



Biomarkers and Disease Trajectories Influencing Women's Health: Results from the UK Biobank Cohort

Haomin Yang^{1,2} · Yudi Pawitan² · Fang Fang³ · Kamila Czene² · Weimin Ye^{1,2}

Received: 3 November 2021 / Revised: 13 April 2022 / Accepted: 15 April 2022 / Published online: 12 May 2022
© The Author(s) 2022

Abstract

Women's health is important for society. Despite the known biological and sex-related factors influencing the risk of diseases among women, the network of the full spectrum of diseases in women is underexplored. This study aimed to systematically examine the women-specific temporal pattern (trajectory) of the disease network, including the role of baseline physical examination indexes, and blood and urine biomarkers. In the UK Biobank study, 502,650 participants entered the cohort from 2006 to 2010, and were followed up until 2019 to identify disease incidence via linkage to the patient registers. For those diseases with increased risk among women, conditional logistic regression models were used to estimate odds ratios (ORs), and the binomial test of direction was further used to build disease trajectories. Among 301 diseases, 82 diseases in women had ORs > 1.2 and $p < 0.00017$ when compared to men, involving mainly diseases in the endocrine, skeletal and digestive systems. Diseases with the highest ORs included breast diseases, osteoporosis, hyperthyroidism, and deformity of the toes. The biomarker and disease trajectories suggested estradiol as a risk predictor for breast cancer, while a high percentage of reticulocyte, body mass index and waist circumference were associated with an increased risk of upper-limb neuropathy. In addition, the risk of cholelithiasis was increased in women diagnosed with dyspepsia and diaphragmatic hernia. In conclusion, women are at an increased risk of endocrine, skeletal and digestive diseases. The biomarker and disease trajectories in women suggested key pathways to a range of adverse outcomes downstream, which may shed light on promising targets for early detection and prevention of these diseases.

Keywords Women health · Disease trajectory · Biomarkers · Phenome wide association study

✉ Haomin Yang
Haomin.yang@ki.se

✉ Weimin Ye
weimin.ye@ki.se

Yudi Pawitan
Yudi.pawitan@ki.se

Fang Fang
fang.fang@ki.se

Kamila Czene
kamila.czene@ki.se

¹ Department of Epidemiology and Health Statistics, School of Public Health and Key Laboratory of Ministry of Education for Gastrointestinal Cancer, Fujian Medical University, Xue Yuan Road 1, University Town, Fuzhou 350122, China

² Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, 17177 Stockholm, Sweden

³ Institute of Environmental Medicine, Karolinska Institutet, 17177 Stockholm, Sweden

Introduction

Women's health is important for themselves, their families and the entire society. However, women experience unique healthcare challenges and are at an increased risk of several diseases, resulting in a lower quality of life (Arrospide et al. 2019). Previous studies on women's diseases separately focused on a limited number of diseases, such as cardiovascular disease (Ketepe-Arachi and Sharma 2017), osteoporosis (Nachtigall et al. 2013), and cancer (Hemminki et al. 2012). The total spectrum of disease risks among women has not been analyzed jointly compared to men.

Despite the known biological and sex-related factors influencing the risk of diseases among women (Pucci et al. 2017), the associations between different diseases among women are still under-explored (Westergaard et al. 2019). Since cost-effective and better health care for women depends on early detection and intervention for adverse outcomes, understanding the temporal pattern of the disease

network may help promote a life-course approach to women's health and identify key indicators to mitigate the risk of future poor outcomes.

While temporal patterns of disease networks have been tested in several studies (Giannoula et al. 2018; Jensen et al. 2014; Siggaard et al. 2020; Westergaard et al. 2019), applications of this approach for exploring clustered and sequential patterns of disease networks in women are scarce. Furthermore, preclinical indicators for the diseases have not been evaluated together with the disease trajectory. As common blood and urine tests are widely performed in clinical practice during disease screening, diagnostic and surveillance procedures, these blood and urine biomarkers may be informative in the disease trajectory and suggest an earlier intervention time window for disease prevention.

This study comprehensively investigated the risk of diseases among women, compared to men, using electronic health records from the UK Biobank. We further investigated temporal patterns of the disease network among women based on pre-disease biomarkers, to identify potential indicators for the early detection of diseases.

Materials and Methods

Study Population

The UK Biobank is a national health project aiming to improve the prevention, diagnosis and treatment of various diseases in the United Kingdom (Hewitt et al. 2016). From 2006 to 2010, 502,650 participants were enrolled in the study across 22 assessment centers in the UK aged 40–69 years. The baseline assessment for the participants included a computer-assisted questionnaire, physical examination, and biological sample (including blood and urine) collection. Disease diagnoses were obtained from the primary diagnoses recorded in the Hospital Episode Statistics database, the Scottish Morbidity Record and the Patient Episode Database in England, Scotland and Wales, respectively, available for all participants since 1997 (Sudlow et al. 2015). Diagnoses are coded according to the International Classification of Diseases (ICD) codes. The ICD-10 system has a clear hierarchical structure for disease classification. All diseases in this study were defined based on the 3-digit ICD-10 codes (A00 to N99 and S00 to T99). Disease ascertainment ended on 31 December 2019, considering the influence of the COVID-19 pandemic.

To assess the association between female sex and disease incidence, we performed a series of nested case–control studies based on the UK Biobank data, using incidence density sampling. We first selected all the incident cases of a specific disease in the UK Biobank beyond three months after the baseline assessment and randomly sampled up to

10 participants for each case from the UK Biobank as the controls, as suggested previously (Richardson 2004) (Supplementary Fig. 1). Individual matching was performed based on age, examination center and year of participation. The controls were alive and free of the studied diseases (including diseases in the same sub-category as the studied diseases) at the date of diagnosis of the matched case (the index date). The full list of disease codes and their sub-categories can be found in Supplementary Material 1. Information on death was obtained by cross-linkage with the National Health Service (NHS) Digital or the NHS Central Register (Sudlow et al. 2015). As this study focused on incident diseases after the baseline assessment, individuals with a previous disease diagnosis before their participation were excluded when analyzing the incidence of that specific disease.

Statistical Analysis

Disease Risk in Women

For each disease, we performed a conditional logistic regression model, treating female sex as the exposure and disease code as the outcome, conditioning on examination center, year of participation and year of birth. In this part of the study, we did not analyze diseases in the genital tract. As suggested by previous studies (Li et al. 2018), we only analyzed ICD codes with more than 200 incident disease cases; this resulted in 301 nested case–control studies. The odds ratio (OR) derived in a nested case–control study is theoretically largely representative of the relative risk derived from a cohort study (Goldstein and Langholz 1992). To avoid a very weak association and ensure a significant difference in relative risk between men and women (Jensen et al. 2017; Westergaard et al. 2019), diseases with ORs larger than 1.2 and p value less than 0.05/301 (the Bonferroni corrected threshold) were considered as diseases associated with the female sex.

Disease Trajectory Analysis

Disease trajectory analysis was proposed to assess the strength and directionality of the association between two diseases (Jensen et al. 2014; Siggaard et al. 2020). In the first step, the association between two incident diseases (denoted by D1 and D2) was tested among women after they participated in the UK Biobank. Women with either D1 or D2 before three months after the date of participation were excluded from the analysis. As we aimed to investigate the female-specific disease trajectories, D1 and D2 should be identified as associated with the female sex in the above analysis or in the female reproductive system. We first constructed the disease pairs (D1–D2) among women. As suggested by previous studies (Han et al. 2021; Siggaard et al.

2020; Yang et al. 2019), only disease pairs with more than 50 women with these two diseases beyond three months after cohort participation were included in the analysis to ensure statistical power. The association between the two diseases was estimated using a case–control study design. We defined women with D2 as cases and randomly matched them with up to 10 participants (controls). The controls were alive and free from D2 at the time when D2 occurred in their matched cases. Cases and controls were individually matched by age at participation, year of birth, and examination center. We then tested whether D2 was associated with D1, using conditional logistic regression models, and calculated the OR to estimate the association between the disease pairs.

In the second step, for each pair of diseases identified in step 1, we used binomial tests to assess the temporal directionality ($D1 \rightarrow D2$) of the association, as recommended by Jensen A. et al. (Jensen et al. 2014). The binomial test for $D1 \rightarrow D2$ tested whether the probability of a patient with D1 diagnosed before D2 was significantly higher than 50% among patients diagnosed with both D1 and D2. $D1 \rightarrow D2$ disease pairs with p values less than the Bonferroni-corrected threshold were then included in the next step of the analysis.

Disease pairs meeting the strength ($OR > 1.2$) and directionality requirement of the association test were combined into disease trajectories. To detect the subgroups of the identified disease trajectories, we further used the Louvain clustering algorithm to subdivide the network of disease trajectories into clusters. The Louvain algorithm was previously used for community identification in social network analysis, which identifies the areas of the neighbor graph to be more densely connected than the overall connectivity (Iliho and Saritha 2019).

Biomarkers and Disease Trajectories

To investigate the associations between preclinical biomarkers and the longitudinal patterns of multi-step disease trajectories, disease pairs meeting the directionality requirement of the association test were first combined into 3-line disease trajectories (e.g. $D1 \rightarrow D2$ and $D2 \rightarrow D3$ were combined into $D1 \rightarrow D2 \rightarrow D3$) with at least 10 patients passing through this trajectory to avoid chance findings. Mediation analyses were further performed to test the potential causal relationship for the 3-line trajectories treating D1 as exposure, D3 as outcome, and D2 as mediator. We used the method suggested by VanderWeele (VanderWeele 2014), which estimated the overall effect of D1 on D3, in the presence of D2, and was decomposed into direct effect, only mediation, only interaction and both mediation and interaction effects. We also estimated the percent of the total association (on the log-odds scale) between D1 and D3 that was mediated by D2. The 65 blood and urinary biomarkers and four physical

examination indexes (body mass index (BMI), waist circumference, blood pressure, and heart rate) (B0) tested at the baseline in the UK Biobank (Supplementary Material 1) were analyzed to reflect the potential preclinical status of the women. As indicated by the UK Biobank, the selection of the biomarkers was based on their scientific relevance for studying a wide range of diseases.

To construct the biomarker and disease trajectory ($B0 \rightarrow D1 \rightarrow D2 \rightarrow D3$) among women, the association between D1 and the biomarker was tested by considering D1 as the outcome and the biomarker (B0) from the blood or urine sample collected at the baseline as the exposure. The incidence density sampling procedure and matching criteria were the same as in the disease risk analysis. The majority of the biomarkers were standardized with a mean of zero and standard deviation of one, except for estradiol, rheumatoid factor in blood and microalbumin in urine, which were dichotomized according to their limit of detection (as more than 50% of the participants were under the limit of detection). Their associations with D1 were tested using conditional logistic regression models. All biomarkers and disease trajectories ($B0 \rightarrow D1 \rightarrow D2 \rightarrow D3$) were mapped out using Cytoscape 3.5 (Shannon et al. 2003).

A flow chart of the methodology used in the analyses is illustrated in Supplementary Fig. 1. Sensitivity analyses included starting the follow-up 1 year after participating in the cohort and separating the analyses by age < 55 and age > 55 .

Statistical analyses were performed using SAS (version 9.4; SAS Institute Inc, Cary, NC, USA), and R software (version 3.6.3; R Foundation for Statistical Computing, Vienna, Austria).

Results

Risk of Diseases in Women Compared with Men

In the community-based cohort of the UK Biobank, 502,650 individuals (including 273,375 women) were included in the analysis, among whom 301 diseases were identified with more than 200 incident cases during the follow-up until the end of 2019. In the diseases analyzed, the risk of 82 diseases was higher among women than men with $OR > 1.2$, and $p < 0.05/301 (= 0.00017)$ (Supplementary Table 1, Fig. 1). Diseases with the highest ORs included breast diseases, osteoporosis, hyperthyroidism, and deformity of the toes, all of which are closely related to the hormones or the lifestyle behavior of women (Fig. 1). An increased risk of anxiety was also found in women. Sensitivity analyses focusing on diseases diagnosed beyond one year after participating in the UK Biobank cohort rendered similar results as those obtained from the main analysis (Supplementary Table 1,

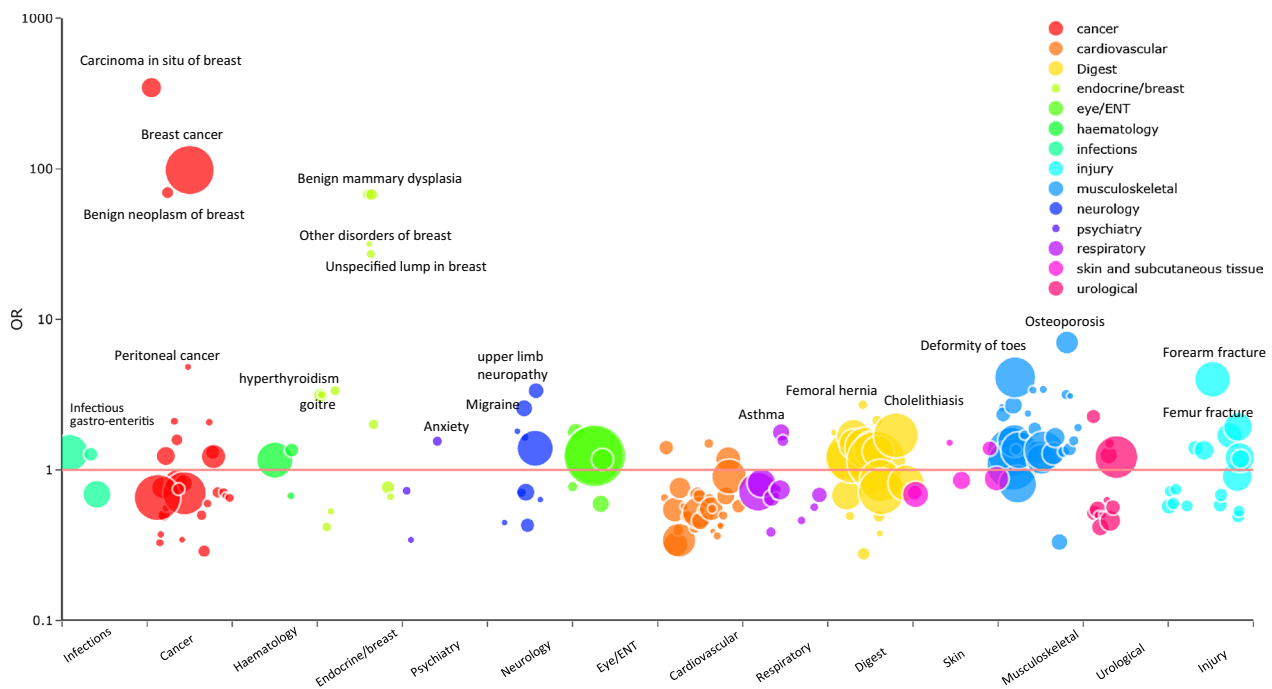


Fig. 1 Significant odds ratios (ORs) of diseases among women, compared to men in the UK Biobank. All risk increases were statistically significant after considering the issue of multiple testing ($p < 0.00017$). The Y-axis shows the odds ratio (on the log scale) of

the disease in women, compared to men. The X-axis shows the disease categories according to ICD codes. Details of the number of cases, odds ratios, and confidence intervals are listed in Supplementary Table 1

only one of the 82 diseases did not pass the threshold). Stratified analysis by age at the participation of the cohort showed that 83% (68 of 82) of the identified associations were consistent in young and old participants. Associations with aging-related diseases, such as cataracts, deformities of limbs and spinal and hip fractures were only found in old participants, while associations with melanocytic nevi and urinary disorders were only found in young participants.

Disease Trajectory Network in Women

Among the 619 disease pairs tested in women from UK Biobank, 77 pairs were associated with each other and passed the directionality test (p value $< 0.05/619 = 0.00009$, Supplementary Table 2). The strongest temporal associations were found for rheumatoid arthritis (M06 \rightarrow M05, OR = 221.46, 95%CI = 90.40–542.52), metastasis (C77 \rightarrow C78, OR = 29.78, 95%CI = 25.15–35.26) and back pain (M54 \rightarrow M47, OR = 27.31, 95%CI = 19.72–37.83).

The overall disease trajectory network is mapped in Supplementary Fig. 2, with the majority of the trajectories (more than 90%) found in both young and old women. Network clustering analysis divided the overall disease trajectory network into nine groups of diseases (Fig. 2). Among the subgroups of the disease trajectory network, the largest group included bone fracture and joint disorders, showing

a trajectory from bone fracture/joint disorders to complications of internal orthopedic implants (Fig. 2a). The second subgroup consisted of digestive diseases from dyspepsia to gastroenteritis and colitis (Fig. 2b), while the third subgroup suggested the development of breast cancer from carcinoma in situ to metastasis (Fig. 2c). Other subgroups included uterine diseases, urethrovaginal disease, degenerative disease, back pain, ovarian diseases, and rheumatoid arthritis.

Biomarkers and Disease Trajectories in Women

Among the 77 disease pairs, four 3-line trajectories (D1 \rightarrow D2 \rightarrow D3) were identified with more than 10 women passing through the trajectory. In the mediation analyses, the trajectory “G56 \rightarrow H25 \rightarrow H26” did not pass the mediation test, and was therefore not considered as a causal relationship (Supplementary Table 3). The three distinct diseases denoted by D1 (C50 breast cancer, G56 upper-limb neuropathy, and K30 dyspepsia) in the 3-line trajectories were further tested for their associations with 65 blood and urine biomarkers and four physical examination indexes at the baseline (Supplementary Table 3, Fig. 3). We further used the identified statistically significant B0 \rightarrow D1 pairs to extend the 3-line disease trajectories, resulting in the 4-line trajectories (B0 \rightarrow D1 \rightarrow D2 \rightarrow D3, shown in Fig. 4).

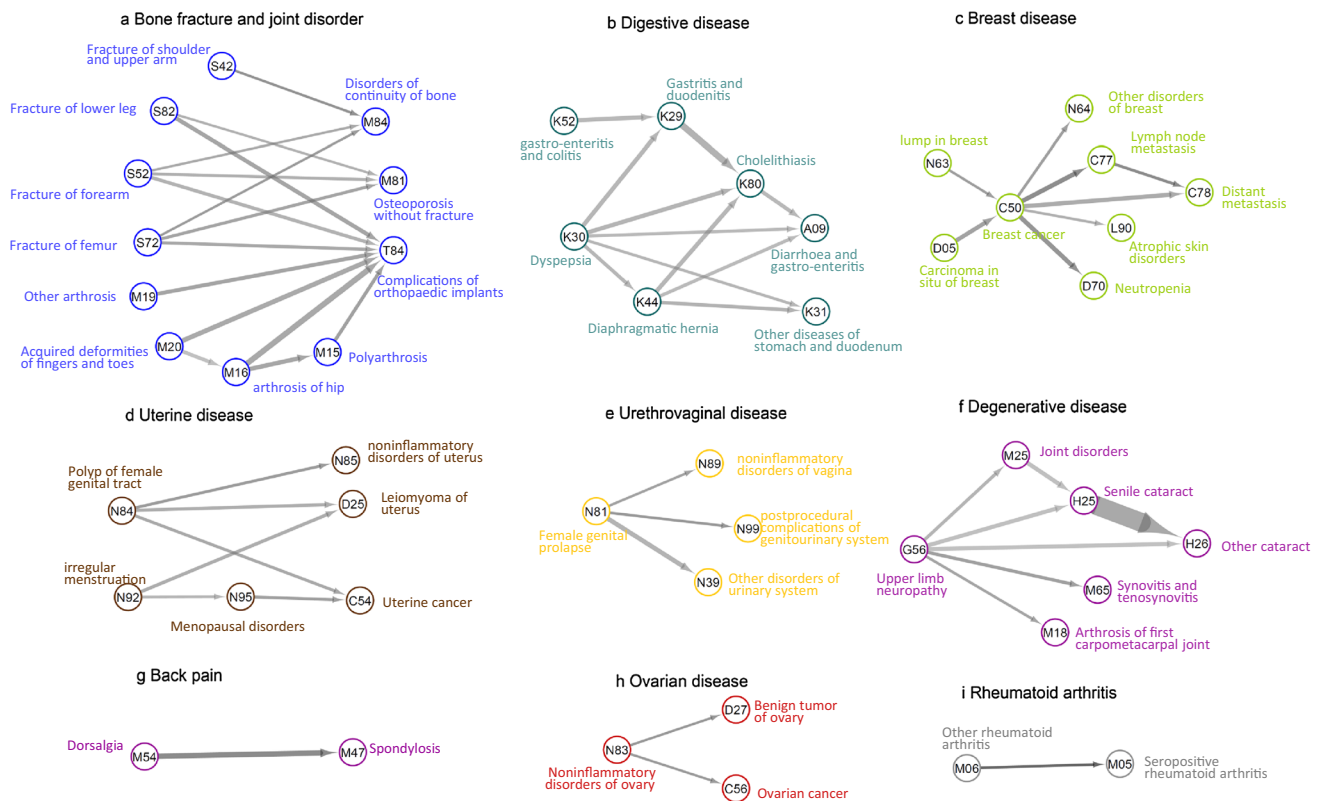


Fig. 2 Subgroups of disease trajectory networks among women in the UK Biobank. This figure illustrates the subgroups of the disease trajectory network in our analysis. The ICD-10 codes for the diseases are shown within the circle. The width of the line connecting two cir-

cles corresponds to the number of women with this disease pair. The color of the line indicates the odds ratio of the association between the two diseases

In the trajectories, a higher level of baseline estradiol (OR = 1.25, 95% CI = 1.12–1.39) was associated with a higher risk of breast cancer, which further revealed the progression of breast cancer from diagnosis to lymph-node and distant metastasis. Other biomarkers such as BMI, neutrophil count, IGF-1, testosterone and sex hormone-binding globulin (SHBG) were also associated with breast cancer risk, but the ORs were below 1.2 (Supplementary Table 4).

A higher level of reticulocyte percentage was associated with a higher risk of upper-limb neuropathy (OR = 1.26, 95%CI = 1.21–1.32) in women. High BMI and waist circumference were also associated with an increased risk of upper-limb neuropathy (OR_{BMI} = 1.39, 95%CI = 1.36–1.43; OR_{waist circumference} = 1.37, 95%CI = 1.33–1.40). In addition, triglycerides, cystatin C, urate, glucose and C-reactive protein were associated with an increased risk of upper-limb neuropathy with an OR close to 1.2. The diagnosis of upper-limb neuropathy was associated with future risk of arthrosis, which might result in complications in the internal orthopedic implants. In addition, the risk of cholelithiasis was increased in women diagnosed with dyspepsia and diaphragmatic hernia.

Sensitivity analyses by starting the follow-ups one year after participation and stratifying the analyses by age groups did not render results that differ greatly from the presented results, with more than 90% of the identified associations confirmed in both age groups (Supplementary Table 4).

Discussion

Using electronic health records in the UK Biobank, we found 82 diseases with increased risk in women as compared with men. Diseases with the largest ORs were those within the endocrine, digestive and skeletal systems. Subdivision of the female disease network revealed nine clusters of diseases, including mainly bone fracture and joint disorders, digestive diseases, and breast diseases. The biomarker and disease trajectory analysis in women showed estradiol as a risk predictor for breast cancer, while reticulocyte percentage, BMI and waist circumference were risk factors for upper-limb neuropathy and subsequent arthrosis.

In this study, we found an increased incidence of many endocrine, digestive and skeletal diseases, confirming previous knowledge (Lane 2006; Stinton et al. 2010; Wang

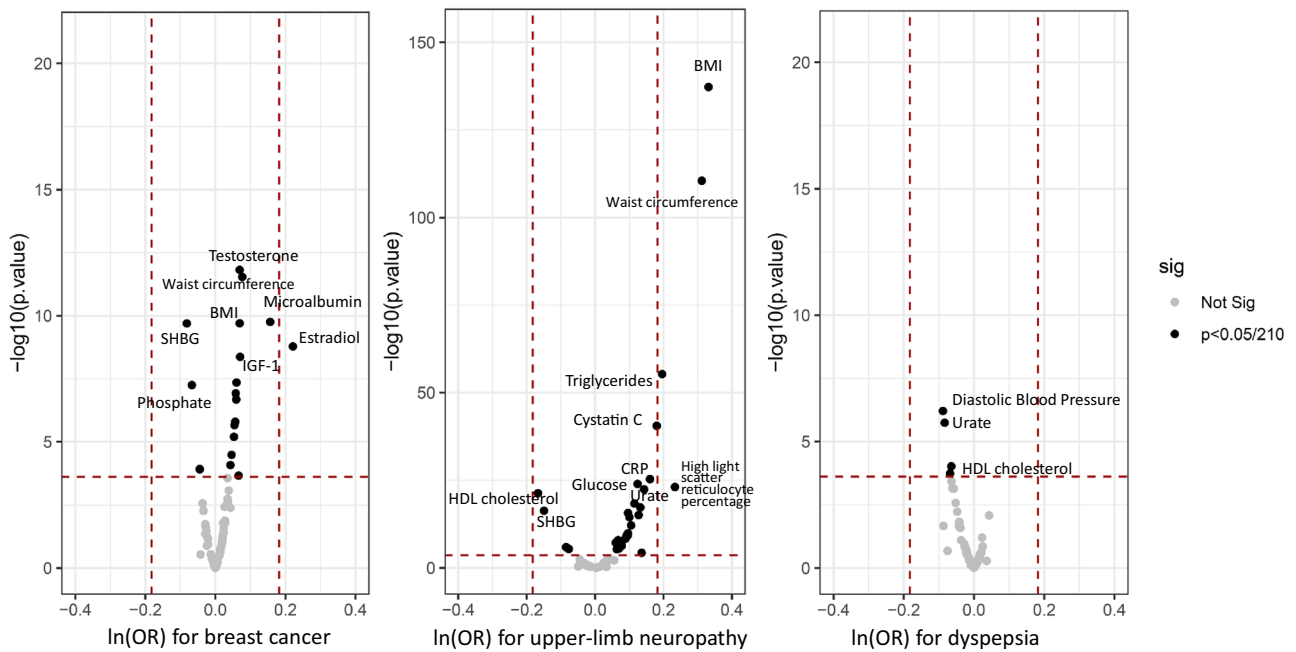
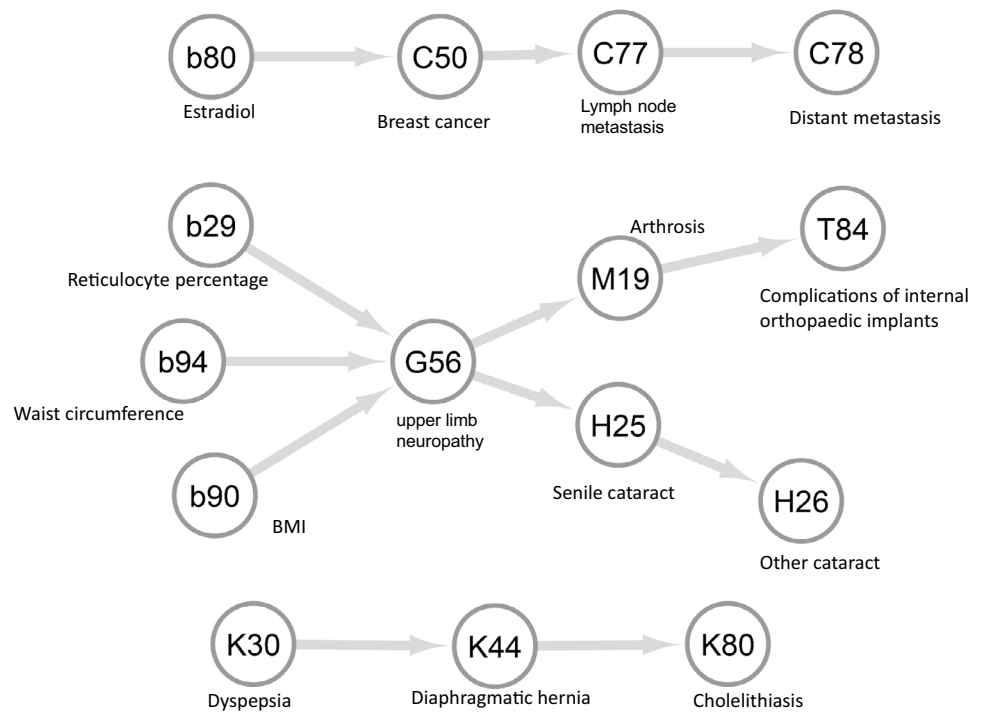


Fig. 3 The association between biomarkers and selected diseases in the UK Biobank. This figure shows the statistically significant associations between blood and urine biomarkers and the D1s of the 3-line disease trajectories (including upper-limb neuropathy, breast

cancer and dyspepsia). All the tested associations are listed in Supplementary Table 3. The horizontal dashed line indicates the p value $< 0.05/210$ ($= 0.00024$), while the vertical dashed line indicates an OR > 1.2 or < 0.8

Fig. 4 Biomarkers and disease trajectories among women in the UK Biobank. This figure shows the identified biomarkers and disease trajectories among women in the UK Biobank. For each pair of the trajectories, the codes in the circle are the ICD-10 codes for the diseases or the codes for biomarkers



and Crapo 1997). Many of the diseases with an increased risk in women, such as breast diseases, osteoporosis and fractures are closely related to the function of estrogen,

considering the expression of estrogen receptors (ER) in breast and bone tissues (Allred et al. 2004; Khalid and Krum 2016). Similarly, estrogen may directly affect human thyroid

cells through ER-dependent mechanisms, or indirectly by increasing thyroxine-binding globulin (Santin and Furlanetto 2011). Digestive diseases in the biliary system might also be influenced by estrogen as it increases the secretion of biliary cholesterol, causing cholesterol supersaturation of the bile, abdominal discomfort diagnosed as dyspepsia, and later-onset cholelithiasis (Wang et al. 2009). All these results indicate the necessity of focusing women's health on diseases associated with estrogen for health promotion.

In addition to the hormonal effects, women are more vulnerable to coping with stressful life events and have less social support (Burani and Nelson 2020; Mandelli et al. 2015), which might explain the increased risk of anxiety observed in women in our study.

Dyspepsia was associated with the risk of diaphragmatic hernia and subsequent cholelithiasis. One of the main causes of dyspepsia is gastroesophageal reflux disease (Blair and Beltz 2006), while diaphragmatic hernia could be a consequence of long-term severe gastroesophageal reflux disease (Gordon et al. 2004). However, although the results on the association between dyspepsia/reflux and cholelithiasis are conflicting (Avidan et al. 2001; Martínez de Pancorbo et al. 1997) in previous studies, a disease trajectory from dyspepsia to cholelithiasis has been confirmed by Danish data (Siggaard et al. 2020).

In this study, the risk of breast cancer was increased in women with high estradiol. Circulating estradiol is a known risk factor for breast cancer. The association between circulating estradiol and breast cancer was supported by previous studies in premenopausal as well as post-menopausal women (Endogenous Hormones and Breast Cancer Collaborative Group et al. 2013; Fuhrman et al. 2012). We further confirmed that high circulating testosterone and low SHBG levels were associated with an increased risk of breast cancer in our study, likely due to aromatase activity in fat tissues for testosterone (de Jong et al. 2001) and the binding effect of SHBG with estrogen (Arthur et al. 2021; Missmer et al. 2004). Additionally, a Mendelian randomization study also suggested SHBG as a causal factor for breast cancer (Dimou et al. 2019). These findings argue for a close surveillance of sex hormones among women at high risk for breast cancer.

In addition to sex hormones, we also confirmed several other biomarkers to be associated with breast cancer, such as waist circumference, BMI and IGF-1. The associations between waist circumference, BMI and breast cancer were probably caused by the aromatization of testosterone to estrogens in fat tissue (de Jong et al. 2001), while women with high BMI and waist circumference have abundant fat tissue cells. The association between IGF-1 and breast cancer was confirmed by cohort and Mendelian randomization studies (Murphy et al. 2020; Tan et al. 2021), likely via biological mechanisms through mitogenic and anti-apoptotic effects (Pollak 2008).

Higher BMI and waist circumference were associated with an increased risk of upper-limb neuropathy in our study, which is consistent with previous findings for different types of upper-limb neuropathy, such as carpal tunnel syndrome and ulnar neuropathies (Rhee et al. 2021; Warner et al. 1994). This association could be explained by the extra fat tissue inside the carpal tunnel which causes hydrostatic pressure or compression of the median nerve (Becker et al. 2002).

Interestingly, our analysis also showed a strong association between a high reticulocyte percentage and an increased risk of upper-limb neuropathy in women. Historically, testing for reticulocyte percentage was only considered for patients with suspected anemia (Piva et al. 2015). Our study suggests it as an indicator for subsequent disorders of anemia, such as neuropathy and arthrosis. These associations are biologically plausible, considering the deficiency of iron, folate acid, vitamin E and vitamin B12 as causes of anemia (World Health Organization 2017). Deficiency of these nutrients could probably influence the function of the myelin sheath that surrounds and protects nerves and result in neuropathy (Staff and Windebank 2014).

The main strengths of our study include the use of a community-based cohort with abundant information on biomarkers and clinical diagnoses from electronic health records. In addition, the novel trajectory analysis provides additional information to the existing literature by clarifying the clusters and the sequential pattern of diseases with increased risk in women. Furthermore, linking the biomarkers with the disease network can assist in the exploration of key pathophysiological pathways, from preclinical biomarkers to late adverse outcomes in women. These findings, if validated by other studies, may suggest further investigation into the biological linkage among different body systems and promote the overall health of women.

We also acknowledge several limitations. The present disease network and trajectory were constructed based on electronic health records. However, the validity of electronic health records in the UK has not been thoroughly evaluated, and we, therefore, used the main diagnosis to minimize the potential bias from misclassification. Notably, although we aimed to explore sequential associations between different diseases, the disease trajectory analysis itself did not provide a substantial basis for establishing a causal relationship. The observed disease patterns might be explained by the confounding factors between diseases, including genetics, lifestyle and other environmental factors. In addition, although 80% of the biomarkers had a missing rate of less than 10%, several biomarkers had missing values because of no data return or value beyond the reportable range. We assumed the missingness as random for no data return and used complete case analysis,

which might have attenuated the associations between biomarkers and diseases. Further studies are also needed to validate the identified biomarkers and disease trajectories in other settings.

Conclusion

This study describes a comprehensive picture of diseases with increased risks among women, including mainly endocrine, skeletal and digestive disorders. The biomarker and disease trajectories in women indicate key pathophysiological pathways from biomarkers to a range of late adverse outcomes. Further investigation into these key diseases or preclinical conditions might therefore be warranted to improve women's health.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s43657-022-00054-1>.

Authors' Contributions HY and WY were responsible for the study concept and design. HY and WY did the data and project management. HY did the data cleaning and analysis. HY, YP, KZ, WY and FF interpreted the data. HY and WY drafted the manuscript. All authors approved the final manuscript as submitted and agreed to be accountable for all aspects of the work.

Funding Open access funding is provided by Karolinska Institute. This work is supported by the Grant of Science and Technology of Fujian, China [2019L3006; 2020L3009]. HY is supported by the Natural Science Foundation of Fujian Province [grant no: 2021J01721], the Startup Fund for High-level Talents of Fujian Medical University [grant no: XRCZX2020007], Startup Fund for Scientific Research, Fujian Medical University [grant no: 2019QH1002], and Laboratory Construction Program of Fujian Medical University [grant no: 1100160208].

Data Availability Data from the UK Biobank (<http://www.ukbiobank.ac.uk/>) are available to all researchers upon making an application. Part of this research was conducted using the UK Biobank Resource under Application 61083.

Declarations

Conflicts of interest WY is the Editorial Board Member of *Phenomics*, and he was not involved in reviewing this paper.

Ethical Approval The UK Biobank has full ethical approval from the NHS National Research Ethics Service (reference number: 16/NW/0274).

Consent to Participate All the participants provided written informed consent.

Consent to Publish The participant's consent to have their data published.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing,

adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Allred DC, Brown P, Medina D (2004) The origins of estrogen receptor alpha-positive and estrogen receptor alpha-negative human breast cancer. *Breast Cancer Res* 6(6):240–245. <https://doi.org/10.1186/bcr938>
- Arrospe A, Machón M, Ramos-Goñi JM, Ibarrondo O, Mar J (2019) Inequalities in health-related quality of life according to age, gender, educational level, social class, body mass index and chronic diseases using the Spanish value set for Euroqol 5D–5L questionnaire. *Health Qual Life Outcomes* 17(1):69. <https://doi.org/10.1186/s12955-019-1134-9>
- Arthur RS, Dannenberg AJ, Rohan TE (2021) The association of prediagnostic circulating levels of cardiometabolic markers, testosterone and sex hormone-binding globulin with risk of breast cancer among normal weight postmenopausal women in the UK Biobank. *Int J Cancer*. <https://doi.org/10.1002/ijc.33508>
- Avidan B, Sonnenberg A, Schnell TG, Sontag SJ (2001) No association between gallstones and gastroesophageal reflux disease. *Am J Gastroenterol* 96(10):2858–2862. [https://doi.org/10.1016/S0002-9270\(01\)02800-3](https://doi.org/10.1016/S0002-9270(01)02800-3)
- Becker J, Nora DB, Gomes I, Stringari FF, Seitens R, Panosso JS, Ehlers JC (2002) An evaluation of gender, obesity, age and diabetes mellitus as risk factors for carpal tunnel syndrome. *Clin Neurophysiol* 113(9):1429–1434. [https://doi.org/10.1016/s1388-2457\(02\)00201-8](https://doi.org/10.1016/s1388-2457(02)00201-8)
- Blair KA, Beltz J (2006) Dyspepsia: is it gastroesophageal reflux disease peptic ulcer disease? *J Nurse Practitioners* 2(3):157–163. <https://doi.org/10.1016/j.nurpra.2006.01.017>
- Burani K, Nelson BD (2020) Gender differences in anxiety: The mediating role of sensitivity to unpredictable threat. *Int J Psychophysiol* 153:127–134. <https://doi.org/10.1016/j.ijpsycho.2020.05.001>
- de Jong PC, Blankenstein MA, van de Ven J, Nortier JW, Blijham GH, Thijssen JH (2001) Importance of local aromatase activity in hormone-dependent breast cancer: a review. *Breast* 10(2):91–99. <https://doi.org/10.1054/brst.2000.0209>
- Dimou NL, Papadimitriou N, Gill D, Christakoudi S, Murphy N, Gunter MJ, Travis RC, Key TJ, Fortner RT, Haycock PC, Lewis SJ, Muir K, Martin RM, Tsilidis KK (2019) Sex hormone binding globulin and risk of breast cancer: a Mendelian randomization study. *Int J Epidemiol* 48(3):807–816. <https://doi.org/10.1093/ije/dyz107>
- Endogenous Hormones and Breast Cancer Collaborative Group, Key TJ, Appleby PN, Reeves GK, Travis RC, Alberg AJ, Barricarte A, Berrino F, Krogh V, Sieri S, Brinton LA, Dorgan JF, Dossus L, Dowsett M, Eliassen AH, Fortner RT, Hankinson SE, Helzlsouer KJ, Hoff Man-Bolton J, Comstock GW, Kaaks R, Kahle LL, Muti P, Overvad K, Peeters PH, Riboli E, Rinaldi S, Rollison DE, Stanczyk FZ, Trichopoulos D, Tworoger SS, Vineis P (2013) Sex hormones and risk of breast cancer in premenopausal women: a

- collaborative reanalysis of individual participant data from seven prospective studies. *Lancet Oncol* 14(10):1009–1019. [https://doi.org/10.1016/S1470-2045\(13\)70301-2](https://doi.org/10.1016/S1470-2045(13)70301-2)
- Fuhrman BJ, Schairer C, Gail MH, Boyd-Morin J, Xu X, Sue LY, Buys SS, Isaacs C, Keefer LK, Veenstra TD, Berg CD, Hoover RN, Ziegler RG (2012) Estrogen metabolism and risk of breast cancer in postmenopausal women. *J Natl Cancer Inst* 104(4):326–339. <https://doi.org/10.1093/jnci/djr531>
- Giannoula A, Gutierrez-Sacristán A, Bravo Á, Sanz F, Furlong LI (2018) Identifying temporal patterns in patient disease trajectories using dynamic time warping: a population-based study. *Sci Rep* 8(1):4216. <https://doi.org/10.1038/s41598-018-22578-1>
- Goldstein L, Langholz B (1992) Asymptotic theory for nested case-control sampling in the Cox regression model. *Ann Stat* 20(4):1903–1928. <https://doi.org/10.1214/aos/1176348895>
- Gordon C, Kang JY, Neild PJ, Maxwell JD (2004) The role of the hiatus hernia in gastro-oesophageal reflux disease. *Aliment Pharm Ther* 20(7):719–732. <https://doi.org/10.1111/j.1365-2036.2004.02149.x>
- Han X, Hou C, Yang H, Chen W, Ying Z, Hu Y, Sun Y, Qu Y, Yang L, Valdimarsdóttir UA, Zhang W, Yang H, Fang F, Song H (2021) Disease trajectories and mortality among individuals diagnosed with depression: a community-based cohort study in UK Biobank. *Mol Psychiatr*. <https://doi.org/10.1038/s41380-021-01170-6>
- Hemminki K, Liu X, Ji J, Forsti A, Sundquist J, Sundquist K (2012) Effect of autoimmune diseases on risk and survival in female cancers. *Gynecol Oncol* 127(1):180–185. <https://doi.org/10.1016/j.ygyno.2012.07.100>
- Hewitt J, Walters M, Padmanabhan S, Dawson J (2016) Cohort profile of the UK Biobank: diagnosis and characteristics of cerebrovascular disease. *BMJ Open* 6(3):e009161. <https://doi.org/10.1136/bmjopen-2015-009161>
- Iliho SSK (2019) Community detection methods in social network analysis. In: Abraham A, Dutta P, Kumar Mandal J, Bhattacharya A, Dutta S (eds) *Emerging technologies in data mining and information security*. Springer, Singapore. https://doi.org/10.1007/978-981-13-1498-8_75
- Jensen AB, Moseley PL, Oprea TI, Ellesøe SG, Eriksson R, Schmock H, Jensen PB, Jensen LJ, Brunak S (2014) Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. *Nat Commun* 5:4022. <https://doi.org/10.1038/ncomms5022>
- Jensen K, Soguero-Ruiz C, Oyvind Mikalsen K, Lindsetmo R-O, Kouskoumvekaki I, Girolami M, Olav Skrovseth S, Augestad KM (2017) Analysis of free text in electronic health records for identification of cancer patient trajectories. *Sci Rep* 7(1):46226. <https://doi.org/10.1038/srep46226>
- Ketepe-Arachi T, Sharma S (2017) Cardiovascular disease in women: understanding symptoms and risk factors. *Eur Cardiol* 12(1):10–13. <https://doi.org/10.15420/ecr.2016.32.1>
- Khalid AB, Krum SA (2016) Estrogen receptors alpha and beta in bone. *Bone* 87:130–135. <https://doi.org/10.1016/j.bone.2016.03.016>
- Lane NE (2006) Epidemiology, etiology, and diagnosis of osteoporosis. *Am J Obstet Gynecol* 194(2):S3–S11. <https://doi.org/10.1016/j.ajog.2005.08.047>
- Li X, Meng X, Spiliopoulou A, Timofeeva M, Wei WQ, Gifford A, Shen X, He Y, Varley T, McKeigue P, Tzoulaki I, Wright AF, Joshi P, Denny JC, Campbell H, Theodoratou E (2018) MR-PheWAS: exploring the causal effect of SUA level on multiple disease outcomes by using genetic instruments in UK Biobank. *Ann Rheum Dis* 77(7):1039–1047. <https://doi.org/10.1136/annrheumdis-2017-212534>
- Mandelli L, Nearchou FA, Vaiopoulos C, Stefanis CN, Vitoratou S, Serretti A, Stefanis NC (2015) Neuroticism, social network, stressful life events: Association with mood disorders, depressive symptoms and suicidal ideation in a community sample of women. *Psychiatry Res* 226(1):38–44. <https://doi.org/10.1016/j.psychres.2014.11.001>
- Martínez de Pancorbo C, Carballo F, Horcajo P, Aldeguer M, de la Villa I, Nieto E, Gaspar MJ, de la Morena J (1997) Prevalence and associated factors for gallstone disease: results of a population survey in Spain. *J Clin Epidemiol* 50(12):1347–1355. [https://doi.org/10.1016/s0895-4356\(97\)00198-4](https://doi.org/10.1016/s0895-4356(97)00198-4)
- Missmer SA, Eliassen AH, Barbieri RL, Hankinson SE (2004) Endogenous estrogen, androgen, and progesterone concentrations and breast cancer risk among postmenopausal women. *J Natl Cancer Inst* 96(24):1856–1865. <https://doi.org/10.1093/jnci/djh336>
- Murphy N, Knuppel A, Papadimitriou N, Martin RM, Tsilidis KK, Smith-Byrne K, Fensom G, Perez-Cornago A, Travis RC, Key TJ, Gunter MJ (2020) Insulin-like growth factor-1, insulin-like growth factor-binding protein-3, and breast cancer risk: observational and Mendelian randomization analyses with ~430 000 women. *Ann Oncol* 31(5):641–649. <https://doi.org/10.1016/j.annonc.2020.01.066>
- Nachtigall MJ, Nazem TG, Nachtigall RH, Goldstein SR (2013) Osteoporosis risk factors and early life-style modifications to decrease disease burden in women. *Clin Obstet Gynecol* 56(4):650–653. <https://doi.org/10.1097/GRF.0b013e3182aa1daf>
- World Health Organization (2017) Nutritional anaemias: tools for effective prevention and control. <https://apps.who.int/iris/rest/bitstreams/1091289/retrieve>
- Piva E, Brugnara C, Spolaore F, Plebani M (2015) Clinical utility of reticulocyte parameters. *Clin Lab Med* 35(1):133–163. <https://doi.org/10.1016/j.cll.2014.10.004>
- Pollak M (2008) Insulin and insulin-like growth factor signalling in neoplasia. *Nat Rev Cancer* 8(12):915–928. <https://doi.org/10.1038/nrc2536>
- Pucci G, Alciadi R, Tap L, Battista F, Mattace-Raso F, Schillaci G (2017) Sex- and gender-related prevalence, cardiovascular risk and therapeutic approach in metabolic syndrome: a review of the literature. *Pharmacol Res* 120:34–42. <https://doi.org/10.1016/j.phrs.2017.03.008>
- Rhee SY, Cho HE, Kim JH, Kim HS (2021) Incidence and reappraisal of known risk factors associated with carpal tunnel syndrome: a nationwide, 11-year, population-based study in South Korea. *J Clin Neurol (Seoul, Korea)* 17(4):524–533. <https://doi.org/10.3988/jcn.2021.17.4.524>
- Richardson DB (2004) An incidence density sampling program for nested case-control analyses. *Occup Environ Med* 61(12):e59–e59. <https://doi.org/10.1136/oem.2004.014472>
- Santin AP, Furlanetto TW (2011) Role of estrogen in thyroid function and growth regulation. *J Thyroid Res* 2011:875125–875125. <https://doi.org/10.4061/2011/875125>
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11):2498–2504. <https://doi.org/10.1101/gr.1239303>
- Siggaard T, Reguant R, Jørgensen IF, Haue AD, Lademann M, Aguayo-Orozco A, Hjaltelin JX, Jensen AB, Banasik K, Brunak S (2020) Disease trajectory browser for exploring temporal, population-wide disease progression patterns in 7.2 million Danish patients. *Nat Commun* 11(1):4952. <https://doi.org/10.1038/s41467-020-18682-4>
- Staff NP, Windebank AJ (2014) Peripheral neuropathy due to vitamin deficiency, toxins, and medications. *Continuum (Minneapolis, Minn)* 20:1293–1306. <https://doi.org/10.1212/01.CON.0000455880.06675.5a>
- Stinton LM, Myers RP, Shaffer EA (2010) Epidemiology of Gallstones. *Gastroenterol Clin North Am* 39(2):157–169. <https://doi.org/10.1016/j.gtc.2010.02.003>

- Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, Liu B, Matthews P, Ong G, Pell J, Silman A, Young A, Sprosen T, Peakman T, Collins R (2015) UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 12(3):e1001779. <https://doi.org/10.1371/journal.pmed.1001779>
- Tan VY, Bull CJ, Biernacka KM, Teumer A, Richardson TG, Sanderson E, Corbin LJ, Dudding T, Qi Q, Kaplan RC, Rotter JI, Friedrich N, Völker U, Mayerle J, Perks CM, Holly JMP, Timpson NJ (2021) Investigation of the Interplay between Circulating Lipids and IGF-I and Relevance to Breast Cancer Risk: An Observational and Mendelian Randomization Study. *Cancer Epidemiol Biomark Prev* 30(12):2207–2216. <https://doi.org/10.1158/1055-9965.Epi-21-0315>
- VanderWeele TJ (2014) A unification of mediation and interaction: a 4-way decomposition. *Epidemiology* 25(5):749–761. <https://doi.org/10.1097/ede.0000000000000121>
- Wang C, Crapo LM (1997) The epidemiology of thyroid disease and implications for screening. *Endocrinol Metab Clin North Am* 26(1):189–218. [https://doi.org/10.1016/S0889-8529\(05\)70240-1](https://doi.org/10.1016/S0889-8529(05)70240-1)
- Wang HH, Liu M, Clegg DJ, Portincasa P, Wang DQ (2009) New insights into the molecular mechanisms underlying effects of estrogen on cholesterol gallstone formation. *Biochem Biophys Acta* 1791(11):1037–1047. <https://doi.org/10.1016/j.bbali.2009.06.006>
- Warner MA, Warner ME, Martin JT (1994) Ulnar neuropathy. Incidence, outcome, and risk factors in sedated or anesthetized patients. *Anesthesiology* 81(6):1332–1340. <https://doi.org/10.1097/00000542-199412000-00006>
- Westergaard D, Moseley P, Sørup FKH, Baldi P, Brunak S (2019) Population-wide analysis of differences in disease progression patterns in men and women. *Nat Commun* 10(1):666. <https://doi.org/10.1038/s41467-019-08475-9>
- Yang H, Pawitan Y, He W, Eriksson L, Holowko N, Hall P, Czene K (2019) Disease trajectories and mortality among women diagnosed with breast cancer. *Breast Cancer Res* 21(1):95. <https://doi.org/10.1186/s13058-019-1181-5>