bmjmedicine

Check for updates

Familial aggregation of multimorbidity in Sweden: national explorative family study

Bengt Zöller [1,2] MirNabi Pirouzifard,[1,2] Björn Holmquist [3] Jan Sundquist,[1,2] Anders Halling,[1,2] Kristina Sundquist [1,2]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjmed-2021-000070).

[1]Department of Clinical Science, Lund University, Malmö, Sweden
[2]Centre for Primary Health Care Research, Lund University, Malmö, Sweden
[3]Department of Statistics, Lund University, Lund, Sweden
Correspondence to: Dr Bengt Zöller, Lund University, Lund 221 00, Sweden; bengt.zoller@med.lu.se

Cite this as: *BMJMED* 2023;2:e000070. doi:10.1136/bmjmed-2021-000070

Received: 24 November 2021
Accepted: 1 June 2023

ABSTRACT

OBJECTIVES To examine whether multimorbidity aggregates in families in Sweden.

DESIGN National explorative family study.

SETTING Swedish Multigeneration Register linked to the National Patient Register, 1997-2015. Multimorbidity was assessed with a modified counting method of 45 chronic non-communicable diseases according to ICD-10 (international classification of diseases, 10th revision) diagnoses.

PARTICIPANTS 2 694 442 Swedish born individuals (48.73% women) who could be linked to their Swedish born first, second, and third degree relatives. Twins were defined as full siblings born on the same date.

MAIN OUTCOME MEASURES Multimorbidity was defined as two or more non-communicable diseases. Familial associations for one, two, three, four, and five or more non-communicable diseases were assessed to examine risks depending on the number of non-communicable diseases. Familial adjusted odds ratios for multimorbidity were calculated for individuals with a diagnosis of multimorbidity compared with relatives of individuals unaffected by multimorbidity (reference). An initial principal component decomposition followed by a factor analysis with a principal factor method and an oblique promax rotation was used on the correlation matrix of tetrachoric correlations between 45 diagnoses in patients to identify disease clusters.

RESULTS The odds ratios for multimorbidity were 2.89 in twins (95% confidence interval 2.56 to 3.25), 1.81 in full siblings (1.78 to 1.84), 1.26 in half siblings (1.24 to 1.28), and 1.13 in cousins (1.12 to 1.14) of relatives with a diagnosis of multimorbidity. The odds ratios for multimorbidity increased with the number of diseases in relatives. For example, among twins, the odds ratios for multimorbidity were 1.73, 2.84, 4.09, 4.63, and 6.66 for an increasing number of diseases in relatives, from one to five or more, respectively. Odds ratios were highest at younger ages: in twins, the odds ratio was 3.22 for those aged ≤20 years, 3.14 for those aged 21-30 years, and 2.29 for those aged >30 years at the end of follow-up. Nine disease clusters (factor clusters 1-9) were identified, of which seven aggregated in families. The first three disease clusters in the principal component decomposition were cardiometabolic disease (factor 1), mental health disorders (factor 2), and disorders of the digestive system (factor 3). Odds ratios for multimorbidity in twins, siblings, half siblings, and cousins for the factor 1 cluster were 2.79 (95% confidence interval 0.97 to 8.06), 2.62 (2.39 to 2.88), 1.52 (1.34 to 1.73), and 1.31 (1.23 to 1.39), and for the factor 2 cluster, 5.79 (4.48 to 7.48) 3.24 (3.13 to 3.36), 1.51 (1.45 to 1.57), and 1.37 (1.341.40).

CONCLUSIONS The results of this explorative family study indicated that multimorbidity aggregated in Swedish families. The findings suggest that map clusters of diseases should be used for the genetic study of common diseases to show new genetic patterns of non-communicable diseases.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Multimorbidity, the coexistence of two or more chronic diseases, is a growing challenge for society
⇒ The cause of multimorbidity is not known but a genetic contribution has been suggested

WHAT THIS STUDY ADDS

⇒ This nationwide explorative Swedish study indicated that multimorbidity aggregated in families depending on the degree of relatedness
⇒ The study suggested a genetic component of multimorbidity, although familial lifestyle factors might also contribute
⇒ Some diseases seemed to be clustered in families

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE, OR POLICY

⇒ More research on the genetic factors for multimorbidity is needed
⇒ Identification of high risk families might be a future opportunity for the prevention of multimorbidity

Introduction

Multimorbidity, defined as the coexistence of two or more chronic or long term diseases or medical conditions, is a major challenge for healthcare systems worldwide.[1 2] Multimorbidity is associated with increased mortality, impaired quality of life, and increased consumption of healthcare resources.[3] A cross sectional study from Scotland showed that 23.2% of 1751841 people registered with 314 medical practices in Scotland had multimorbidity.[4] Multimorbidity of non-communicable diseases, however, is a major medical problem for high income countries but also a global burden, affecting low and middle income countries.[5] Factors associated with multimorbidity in the Scottish study were old age, mental health disorders, and living in socioeconomically deprived areas.[4] Apart from old age, female sex and low socioeconomic status seem to be associated

with multimorbidity.[6] Other suggested associated factors for multimorbidity are smoking, physical inactivity, and high body mass index, as well as hypertension and a low educational achievement in men.[6]

Although multimorbidity is expected to increase with longer lifespans, multimorbidity is not a new concept.[7] Moreover, diseases often do not cluster randomly, and diseases with common risk factors are expected to cluster.[7] These common risk factors might be acquired or have a genetic basis, although a genetic cause has yet to be determined.[3]

A key concept in genetic epidemiology is whether evidence exists for disease aggregation in families.[8] Clustering of disease is necessary, although not sufficient, to infer a genetic basis for disease. Whether multimorbidity clusters in families needs to be established, irrespective of whether the cause is genetic or shared family environments.[3] Genome-wide association studies have identified that many genetic variants are associated with several diseases, a phenomenon known as pleiotropy.[9 10] A study of electronic primary healthcare records by Amell et al suggested that shared genetic factors among diseases are linked to certain multimorbidities.[11] A study of hospital inpatient data of 3 85 335 patients in the UK Biobank, which also included genetic data, suggested a shared genetic component for multimorbidity.[12]

In this explorative study, our aim was to determine familial aggregation of multimorbidity based on clinical hospital diagnoses. We used the large and comprehensive family database, the nationwide Swedish Multigeneration Register,[13] to study the inheritance of multimorbidity in Swedish born individuals and their Swedish born first, second, and third degree relatives.

Material and methods

General description of Swedish registers

The Swedish national registers used for data extraction were: Swedish Multigeneration Register, containing data on familial relationships and index persons born in 1932 or later; National Patient Register, giving all hospital discharge diagnoses from 1964 to 2015, with nationwide coverage from 1987 and hospital outpatient diagnoses from 2001 to 2015; Total Population Register, containing data on date of death, marital status, education, and migration with high degree of coverage; and the Swedish Cause of Death Register (1961-2015).[13–18] The databases were linked, as previously described,[19–21] through the personal identity number, which is seldom incorrect.[18] Up to January 2008, an estimated 75 638 (0.56%) individuals had changed their personal identity number compared with the estimated 13 500 000 personal identity numbers in Sweden from 1969 to 31 December 2007.[18]

Study design, study population, and study period

This nationwide retrospective cohort family study was conducted from 1997 to 2015. This study period was chosen because complete nationwide coverage of the National Patient Register was available and we also wanted to avoid problems with the ICD (international classification of disease) codes by using ICD-10 (10th version) only. A common cause of an incorrect personal number is immigration.[18] We therefore only included Swedish born individuals that could be linked to their Swedish born first, second, and third degree relatives to minimise this source of error. Only individuals that could be linked to both of their biological Swedish born parents were included to further minimise the problem of incorrect personal identity numbers.

In the national Swedish Multigeneration Register, families with two full siblings were identified. We also retrieved data on pairs of children born in Sweden between 1948 and 2005 to parents born between 1932 and 1985 in Sweden. Age selection of parents was done so that cousins could be included. We excluded families with members who died or emigrated before 1997 or emigrated before the age of 17. Both biological parents had to be known. Full siblings were then linked to related families and to the half siblings and cousins of the full siblings (we refer to full siblings as siblings in this paper). The same criteria were applied for related families: we excluded all half siblings or cousins who were not born in Sweden, had emigrated before the age of 17 years, or had non-Swedish born parents.

Four different datasets were created: twins, siblings, half siblings, and cousins. Twins were dizygotic and monozygotic twins and were defined as full siblings born on the same date. In the datasets, all relative pairs were entered twice (ie, all sibling pairs, all twin pairs, all half sibling pairs, and all cousin pairs, as described previously).[19 20 22 23] This double entry approach is a common procedure in genetics.[22 23] Zygosity was assessed only from sex (ie, twins of different sexes were considered to be dizygotic twins). We had no access to zygosity data. We allowed the same person to be included in more than one family relationship.

Multimorbidity score

No standard approach for the measurement of multimorbidity exists,[4] and therefore the selection and definition of morbidities were partly subjective and dependent on the data available.[4] Any disease recommended as a core for a multimorbidity measure and major long term disorder was therefore included, based on the cross sectional study of Barnett et al.[4] The multimorbidity score was modified and adapted to the Swedish ICD-10 codes after Barnett et al (online supplemental table S1).[4] Infectious diseases (ie, viral hepatitis) were not included because their main cause is not genetic. To compensate for the

exclusion of viral hepatitis and to increase the number of diseases, we added five more common non-communicable diseases or conditions associated with considerable morbidity: osteoporosis, arthrosis, gout, obesity, and pancreatic diseases. All five conditions also have a potential polygenic background. The clinicians in our team (BZ, JS, AH, and KS) agreed on the added diagnoses after discussion. Also, psoriasis was counted as a non-communicable disease separate from dermatitis and eczema.

Online supplemental table S1 lists the diagnoses and ICD-10 codes used. One point was awarded for each of the 45 non-communicable disorders and theoretically, individuals could be assigned a multimorbidity score between 0 and 45 points. No patient, however, had more than 20 points (online supplemental table S2). Individuals with two or more points were considered to have multimorbidity, in agreement with recent definitions.[1 2 4] Familial associations for one, two, three, four, and five or more non-communicable diseases were also assessed. Diagnoses in the National Patient Register are coded according to the Swedish ICD-10 system (adapted from the World Health Organizationn ICD classification system).[15] The Swedish Hospital Discharge Register has almost 90% overall validity or positive predictive values.[15]

Statistical analysis

We investigated the crude and adjusted familial associations between multimorbidity scores (one, two, three, four, and five or more diseases) for twins, siblings, half siblings, and cousins and multimorbidity (yes/no) with logistic regression.[24] The ambiguity in unselected samples for which trait in twins, siblings, half siblings, and cousins should be used as the dependent variable and which should be used as the independent variable is frequently resolved with double entry.[22 23] Each twin, sibling, half sibling, or cousin is entered twice in the data, and each member of a twin, sibling, half sibling, or cousin pair provides a dependent variable once and an explanatory variable once. Although the consistency of the regression estimates for heritability and environmental influences is not affected by double entry, the standard errors of the coefficients are biased and need to be adjusted.[22 23] In the logistic regression models in this study, we used the variance covariance cluster method with Stata, which calculated robust standard errors with families as clusters.[25] The variance covariance cluster method specifies that the standard errors allow for intragroup correlation, relaxing the usual requirement that the observations are independent. The results obtained with this technique are also (in this case) consistent with the results obtained with the standard technique to double the variances obtained from the double entry approach to correct for the dependence between the two reversed order entries.[22 23] These latter results are thus not reported.

The study of inheritance starts with an individual who is affected by a genetic condition or who is concerned that they are at risk of a condition. This person is referred to as the proband. In this study, we compared the familial odds of multimorbidity in people with multimorbidity compared with people without multimorbidity. Thus a proband could be affected or unaffected by multimorbidity. Results are reported as familial odds ratios (95% confidence intervals); that is, the odds ratio comparing relatives with an affected relative (ie, affected proband) with relatives with an unaffected relative (ie, unaffected proband).[24] Familial odds ratios for multimorbidity were calculated for relatives of individuals who had a multimorbidity score of two or more (affected proband) compared with relatives of individuals with a multimorbidity score of none or, at most, one disease (unaffected proband). Familial odds ratios were also calculated according to sex and age. Complex traits have been reported to show stronger inheritance at younger ages.[8]

Models were adjusted for year of birth, sex, region of birth (county, online supplemental table S3), and level of educational achievement (as an indicator of socioeconomic status and lifestyle related factors). Level of educational achievement was categorised into four groups: unknown; ≤9 years of education (elementary school); 10-11 years of education (vocational upper secondary education); and (4) ≥12 years of education (secondary school, college, or university). Unknown education was kept as one category and was not added because of the possibility of missing not at random (ie, the probability that missing data for a specific variable could be related to the values of this variable itself).[26]

To determine whether the observed odds ratios could be explained by unobserved confounders rather than causal effects, E values were calculated.[27] An E value is defined as the minimum strength of association that an unmeasured confounder would need to have with both the predictor and the outcome to fully explain a specific predictor-outcome association.[27] A large E value indicates that considerable unmeasured confounding would be needed to explain an effect estimate. A small E value indicates that little unmeasured confounding would be needed to explain an effect estimate.[27] The lower 95% confidence interval limits for the E values were also calculated. The average genetic resemblance for twins, both monozygotic and dizygotic, was determined as 0.66 with Weinberg's differential method. Each relative pair was assigned their average genetic resemblance (ie, 0.66 for twin pairs, 0.5 for sibling pairs, 0.25 for half sibling pairs, and 0.125 for cousin pairs).

Principal component analysis is a statistical technique for exploratory studies, features extraction, dimensionality reduction, and data compression.[28 29] Briefly, to identify disease clusters, an initial principal

component decomposition followed by a factor analysis with a principal factor method and an oblique promax rotation was used on the correlation matrix of tetrachoric correlations between 45 diagnoses in patients.[28 29] The statistical section in the online supplemental file provides a detailed description. Statistical significance was set at P<0.05 and all tests were two tailed. Data were analysed with SAS version 9.4 (SAS Institute, Cary, NC) and Stata version 16.1 (StataCorp, College Station, TX).

Patient and public involvement

Owing to regulatory and study design constraints, neither patients or members of the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. The patients' data were anonymised so we will not be able to disseminate the results to individual participants. The research team will use academic and public dissemination channels to inform the public of the results. Our findings will be shared on social media and relevant websites.

Results

The study population (n=2 694 442 unique individuals) comprised 1 570 128 siblings, of whom 24 020 were twins. Table 1 summarises the personal characteristics of the individuals in the study sample. We found that 440 742 (16.36% of the total study population) people had a diagnosis of multimorbidity (a score of two or more) during the period 1997-2015. Table 1 shows the details of the specific study subsets; 58% (n=253 931) of the multimorbidity individuals were women, with a median age of 35 years (interquartile range 24-46) at the end of the study. Data for sex were taken from information in the Swedish national registers rather than from patient reported gender. The proportion of women increased with higher multimorbidity scores. Lower educational level, older age at the end of the study, and earlier birth year were also associated with multimorbidity. We found the same pattern of the association between multimorbidity scores and sex, educational level, age at the end of the study, and birth year among all studied relative pairs (ie, twins, siblings, half siblings, and cousins, online supplemental table S4).

Familial risk of multimorbidity

Table 2 shows the familial crude and adjusted odds ratios for multimorbidity (score of ≥2) and corresponding E values for twins, siblings, half siblings, and cousins of affected probands with one to five or more diseases (proband) in analyses that were compared with relatives with no diseases (unaffected proband). We found a clear progressive response, with higher familial odds ratios when the relatives of the affected proband had higher scores. For example, among twins, the adjusted odds ratio

Table 1 | Characteristics of all study participants, grouped by multimorbidity scores for number of unique individuals, sex, education, age at the end of study, and birth date

Characteristics	Multimorbidity score								
	All	0	1	2	3	4	≥5	≤1	≥2
No (%) individuals	2 694 442	1 616 213 (59.98)	637 487 (23.66)	254 210 (9.43)	104 634 (3.88)	45 585 (1.69)	36 313 (1.35)	2 253 700 (83.64)	440 742 (16.36)
No (%) women	1 312 989 (48.73)	731 337 (45.25)	327 721 (51.41)	141 488 (55.66)	61 808 (59.07)	27 661 (60.68)	22 974 (63.27)	1 059 058 (46.99)	253 931 (57.61)
No (%) with ≥12 years of education	818 146 (30.36)	516 774 (31.97)	189 127 (29.67)	69 050 (27.16)	25 946 (24.80)	10 232 (22.45)	7017 (19.32)	705 901 (31.32)	112 245 (25.47)
Median (IQR, range)* age (years) at end of study	32 (22-43, 0-68)	31 (22-42, 0-67)	32 (22-43, 0-67)	33 (23-45, 0-67)	36 (25-46, 1-66)	38 (27-48, 1-67)	42 (31-50, 3-68)	31 (22-42, 0-67)	35 (24-46, 0-68)
Median (IQR, range)* year of birth	1983 (1972-93, 1947-2005)	1984 (1973-93, 1948-2005)	1983 (1972-93, 1948-2005)	1981 (1970-92, 1948-2005)	1979 (1968-90, 1949-2005)	1977 (1967-88, 1948-2005)	1972 (1964-84, 1947-2005)	1984 (1973-93, 1948-2005)	1980 (1969-90, 1947-2005)

IQR=interquartile range.
*Range=minimum–maximum value.

Table 2 | Odds ratios for multimorbidity (two or more diseases) in relatives according to multimorbidity score (0 to ≥5) in proband

	Twins (n=24 020)		Siblings* (n=1 546 108)		Half siblings (n=984 976)		Cousins (n=6 623 156)	
Proband score	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Score 0	1 (Reference)	1 (Reference)	1 (Reference)	1 (Reference)	1 (Reference)	1 (Reference)	1 (Reference)	1 (Reference)
Score 1	1.77 (1.62 to 1.93)	1.73 (1.59 to 1.89)	1.37 (1.36 to 1.39)	1.34 (1.33 to 1.36)	1.14 (1.13 to 1.15)	1.12 (1.11 to 1.14)	1.09 (1.08 to 1.09)	1.07 (1.06 to 1.07)
E value: PE (low CI)	2.94 (2.62)	2.85 (2.56)	2.08 (2.06)	2.02 (1.99)	1.54 (1.51)	1.49 (1.46)	1.40 (1.37)	1.34 (1.31)
Score 2	2.95 (2.56 to 3.40)	2.84 (2.46 to 3.28)	1.85 (1.82 to 1.88)	1.77 (1.74 to 1.80)	1.29 (1.27 to 1.32)	1.25 (1.23 to 1.28)	1.16 (1.15 to 1.17)	1.12 (1.11 to 1.13)
E value: PE (low CI)	5.35 (4.56)	5.13 (4.36)	3.10 (3.04)	2.94 (2.90)	1.90 (1.86)	1.81 (1.76)	1.59 (1.57)	1.49 (1.46)
Score 3	4.38 (3.66 to 5.23)	4.09 (3.41 to 4.91)	2.27 (2.22 to 2.32)	2.11 (2.07 to 2.16)	1.40 (1.37 to 1.44)	1.33 (1.30 to 1.37)	1.24 (1.22 to 1.26)	1.17 (1.16 to 1.19)
E value: PE (low CI)	8.23 (6.78)	7.65 (6.28)	3.97 (3.87)	3.64 (3.56)	2.15 (2.08)	1.99 (1.92)	1.79 (1.74)	1.62 (1.59)
Score 4	5.18 (4.07 to 6.58)	4.63 (3.61 to 5.93)	2.70 (2.62 to 2.78)	2.43 (2.36 to 2.51)	1.51 (1.46 to 1.56)	1.39 (1.35 to 1.44)	1.33 (1.31 to 1.36)	1.23 (1.21 to 1.25)
E value: PE (low CI)	9.83 (7.61)	8.73 (6.68)	4.84 (4.68)	4.29 (4.15)	2.39 (2.28)	2.13 (2.04)	1.99 (1.95)	1.76 (1.71)
Score ≥5	7.96 (5.98 to 10.60)	6.66 (4.97 to 8.93)	3.20 (3.10 to 3.31)	2.71 (2.62 to 2.81)	1.65 (1.59 to 1.71)	1.47 (1.41 to 1.52)	1.39 (1.36 to 1.42)	1.23 (1.21 to 1.26)
E value: PE (low CI)	15.40 (11.44)	12.80 (9.41)	5.85 (5.65)	4.86 (4.70)	2.69 (2.56)	2.30 (2.17)	2.13 (2.06)	1.76 (1.71)

Data show odds ratios with 95% confidence intervals for multimorbidity scores among relatives unless stated otherwise. References are relatives with probands with no diseases (score=0). The E value is defined as the minimum strength of association that an unmeasured confounder would need to have with both the predictor and outcome to fully explain the association, conditional on the measured covariates.

Low CI is the lower limit of the 95% confidence interval of the E value.

Odds ratios were derived from double entry.

Model 1 a crude model (univariate). Model 2 is an adjusted model (multivariate), with adjustments for sex, year of birth, county, and educational achievement. In the logistic regression models, the variance covariance cluster method in Stata was used, which calculated robust standard errors with families as clusters.

*Excluding twins as siblings.

CI, confidence interval; PE, point estimate.

for multimorbidity was 1.73 for twin relatives with a multimorbidity score of one, 2.84 for a multimorbidity score of two, 4.09 for a multimorbidity score of three, 4.63 for a multimorbidity score of four, and 6.66 for a multimorbidity score of five or more. A declining response was also seen depending on the average genetic resemblance. The adjusted odds ratio was 6.66 (95% confidence interval 4.97 to 8.93) for multimorbidity (two or more diseases) if the multimorbidity score was five or more among twins, 2.71 (2.62 to 2.81) for siblings, 1.47 (1.41 to 1.52) for half siblings, and 1.23 (1.21 to 1.26) for cousins. The corresponding adjusted E values with the lower 95% confidence interval limits were 12.80 (9.41) for twins, 4.86 (4.70) for siblings, 2.30 (2.17) for half siblings, and 1.76 (1.71) for cousins (table 2).

Table 3 shows the odds ratios and corresponding E values for multimorbidity (≥2 score) if the proband relative had multimorbidity (two or more diseases). The odds ratios in the adjusted models for multimorbidity were 2.89 (95% confidence interval 2.56 to 3.25) in twins, 1.81 (1.78 to 1.84) in siblings, 1.26 (1.24 to 1.28) in half siblings, and 1.13 (1.12 to 1.14) in cousins in those having a diagnosis of an index score of more than one (table 3). The corresponding adjusted E values were 5.23 (lower 95% confidence interval limit 4.56) for twins, 3.02 (2.96) for siblings, 1.83 (1.79) for half siblings, and 1.51 (1.49) for cousins (table 3).

Risk of multimorbidity by age and sex

The familial odds ratios were highest at younger ages among all relatives (online supplemental tables S5–S8). For example, in twins, the adjusted odds ratio was 3.22 (95% confidence interval 2.65 to 3.92) for those aged ≤20 years, 3.14 (2.50 to 3.93) for those aged 21-30 years, and 2.29 (1.87 to 2.80) for those aged >30 years at the end of follow-up.

Among twins, the strongest association was observed between male twins (adjusted odds ratio 3.44, 95% confidence interval 2.74 to 4.32); in female twins, the adjusted odds ratio was 3.60 (2.97 to 4.37; online supplemental table S5). Twins of the same sex were monozygotic or dizygotic twins. In twins of opposite sex (dizygotic), odds ratios were lower (1.92 if male proband and 1.91 if female proband). No major differences between the sexes were observed in siblings, half siblings, or cousins (online supplemental tables S6–S8).

Principal component analysis followed by factor analysis

For principal component analysis followed by factor analysis, disease clusters among individual patients and not between family members were used. The full sibling dataset was used for this analysis. The number of diseases was reduced to nine different disease clusters (online supplemental table S9 and online supplemental figure S1). We used visual inspection to find a

Table 3 | Odds ratios for multimorbidity (two or more diseases) according to multimorbidity score in proband relatives

	Twins (n=24 020)		Siblings* (n=1 546 108)		Half siblings (n=984 976)		Cousins (n=6 623 156)	
Proband score	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Score ≤1	1 (Reference)	1 (Reference)	1 (Reference)	1 (Reference)	1 (Reference)	1 (Reference)	1 (Reference)	1 (Reference)
Score ≥2	3.06 (2.72 to 3.45)	2.89 (2.56 to 3.25)	1.92 (1.89 to 1.95)	1.81 (1.78 to1.84)	1.32 (1.30 to 1.35)	1.26 (1.24 to 1.28)	1.19 (1.18 to 1.20)	1.13 (1.12 to 1.14)
E value: PE (low CI)	5.57 (4.88)	5.23 (4.56)	3.25 (3.19)	3.02 (2.96)	1.97 (1.92)	1.83 (1.79)	1.67 (1.64)	1.51 (1.49)

PE=point estimate; CI=confidence interval.
Data show odds ratios with 95% confidence intervals for multimorbidity scores among relatives unless stated otherwise. References are relatives with probands with no or one disease (score ≤1). The E value is defined as the minimum strength of association that an unmeasured confounder would need to have with both the predictor and outcome to fully explain the association, conditional on the measured covariates.
Low CI is the lower limit of the 95% confidence interval of the E value.
Odds ratios were derived from double entry.
Model 1 is a crude model (univariate). Model 2 is an adjusted model (multivariate), with adjustments for sex, year of birth, county, and educational achievement. In the logistic regression models, the variance covariance cluster method in Stata was used, which calculated robust standard errors with families as clusters.
*Excluding twins as siblings.

Table 4 | Rotated factor loadings (pattern matrix) and unique variances sorted

Variables	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9	Uniqueness
Hypertension	0.84									0.23
Heart failure	0.77									0.31
Coronary heart disease	0.74									0.39
Diabetes	0.73									0.46
Obesity	0.62									0.44
Atrial fibrillation	0.61									0.54
Gout	0.52									0.66
Atherosclerosis	0.34									0.58
Renal disease	0.34									0.67
Affective disorders		0.88								0.21
Anxiety		0.86								0.24
Psychoactive substance misuse		0.82								0.26
Alcohol misuse disorders		0.71								0.38
Anorexia or bulimia		0.61								0.35
Schizophrenia		0.60								0.56
Inflammatory bowel disease			0.75							0.45
Liver disease			0.64							0.39
Pancreatic disease			0.56							0.42
Ulcers			0.44							0.56
Epilepsy				0.71						0.45
Blindness and poor vision				0.66						0.49
Cerebrovascular disease				0.61						0.38
Cancer				0.31						0.63
Impaired or hearing loss				0.29						0.84
Connective tissue disease					0.62					0.54
Osteoporosis					0.59					0.42
Thyroid disorders					0.56					0.61
Psoriasis					0.36					0.77
Prostate disease						0.66				0.55
Arthrosis						0.55				0.54
Painful back condition						0.52				0.62
Diverticular disease of intestine						0.50				0.63
Chronic sinusitis						0.40				0.68
Bronchiectasis							−0.99			0.03
Parkinson's disease							0.55			0.49
Glaucoma							0.50			0.57
Learning disability							0.48			0.50
Irritable bowel syndrome							0.45			0.45
Asthma								0.87		0.33
Dermatitis and eczema								0.47		0.69
Constipation								0.46		0.61
Chronic obstructive pulmonary disease								0.39		0.44
Migraine								0.27		0.73
Multiple sclerosis									−0.87	0.11
Dementia									0.84	0.6

An oblique promax rotation was used on the correlation matrix of tetrachoric correlations between the 45 diagnoses in patients to identify disease clusters. Rotated factor loading put the diagnoses into nine different factors. Uniqueness is the proportion of the common variance of the variable not associated with the factors. Uniqueness=1−communality.

point at which the amount of variance explained by subsequent principal component dropped off (online supplemental figure S1). After nine factors, the eigenvalues levelled off (elbow) and nine factors were therefore extracted. In table 4, the rotated factor loadings (pattern matrix) and unique variances are sorted. The loading pattern determines the factor that has the most influence on each variable. Loadings close to −1 or 1 indicate that the factor strongly influences the variable. Loadings close to 0 indicate that the factor has a weak influence on the variable. Online supplemental table S10 shows the distribution of diseases in these nine disease clusters (factor clusters 1-9). As part of this method, all conditions were allocated to one of the nine clusters. For example, in the factor 1 cluster (mainly cardiometabolic disorders),

hypertension, heart failure, coronary heart disease, and diabetes had the strongest contributions in this group (table 4). In the factor 2 cluster (mental health disorders), affective disorders, anxiety, psychoactive substance misuse, and alcohol misuse disorders were the most important in this group (ie, with the largest positive loadings; table 4).

For seven of the nine disease groups identified by factor analysis (factors 1-6, and factor 8), the odds ratios were closely related to the degree of average genetic resemblance, with the highest in twins and then decreasing stepwise with the degree of relatedness; the lowest were in cousins but these values were still significant (table 5). Online supplemental tables S11–S19 show the familial crude and adjusted odds ratios for multimorbidity (two or more diseases) in twins, siblings, half siblings, and cousins of probands with one to five or more diseases in analyses that were compared with relatives with a proband with no diseases. The two most important disease cluster groups identified (ie, with the highest eigenvalue (online supplemental table S9), were factor 1 (hypertension, heart failure, coronary heart disease, diabetes, obesity, atrial fibrillation, gout, atherosclerosis, and renal disease) and factor 2 (affective disorders, anxiety, psychoactive substance misuse, alcohol misuse disorders, anorexia or bulimia, and schizophrenia disorders) (table 5). These two disease groups were strongly clustered, with odds ratios determined by average genetic resemblance (table 5 and online supplemental tables S11 and S12). The disease groups factor 3 (inflammatory bowel disease, liver disease, pancreatic disease, and ulcers), factor 4 (epilepsy, blindness and poor vision, cerebrovascular disease, cancer, and impaired or hearing loss), factor 5 (connective tissue disease, osteoporosis, thyroid disorders, and psoriasis), factor 6 (prostate disease, arthrosis, painful back condition, diverticular disease of intestine, and chronic sinusitis), and factor 8 (asthma, dermatitis and eczema, constipation, chronic obstructive pulmonary disease, and migraine) were also clustered (table 5 and online supplemental tables S13–S16 and S18). Only multimorbidity of the disease clusters factor 7 (bronchiectasis, Parkinson's disease, glaucoma, learning disability, and irritable bowel syndrome) and factor 9 (multiple sclerosis and dementia) were not clustered (table 5 and online supplemental tables S17 and S19).

Additional information

Online supplemental tables S20-S22 show the number of observations and individuals used for the calculations in table 2, table 3, and table 5 (ie, number of individuals and number of persons at risk).

Discussion

Principal findings

In this large nationwide explorative family study, based on complete and validated data sources, we have provided robust estimates correlating with average genetic resemblance, suggesting a hereditary component of multimorbidity. Hence our results support other studies indicating a genetic contribution to multimorbidity.[11 12] According to a study by Khoury et al,[30] even with complete correlation in exposure among first degree relatives, environmental risk factors with relative risks of <10 gave modest familial relative risks (1-2) and low recurrence risks, suggesting that the high risk of disease specific multimorbidity has a major genetic contribution. Moreover, the calculated E values in our study suggest that strong unmeasured confounders are necessary to explain our findings.[27] Nevertheless, we cannot rule out familial lifestyle factors contributing to the familial aggregation of multimorbidity. This study showed the inheritance of specific multimorbidity clusters, suggesting that further genetic research, such as genome wide association studies, could be worthwhile. The study also showed that several disease clusters aggregated in families.

A plausible cause of the familial aggregation of multimorbidity is the abundance of pleiotropy for complex traits observed in genome wide association studies.[9] Seven of the nine disease clusters identified with the factor analysis in individuals showed familial aggregation. The strongest aggregation of multimorbidity was observed in twins and pairs of full siblings, whereas these associations, although significant, were reduced in half siblings and cousins. The associations correlated with average genetic resemblance, which suggests that genetic factors are important in the familial aggregation of multimorbidity.[19–21] Moreover, multimorbidity inheritance was graded in terms of the number of diseases in probands. The higher odds ratios in younger than in older individuals also suggests a genetic basis.[8 31] Also, third degree relatives usually do not share a household and the increased familial odds ratios among cousins suggests a genetic contribution to the familial aggregation of multimorbidity. Older individuals, women, and those with a lower educational achievement had a higher multimorbidity risk across all of the family relationships. These findings confirm previous studies and the validity of this study.[4 6]

The first step in identifying genetic determinants in complex diseases is to study the familial aggregation of the phenotype.[8 31] The Swedish Multigeneration Register offers the opportunity to explore the influence of genetic and non-genetic familial factors in a large number of families by observing first, second, and third degree relatives and the occurrence of specific phenotypes.[19–21] First degree relatives (eg, full siblings) share 50% of their genes, as well as shared environmental exposures common to their

Table 5 | Odds ratios for multimorbidity (two or more diseases) according to multimorbidity score in proband relatives

	Twins (n=24 020)		Siblings* (n=1 546 108)		Half siblings (n=984 976)		Cousins (n=6 623 156)	
Proband score	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Score ≤1	1 (Reference)	1 (Reference)	1 (Reference)	1 (Reference)	1 (Reference)	1 (Reference)	1 (Reference)	1 (Reference)
Factor 1: Hypertension, heart failure, coronary heart disease, diabetes, obesity, atrial fibrillation, gout, atherosclerosis, and renal disease								
Score ≥2	9.61 (3.79 to 24.39)	2.79 (0.97 to 8.06)	6.20 (5.67 to 6.78)	2.62 (2.39 to 2.88)	2.69 (2.36 to 3.06)	1.52 (1.34 to 1.73)	2.21 (2.08 to 2.35)	1.31 (1.23 to 1.39)
E value: PE (low CI)	18.71 (7.04)	5.02 (1.00)	11.88 (10.82)	4.68 (4.21)	4.82 (4.15)	2.41 (2.02)	3.85 (3.58)	1.95 (1.76)
Factor 2: Affective disorders, anxiety, psychoactive substance misuse, alcohol misuse disorders, anorexia or bulimia, and schizophrenia disorders								
Score ≥2	6.60 (5.15 to 8.45)	5.79 (4.48 to 7.48)	3.29 (3.18 to 3.41)	3.24 (3.13 to 3.36)	1.54 (1.48 to 1.60)	1.51 (1.45 to 1.57)	1.38 (1.35 to 1.41)	1.37 (1.34 to 1.40)
E value: PE (low CI)	12.68 (9.77)	11.06 (8.43)	6.04 (5.81)	5.93 (5.71)	2.45 (2.32)	2.39 (2.26)	2.10 (2.04)	2.08 (2.02)
Factor 3: Inflammatory bowel disease, liver disease, pancreatic disease, and ulcers								
Score ≥2	NA	NA	9.80 (4.87 to 19.74)	6.86 (3.39 to 13.88)	3.63 (1.16 to 11.30)	2.44 (0.78 to 7.57)	1.24 (0.56 to 2.76)	1.02 (0.46 to 2.26)
E value: PE (low CI)	NA	NA	19.09 (9.21)	13.20 (6.24)	6.72 (1.59)	4.31 (1.00)	1.79 (1.00)	1.16 (1.00)
Factor 4: Epilepsy, blindness and poor vision, cerebrovascular disease, cancer, and impaired or hearing loss								
Score ≥2	9.34 (2.21 to 39.41)	5.38 (1.07 to 27.19)	3.50 (2.46 to 4.97)	2.59 (1.82 to 3.70)	0.90 (0.45 to 1.80)	0.75 (0.38 to 1.51)	1.17 (0.90 to 1.53)	1.01 (0.77 to 1.32)
E value: PE (low CI)	18.17 (3.85)	10.23 (1.34)	6.46 (4.36)	4.62 (3.04)	1.46 (1.00)	2.00 (1.00)	1.62 (1.00)	1.11 (1.00)
Factor 5: Connective tissue disease, osteoporosis, thyroid disorders, and psoriasis								
Score ≥2	46.79 (5.85 to 374.12)	27.42 (2.21 to 340.29)	7.32 (4.88 to 10.98)	5.20 (3.43 to 7.89)	5.29 (2.99 to 9.36)	4.01 (2.24 to 7.19)	2.12 (1.53 to 2.94)	1.66 (1.20 to 2.31)
E value: PE (low CI)	93.08 (11.18)	54.36 (3.85)	14.12 (9.23)	9.87 (6.32)	10.05 (5.43)	7.48 (3.91)	3.66 (2.43)	2.71 (1.69)
Factor 6: Prostate disease, arthrosis, painful back condition, diverticular disease of intestine, and chronic sinusitis								
Score ≥2	31.92 (9.34 to 109.08)	4.28 (1.17 to 15.61)	7.15 (6.00 to 8.54)	2.33 (1.94 to 2.79)	3.45 (2.71 to 4.38)	1.61 (1.26 to 2.04)	2.22 (1.95 to 2.53)	1.18 (1.04 to 1.34)
E value: PE (low CI)	63.34 (18.17)	8.03 (1.62)	13.78 (11.48)	4.09 (3.29)	6.36 (4.86)	2.60 (1.83)	3.87 (3.31)	1.64 (1.24)
Factor 7: Bronchiectasis, Parkinson's disease, glaucoma, learning disability, and irritable bowel syndrome								
Score ≥2	NA	NA	NA	NA	NA	NA	NA	NA
E value: PE (low CI)	NA	NA	NA	NA	NA	NA	NA	NA
Factor 8: Asthma, dermatitis and eczema, constipation, chronic obstructive pulmonary disease, and migraine								
Score ≥2	12.31 (9.47 to 16.01)	9.07 (6.92 to 11.90)	6.28 (5.98 to 6.60)	4.40 (4.18 to 4.63)	2.08 (1.89 to 2.29)	1.90 (1.73 to 2.10)	1.76 (1.67 to 1.86)	1.39 (1.32 to 1.46)
E value: PE (low CI)	24.11 (18.43)	17.63 (13.32)	12.04 (11.44)	8.27 (7.83)	3.58 (3.19)	3.21 (2.85)	2.92 (2.73)	2.13 (1.97)
Factor 9: Multiple sclerosis and dementia								
Score ≥2	NA	NA	NA	NA	NA	NA	NA	NA
E value: PE (low CI)	NA	NA	NA	NA	NA	NA	NA	NA

NA=not applicable because of too few individuals; PE=point estimate.

Data show odds ratios with 95% confidence intervals for multimorbidity scores among relatives unless stated otherwise. References are relatives with probands with no or one disease (score ≤1). The E value is defined as the minimum strength of association that an unmeasured confounder would need to have with both the predictor and outcome to fully explain the association, conditional on the measured covariates.

Low CI is the lower limit of the 95% confidence interval of the E value.

Odds ratios were derived from double entry.

Model 1 is a crude model (univariate). Model 2 is an adjusted model (multivariate), with adjustments for sex, year of birth, county, and educational achievement. In the logistic regression models, the variance covariance cluster method in Stata was used, which calculated robust standard errors with families as clusters.

*Excluding twins as siblings.

family. Second degree relatives (eg, half siblings) share 25% of their genes, and third degree relatives (eg, first cousins) share 12.5% of their genes.[19–21] In this study, the increasing probability of multimorbidity among relatives of affected individuals was linked to the closeness of the relationships, being strongest in twins (66% average genetic resemblance). Hence our observations support multimorbidity heredity, although the specific underlying genetic mechanisms are not known.

The agnostic method, based on principal component analysis and the oblique rotation in the factor analysis to identify disease clusters at the individual level, identified familial aggregation of multimorbidity in seven of the nine disease groups (tables 4 and 5). This finding agrees with studies showing that diseases with a higher probability of concurrency tend to share more associated genes.[32 33] The cardiometabolic diseases in the factor 1 cluster were previously described to be interrelated.[34] Some of the conditions included in the factor 1 cluster were in a causal pathway with coronary heart disease (eg, obesity, diabetes, hypertension, and atherosclerosis). This finding could contribute to the clustering in the factor 1 group. The mental health disorders in the factor 2 cluster are also well known to be linked.[35] Our study agrees with other studies identifying mental health disorders and cardiometabolic conditions as the two most reproducible groups of multimorbidity.[36] Inflammatory bowel disease and autoimmune liver disease are also known to be linked in the factor 3 cluster.[37] A well known linkage exists between autoimmune disorders, such as certain connective tissue disorders, thyroid disease, and psoriasis (factor 5 cluster).[38] Thus the high familial aggregation of the factor 5 cluster is not unexpected.

Eczema and asthma in the factor 8 cluster are well known to be associated.[39] Chronic obstructive pulmonary disease clustered in the same group as asthma, which is a common feature for many multimorbidity studies.[36] The reason for the inclusion of constipation in the factor 8 cluster is not obvious but a recent report has described an association between asthma and constipation.[40] Migraine has also been shown to be associated with asthma.[41] Other associations might be acquired from treatments, such as osteoporosis clustering with connective tissue diseases (ie, because of treatment with corticosteroids).[42] Many studies have reported that musculoskeletal disorders cluster together (eg, arthrosis and painful back condition, factor 6 cluster in tables 4 and 5).[36] These examples show the strength and proof of concept of our study design for the use of principal component analysis and factor analysis for studying disease clustering. Nevertheless, some associations are less intuitive (eg, why cerebrovascular disease clustered in the factor 4 cluster with hearing loss instead of the cardiometabolic factor 1 cluster), although a recent systematic review and meta-analysis found an association between sensorineural hearing loss and an increased risk of stroke.[43] Also, hearing loss has been reported to be related to the risk of stroke but not coronary heart disease.[44]

Strengths and limitations of this study

In common with all epidemiological studies, the interpretation of results is constrained by time (1997-2015) and geographical location (Sweden), resulting in bias in time period and location. Our inclusion and exclusion criteria ensured that disease events in a family could be registered in the Swedish patient registers (1997-2015). This method is a limitation because families with diseases before 1997 were not included, although this bias is most likely non-differential. Asking a proband about the family history of a disease is a major source of self-report and recall bias, however, which likely exceeds the study limitations of this study design.[45] Another limitation was the lack of information about lifestyle factors, such as smoking, consumption of alcohol, and physical activity. We adjusted for education that correlates with lifestyle factors, such as smoking,[46] but residual bias is likely to exist. Thus we cannot rule out the contribution of family environment and lifestyle factors, such as smoking, to the familial aggregation of multimorbidity. Another limitation was that we did not distinguish between dizygotic and monozygotic twins.

We did not consider the time courses of the diseases in individuals or which of the 45 diseases in the multimorbidity scores a patient was affected by. Different diseases could cluster in different families. With the principal component analysis and factor analysis method to identify diseases that cluster in individual patients, however, we identified disease clusters with familial aggregation. Further studies to establish smaller disease clusters could be worthwhile, although our study agrees with many other reports, including a systematic review of the two major groups of diseases (mental health disorders and cardiometabolic disease).[34–42]

We believe that the use of principal component analysis for studying disease clusters is an advantage over k means. The k means algorithm requires some initialisation of the centroid positions.[47] For most algorithms, these centroids are randomly initialised with some method, such as the Forgy method, or random partitioning, which means that repeated iterations of the algorithm can converge to vastly different results.[47] We also saw this effect and therefore we did not use k means because the clustering was not reproducible, whereas principal component analysis gave identical results. Nevertheless, we found some similarities with our grouping results between k means and principal component analysis. We also tried latent class analysis for cluster analysis but because of the large numbers, the results did not converge for the latent class analysis. A known limitation of latent class analysis for cluster

analysis is that it is computationally expensive, which might be inconvenient with large datasets.[48] Moreover, selecting the number of clusters is a challenging task involving inevitable subjective analytical choices.[48] With the principal component analysis method, however, we reproduced the two most commonly reported clusters of multimorbidity (mental health disorders and cardiometabolic disease), which shows the strength of the principal component analysis method.[34–42]

Some of the 45 conditions can be acute as well as chronic (eg, constipation). We could not take this limitation into consideration or the order in which patients were affected by different diseases. Other methods for studying disease clustering could give divergent results but our two main groups (mental health disorders and cardiometabolic disease) were similar to many other studies.[36] A systematic review showed that cardiovascular and metabolic diseases, mental health problems, and allergic diseases were major disease clusters, regardless of the method used for studying disease clustering and multimorbidity patterns.[49]

The large size of our study was a major advantage. A limitation but also a strength was the use of nationwide registers with almost complete data coverage and high validity.[13–18] A previous validation of the National Patient Register by the National Board of Health and Welfare showed that 85-95% of all diagnoses were in the National Patient Register.[15] The use of clinical hospital diagnoses allowed for the elimination of recall bias. Recall and self-report bias are common problems in many family studies.[45] The objective data in the Swedish Multigeneration Register could therefore be an advantage compared with self-reported data.[13] Age selection of parents was done so that cousins could be included. Median age at the end of follow-up was 32 years (range 0-68), limiting the applicability of the results to older people. Heritability in complex traits is dependent on age, however, and young age is an advantage in an explorative study of heredity,[31] including multimorbidity inheritance. The observed age dependence of the familial odds ratio in this study suggests a genetic cause and that multimorbidity is a complex trait.[8 31] Acquired factors of multimorbidity would be expected to be relatively more influential with increasing age. Another limitation is that in Swedish registers, information on ethnic groups is not available. To reduce the number of incorrect personal numbers because of immigration,[18] only Swedish born individuals with Swedish born parents were included, limiting the generalisability of the study.

Conclusions

The results of our study suggest that the risk of multimorbidity among relatives of affected individuals depends on the closeness of the relationship, being strongest in twins and full siblings, but still significant in third degree relatives, indicating genetic components in susceptibility to multimorbidity. We cannot rule out familial lifestyle factors contributing to the familial aggregation of multimorbidity. Principal component analysis and factor analysis were used to identify smaller disease clusters. This explorative family study suggests that map clusters of diseases should be used for the genetic study of common diseases to show new genetic patterns of non-communicable diseases.

Acknowledgements We thank Patrick O'Reilly, science editor at the Centre for Primary Health Care Research, Department of Clinical Sciences, Lund University, and Region Skåne, Malmö, Sweden, for his comments on the language in the text. The registers used in this study are maintained by Statistics Sweden and the National Board of Health and Welfare.

Contributors All authors contributed to the conception and design of the study. JS and KS contributed to the acquisition of data. All authors contributed to the analysis and interpretation of the data. BZ drafted the manuscript. All authors revised the manuscript critically and approved the final version. All authors had full access to all of the data (including statistical reports and tables) and take responsibility for the integrity of the data and the accuracy of their analysis. BZ is the guarantor of the study. The corresponding author (BZ) attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. Transparency: The lead author (the guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Funding The study was funded by ALF-funding from Region Skåne, the Swedish Research Council (2020-01824), and Sparbanken Skåne. This work was supported by grants awarded to BZ and KS by ALF-funding from Region Skåne, and grants awarded to BZ by the Swedish Research Council and Sparbanken Skåne. The funders had no role in considering the study design or in the collection, analysis, interpretation of data, writing of the report, or decision to submit the article for publication.

Competing interests All authors have completed the ICMJE uniform disclosure form at www.icmje.org/disclosure-of-interest/ and declare: support from ALF-funding, the Swedish Research Council, and Sparbanken Skåne for the submitted work; BZ and KS report receiving grants from ALF-funding from Region Skåne for the submitted work; BZ reports receiving grants from the Swedish Research Council and Sparbanken Skåne for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethics approval This study was approved by the regional ethical review board, Lund, Sweden (registration No 2012/795) and later amendments. Guidelines of the Helsinki Declaration were followed. Data were obtained from the national registers collected by various Swedish health and social agencies. The agency Statistics Sweden linked the (pseudonymised) unique personal identification number assigned to all Swedish residents at birth or migration. According to current Swedish regulations, the use of national data for research purposes does not require informed consent from individuals whose data are held in these registers.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. The two Swedish government agencies, Statistics Sweden and the National Board of Health and Welfare, provided all data. Request to use the data can be applied for from the Swedish National Board of Health and Welfare (https://www.socialstyrelsen.se/en/statistics-and-data/statistics/statistical-database/) and Statistics Sweden (https://www.scb.se/en/About-us/contact-us/)

ORCID iDs

Bengt Zöller http://orcid.org/0000-0002-8250-5613
Björn Holmquist http://orcid.org/0000-0003-4316-2483
Kristina Sundquist http://orcid.org/0000-0001-8031-279X

REFERENCES

1 Pearson-Stuttard J, Ezzati M, Gregg EW. Multimorbidity-a defining challenge for health systems. Lancet Public Health 2019;4:e599–600. 10.1016/S2468-2667(19)30222-1
2 Wallace E, Salisbury C, Guthrie B, *et al*. Managing patients with multimorbidity in primary care. BMJ 2015;350:h176. 10.1136/bmj.h176
3 Gijsen R, Hoeymans N, Schellevis FG, *et al*. Causes and consequences of comorbidity: a review. J Clin Epidemiol 2001;54:661–74. 10.1016/s0895-4356(00)00363-2
4 Barnett K, Mercer SW, Norbury M, *et al*. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet 2012;380:37–43. 10.1016/S0140-6736(12)60240-2
5 Banerjee A, Hurst J, Fottrell E, *et al*. Multimorbidity: not just for the West. Glob Heart 2020;15:45. 10.5334/gh.835
6 Navickas R, Petric V-K, Feigl AB, *et al*. Multimorbidity: What do we know? What should we do? J Comorb 2016;6:4–11. 10.15256/joc.2016.6.72
7 Whitty CJM, Watt FM. Map clusters of diseases to tackle multimorbidity. Nature 2020;579:494–6. 10.1038/d41586-020-00837-4
8 Burton PR, Tobin MD, Hopper JL. Key concepts in genetic epidemiology. Lancet 2005;366:941–51. 10.1016/S0140-6736(05)67322-9
9 Visscher PM, Yang J. A plethora of pleiotropy across complex traits. Nat Genet 2016;48:707–8. 10.1038/ng.3604
10 23andMe Research Team, Social Science Genetic Association Consortium, Turley P, *et al*. Multi-trait analysis of genome-wide association summary statistics using MTAG. Nat Genet 2018;50:229–37. 10.1038/s41588-017-0009-4
11 Amell A, Roso-Llorach A, Palomero L, *et al*. Disease networks identify specific conditions and pleiotropy influencing multimorbidity in the general population. Sci Rep 2018;8:15970. 10.1038/s41598-018-34361-3
12 Dong G, Feng J, Sun F, *et al*. A global overview of genetically interpretable multimorbidities among common diseases in the UK Biobank. Genome Med 2021;13:110. 10.1186/s13073-021-00927-6
13 Ekbom A. The Swedish multi-generation register. Methods Mol Biol 2011;675:215–20. 10.1007/978-1-59745-423-0_10
14 Rosen M, Hakulinen T. Use of disease registers. In: Ahrens W, Pigeot I, eds. Handbook of epidemiology. Berlin: Springer-Verlag, 2005: 231–52. 10.1007/978-3-540-26577-1
15 Ludvigsson JF, Andersson E, Ekbom A, *et al*. External review and validation of the Swedish National inpatient register. BMC Public Health 2011;11:450. 10.1186/1471-2458-11-450
16 Zöller B. Nationwide family studies of cardiovascular diseases – clinical and genetic implications of family history. EMJ Cardiol 2013;1:102–13. 10.33590/emjcardiol/10312042
17 Ludvigsson JF, Almqvist C, Bonamy A-KE, *et al*. Registers of the Swedish total population and their use in medical research. Eur J Epidemiol 2016;31:125–36. 10.1007/s10654-016-0117-y
18 Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, *et al*. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. Eur J Epidemiol 2009;24:659–67. 10.1007/s10654-009-9350-y
19 Zöller B, Ohlsson H, Sundquist J, *et al*. Familial risk of venous thromboembolism in first-, second- and third-degree relatives: a nationwide family study in Sweden. Thromb Haemost 2013;109:458–63. 10.1160/TH12-10-0743
20 Waehrens R, Ohlsson H, Sundquist J, *et al*. Risk of irritable bowel syndrome in first-degree, second-degree and third-degree relatives of affected individuals: a nationwide family study in Sweden. Gut 2015;64:215–21. 10.1136/gutjnl-2013-305705
21 Zöller B, Svensson PJ, Sundquist J, *et al*. Postoperative joint replacement complications in Swedish patients with a family history of venous thromboembolism. JAMA Netw Open 2018;1:e181924. 10.1001/jamanetworkopen.2018.1924
22 Kohler HP, Rodgers JL. DF-analyses of heritability with double-entry twin data: asymptotic standard errors and efficient estimation. Behav Genet 2001;31:179–91. 10.1023/A:1010253411274
23 Hindsberger C, Bryld LE. Analysis of twin data ascertained through probands: the double-entry approach. Genet Epidemiol 2003;25:225–35. 10.1002/gepi.10253
24 Thomas DC. Statistical methods in genetic epidemiology. New York: Oxford University Press, 2004.
25 McLoughlin G, Palmer JA, Rijsdijk F, *et al*. Genetic overlap between evoked frontocentral theta-band phase variability, reaction time variability, and attention-deficit/hyperactivity disorder symptoms in a twin study. Biol Psychiatry 2014;75:238–47. 10.1016/j.biopsych.2013.07.020
26 Schober P, Vetter TR. Missing data and imputation methods. Anesth Analg 2020;131:1419–20. 10.1213/ANE.0000000000005068
27 VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. Ann Intern Med 2017;167:268–74. 10.7326/M16-2607
28 Tinkler KJ. The physical interpretation of eigenfunctions of dichotomous matrices. Transactions of the Institute of British Geographers 1972;55:17. 10.2307/621721
29 Sainani KL. Introduction to principal components analysis. PM&R 2014;6:275–8. 10.1016/j.pmrj.2014.02.001
30 Khoury MJ, Beaty TH, Liang KY. Can familial aggregation of disease be explained by familial aggregation of environmental risk factors? Am J Epidemiol 1988;127:674–83. 10.1093/oxfordjournals.aje.a114842
31 Lander ES, Schork NJ. Genetic dissection of complex traits. Science 1994;265:2037–48. 10.1126/science.8091226
32 Park J, Lee D-S, Christakis NA, *et al*. The impact of cellular networks on disease comorbidity. Mol Syst Biol 2009;5:262. 10.1038/msb.2009.16
33 Melamed RD, Emmett KJ, Madubata C, *et al*. Genetic similarity between cancers and comorbid mendelian diseases identifies candidate driver genes. Nat Commun 2015;6:7033. 10.1038/ncomms8033
34 Global Burden of Metabolic Risk Factors for Chronic Diseases Collaboration. Cardiovascular disease, chronic kidney disease, and diabetes mortality burden of cardiometabolic risk factors from 1980 to 2010: a comparative risk assessment. Lancet Diabetes Endocrinol 2014;2:634–47. 10.1016/S2213-8587(14)70102-0
35 Assanangkornchai S, Edwards JG. Clinical and epidemiological assessment of substance misuse and psychiatric comorbidity. Curr Opin Psychiatry 2012;25:187–93. 10.1097/YCO.0b013e3283523d27
36 Busija L, Lim K, Szoeke C, *et al*. Do replicable profiles of multimorbidity exist? Systematic review and synthesis. Eur J Epidemiol 2019;34:1025–53. 10.1007/s10654-019-00568-5
37 DeFilippis EM, Kumar S. Clinical presentation and outcomes of autoimmune hepatitis in inflammatory bowel disease. Dig Dis Sci 2015;60:2873–80. 10.1007/s10620-015-3699-4
38 Heward J, Gough SC. Genetic susceptibility to the development of autoimmune disease. Clin Sci (Lond) 1997;93:479–91. 10.1042/cs0930479
39 Burgess JA, Lowe AJ, Matheson MC, *et al*. Does eczema lead to asthma? J Asthma 2009;46:429–36. 10.1080/02770900902846356
40 Huang Y-C, Wu M-C, Wang Y-H, *et al*. The influence of constipation on asthma: a real-world, population-based cohort study. Int J Clin Pract 2021;75:e14540. 10.1111/ijcp.14540
41 Wang L, Deng Z-R, Zu M-D, *et al*. The comorbid relationship between migraine and asthma: a systematic review and meta-analysis of population-based studies. Front Med (Lausanne) 2020;7:609528. 10.3389/fmed.2020.609528
42 Smith R. Corticosteroids and osteoporosis. Thorax 1990;45:573–8. 10.1136/thx.45.8.573
43 Khosravipour M, Rajati F. Sensorineural hearing loss and risk of stroke: a systematic review and meta-analysis. Sci Rep 2021;11:11021. 10.1038/s41598-021-89695-2
44 Yang L, Fang Q, Zhou L, *et al*. Hearing loss is associated with increased risk of incident stroke but not coronary heart disease among middle-aged and older Chinese adults: the Dongfeng-Tongji cohort study. Environ Sci Pollut Res 2022;29:21198–209. 10.1007/s11356-021-17324-6
45 Weinstock MA, Brodsky GL. Bias in the assessment of family history of melanoma and its association with dysplastic nevi in a case-control study. J Clin Epidemiol 1998;51:1299–303. 10.1016/s0895-4356(98)00070-5
46 Ochieng BMN. Factors affecting choice of a healthy lifestyle: implications for nurses. Br J Community Nurs 2006;11:78–81. 10.12968/bjcn.2006.11.2.20445
47 Celebi ME, Kingravi HA, Vela PA. A comparative study of efficient Initialization methods for the K-means clustering algorithm.

Expert Systems with Applications 2013;40:200–10. 10.1016/j.eswa.2012.07.021
48 Lezhnina O, Kismihók G. Latent class cluster analysis: selecting the number of clusters. MethodsX 2022;9:101747. 10.1016/j.mex.2022.101747
49 Ng SK, Tawiah R, Sawyer M, *et al.* Patterns of multimorbid health conditions: a systematic review of analytical methods and comparison analysis. Int J Epidemiol 2018;47:1687–704. 10.1093/ije/dyy134

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjmed-2021-000070).