumour suppression and is associated with height. *PML* is not expressed in breast tissue, but highly expressed in adipose tissue, suggesting that inverse (protective) associations observed do not derive from direct oncosuppression.

b Excluding one outlying SNP identified by MR-PRESSO: rs2764261 (for analyses modelling the association with in situ tumours)

c Causal odds ratios were estimated by inverse-variance weighted Mendelian randomization, using SNPs identified in a GWAS of physical activity by Klimentidis et al (10)

d *p*-value associated with the heterogeneity test statistic (Cochran's Q statistic) measuring heterogeneity of causal effects between SNPs

e vs pre/perimenopausal controls (n=17,686), assigned using age (<50 years) if menopause status was unknown

f vs postmenopausal controls (n=36,766), assigned using age (≥50 years) if menopause status was unknown

-- No outlying SNPs were identified by MR-PRESSO.

Table 4. Association between instrumental genetic variables for sedentary time (per standard deviation in percent time spent sedentary) and risk of breast cancer

Type of breast cancer	N cases (vs. 54,452 controls)	Odds ratios (95% CI) [a]	*P* for heterogeneity [b]
Invasive cancers			
All invasive	69,838	1.20 (0.93-1.55)	0.962
Pre/perimenopausal	[c] 23,999	1.22 (0.78-1.90)	0.589
Postmenopausal	[d] 45,839	1.21 (0.89-1.65)	0.983
By receptor status			
ER+	46,528	1.19 (0.90-1.57)	0.992
ER-	11,246	1.43 (0.90-2.26)	0.926
PR+	34,891	1.19 (0.87-1.63)	0.386
PR-	16,432	1.40 (0.94-2.09)	0.435
HER2+	6,945	1.17 (0.67-2.06)	0.718
HER2-	33,214	1.27 (0.93-1.74)	0.955
Combined hormone receptor- and/or HER2-defined subtypes			
ER+ or PR+; HER2+	4,816	0.86 (0.44-1.67)	0.585
ER+ or PR+; HER2-	27,874	1.12 (0.80-1.56)	0.801
ER-; PR-; HER2+	1,974	1.94 (0.71-5.25)	0.646
ER-; PR-; HER2-	4,964	2.04 (1.06-3.93)	0.500
ER- and PR- (all)	9,215	1.77 (1.07-2.92)	0.819
By morphology			
Ductal	42,223	1.21 (0.91-1.62)	0.992
Lobular	8,795	1.12 (0.66-1.91)	0.695
By stage at diagnosis			
Stage I	17,583	1.62 (0.99-2.65)	0.187
Stage II	15,992	1.23 (0.79-1.90)	0.820
Stage III/IV	4,553	0.91 (0.45-1.84)	0.640
By tumour grade			
Grade 1/2	34,647	1.15 (0.84-1.57)	0.901
Grade 3	16,432	1.32 (0.88-1.97)	0.967
In situ cancers			
All in situ	6,667	1.75 (1.00-3.07)	0.933
Ductal carcinoma in situ	3,510	2.11 (0.99-4.49)	0.487

Abbreviations: CI, confidence interval; ER+/-, oestrogen receptor positive/negative; GWAS, genome wide association study; HER2+/-, human epidermal growth factor receptor 2 positive/negative; PR+/-, progesterone receptor positive/negative; SNP, single nucleotide polymorphism.

a Causal odds ratios were estimated by inverse-variance weighted Mendelian randomization, using six SNPs identified in a GWAS of accelerometer-measured movement traits by Doherty et al (9)

b *p*-value associated with the heterogeneity test statistic (Cochran's Q statistic) measuring heterogeneity of causal effects between SNPs

c vs pre/perimenopausal controls (n=17,686), assigned using age (<50 years) if menopause status was unknown

d vs postmenopausal controls (n=36,766), assigned using age (≥50 years) if menopause status was unknown

References

1. International Agency for Research on Cancer. Weight Control and Physical Activity. Lyon; 2002.
2. Physical Activity Guidelines Advisory Committee. Physical Activity Guidelines Advisory Committee Report, 2008. Washington, DC: U.S. Department of Health and Human Services; 2008.
3. World Cancer Research Fund International / American Institute for Cancer Research. Continuous Update Project Report: Diet, Nutrition, Physical Activity and Breast Cancer. 2017.
4. Lynch BM, Mahmood S , T B. Chapter 10: Sedentary Behaviour and Cancer. In: Leitzmann M, Jochem C, Schmid D, editors. Sedentary Behaviour Epidemiology. Springer Series on Epidemiology and Public Health. Cham, Switzerland: Springer International Publishing; 2018. p. 245-98.
5. Chong F, Wang Y, Song M, Sun Q, Xie W, Song C. Sedentary behavior and risk of breast cancer: a dose-response meta-analysis from prospective studies. Breast cancer (Tokyo, Japan). 2020.
6. Davey Smith G, Hemani G. Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. Human molecular genetics. 2014;23(R1):R89-R98.
7. Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol. 2000;29(4):722-9.
8. Papadimitriou N, Dimou N, Tsilidis KK, Banbury B, Martin RM, Lewis SJ, et al. Physical activity and risks of breast and colorectal cancer: A Mendelian randomisation analysis. Nature communications. 2020;11(1):597.
9. Doherty A, Smith-Byrne K, Ferreira T, Holmes MV, Holmes C, Pulit SL, et al. GWAS identifies 14 loci for device-measured physical activity and sleep duration. Nature communications. 2018;9(1):5257.
10. Klimentidis YC, Raichlen DA, Bea J, Garcia DO, Wineinger NE, Mandarino LJ, et al. Genome-wide association study of habitual physical activity in over 377,000 UK Biobank participants identifies multiple variants including CADM2 and APOE. Int J Obes. 2018;42(6):1161-76.
11. Choi KW, Chen CY, Stein MB, Klimentidis YC, Wang MJ, Koenen KC, et al. Assessment of bidirectional relationships between physical activity and depression among adults: A 2-sample Mendelian randomization study. JAMA psychiatry. 2019;76(4):399-408.
12. Doherty A, Jackson D, Hammerla N, Plötz T, Olivier P, Granat MH, et al. Large scale population assessment of physical activity using wrist worn accelerometers: The UK Biobank study. PLoS One. 2017;12(2):e0169649.
13. Willetts M, Hollowell S, Aslett L, Holmes C, Doherty A. Statistical machine learning of sleep and physical activity phenotypes from sensor data in 96,220 UK Biobank participants. Scientific reports. 2018;8(1):7961.
14. Hildebrand M, VT VANH, Hansen BH, Ekelund U. Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. Medicine and science in sports and exercise. 2014;46(9):1816-24.
15. Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA, et al. The OncoArray consortium: A network for understanding the genetic architecture of common cancers. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2017;26(1):126-35.
16. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature. 2011;478(7367):103-9.
17. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genetic epidemiology. 2013;37(7):658-65.
18. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. Int J Epidemiol. 2017:1-6.

19. Burgess S, Davey Smith G, Davies N, Dudbridge F, Gill D, Glymour MM, et al. Guidelines for performing Mendelian randomization investigations [version 2]. Wellcome Open Res 2020;4:186.
20. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genetic epidemiology. 2016;40(4):304-14.
21. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44(2):512-25.
22. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nature genetics. 2018;50(5):693-8.
23. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic acids research. 2019;47(D1):D1005-D12.
24. Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, et al. PhenoScanner: A database of human genotype-phenotype associations. Bioinformatics (Oxford, England). 2016;32(20):3207-9.
25. Kamat MA, Blackshaw JA, Young R, Surendran P, Burgess S, Danesh J, et al. PhenoScanner V2: An expanded tool for searching human genotype-phenotype associations. Bioinformatics (Oxford, England). 2019;35(22):4851-3.
26. Brion M-JA, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. International journal of epidemiology. 2012;42(5):1497-501.
27. Neale B. UK Biobank GWAS round 2 2018 [Available from: http://www.nealelab.is/uk-biobank/.
28. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. Nature. 2016;533(7604):539-42.
29. Matthews CE, Keadle SK, Berrigan D, Staudenmayer J, P FS-M, Troiano RP, et al. Influence of accelerometer calibration approach on moderate-vigorous physical activity estimates for adults. Medicine and science in sports and exercise. 2018;50(11):2285-91.
30. Greenland S. Noncollapsibility, confounding, and sparse-data bias. Part 1: The oddities of odds. Journal of clinical epidemiology. 2021;138:178-81.
31. Greenland S. Noncollapsibility, confounding, and sparse-data bias. Part 2: What should researchers make of persistent controversies about the odds ratio? Journal of clinical epidemiology. 2021;139:264-8.
32. Neilson HK, Farris MS, Stone CR, Vaska MM, Brenner DR, Friedenreich CM. Moderate-vigorous recreational physical activity and breast cancer risk, stratified by menopause status: A systematic review and meta-analysis. Menopause (New York, NY). 2017;24(3):322-44.
33. McTiernan A, Friedenreich CM, Katzmarzyk PT, Powell KE, Macko R, Buchner D, et al. Physical activity in cancer prevention and survival: A systematic review. Medicine and science in sports and exercise. 2019;51(6):1252-61.
34. Munafò MR, Davey Smith G. Robust research needs many lines of evidence. Nature. 2018;553(7689):399-401.
35. Lynch BM, Neilson HK, Friedenreich CM. Physical activity and breast cancer prevention. In: Courneya KS, Friedenreich CM, editors. Recent Results in Cancer Research. Physical Activity and Cancer. Berlin: Springer-Verlag; 2011.
36. Lynch B, Leitzmann M. An evaluation of the evidence relating to physical inactivity, sedentary behavior, and cancer incidence and mortality. Curr Epidemiol Rep. 2017;4(3):221-31.
37. Neilson HK, Conroy SM, Friedenreich CM. The influence of energetic factors on biomarkers of postmenopausal breast cancer risk. Current nutrition reports. 2014;3:22-34.

38. Dempsey PC, Matthews CE, Dashti SG, Doherty AR, Bergouignan A, van Roekel EH, et al. Sedentary behavior and chronic disease: Mechanisms and future directions. Journal of physical activity & health. 2020;17(1):52-61.
39. Lynch BM. Sedentary behavior and cancer: A systematic review of the literature and proposed biological mechanisms. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2010;19(11):2691-709.