

Decoding multimorbidity: Understanding cancers, inflammatory chronic diseases contribution to death.

Kelly Larissa Vomo Donfack^{1,2, 3}, Burak Gönüllü⁴, Frédéric Pamoukjian^{3,5}, Eurydice Angeli^{3,6}, Grégory Ginot¹, Ian Morilla^{1,2#}, Guilhem Bousquet^{3,6##}, Géraldine Falgarone^{3,7#},

¹Université Sorbonne Paris Nord, LAGA, CNRS, UMR 7539, Laboratoire d'excellence Infibrex, Villetaneuse, France

²Instituto de Hortofruticultura Subtropical y Mediterránea “La Mayora” (IHSM), UMA-CSIC, Málaga, Spain

³Unité de Recherche SynKoMIC, Université Sorbonne Paris Nord, Bobigny, Fr-93017 France

⁴Faculty of Medicine, Yeditepe University, Istanbul 34755, Turkey

⁵Hôpital Avicenne, APHP, Service de Gériatrie, Bobigny, Fr-93009, France

⁶Hôpital Avicenne, APHP, Service d'Oncologie médicale, Bobigny, FR-93009, France

⁷Hôpital Avicenne, APHP, Unité de Médecine Ambulatoire (UMA), Bobigny, FR-93009 France

[#]co-senior authorship

*Correspondence: Professor Guilhem Bousquet, Université Sorbonne Paris Nord, UFR SMBH, 1 rue Chablis 93000 Bobigny. E-mail: (guilhem.bousquet@aphp.fr), ORCID: 0000-0001-5594-6694. Telephone: +33681845379

The authors have declared that no conflict of interest exists

Abstract

Importance

Thanks to Structural equation modeling (SEM), showing importance of complex interaction among genetic predisposition, chronic diseases and aging-related inflammation that influences cancer outcome and/or death, why these relationships are difficult to study using conventional statistical methods.

Objective

Construct and validate latent variables representing inflammaging, familial cancer risk, familial vascular risk, and early chronic disease, and to evaluate their relationships with cancer severity, metastasis, and mortality.

Design, Setting, and Participants

This prospective, monocentric cohort study includes 677 adult patients with cancer enrolled between January 2016 and December 2022. Patients were followed for a median of 2.14 years and subsequently 3.03 years. SEM was used to analyze relationships among clinical, behavioral, and familial variables. Sex-stratified analyses were also performed.

Exposures

Sex, smoking history, body mass index (BMI), personal and familial history of cancer, autoimmune diseases, cardiovascular diseases, and cardiovascular risk factors were collected.

Main Outcomes and Measures

Cancer severity (early or multiple cancers), metastatic disease, and mortality.

Results

Four latent variables—Inflammaging, Familial Cancer Risk, Familial Vascular Risk, and Early Chronic Disease—showed good statistical reliability and model fit (comparative fit index [CFI] up to 0.926; root mean square error of approximation (RMSEA) as low as 0.043). Inflammaging, defined using sex, smoking, and BMI, was strongly associated with cancer (loading, 0.80) and significantly associated with metastatic disease and mortality at shorter follow-up. Familial Vascular Risk was associated with Familial Cancer Risk, which in turn was associated with Inflammaging. Sex-specific analyses revealed distinct pathways in women, in whom familial risk and early chronic disease directly contributed to metastatic disease and death, while Inflammaging was directly associated with mortality.

Conclusions and Relevance

Latent clinical constructs integrating familial risk, chronic disease, and aging-related factors provide a coherent framework for understanding cancer outcomes. Marked sex-specific differences suggest that genetic predisposition and chronic disease burden play a central role in cancer progression in women.

Introduction

With the population ageing, cancers and cardiovascular diseases are the main causes of mortality worldwide and major public health concerns, requiring to better understand their pathophysiology and links, beyond cardiotoxicity, to identify new therapeutic targets (1).

The co-occurrence of 2 or more chronic conditions in the same individual, named multimorbidity, is the dominant pattern of disease in ageing populations and is associated with reduced quality of life, treatment burden, and premature mortality. Beyond simple disease counts, multimorbidity reflects dynamic interactions among genetic susceptibility, lifetime environmental exposures, and social determinants of health. A recent viewpoint characterized multimorbidity as a complex, person-level challenge that demands integrated models able to capture shared risk factors, disease clustering, and heterogeneous trajectories over the life course (2).

Most diseases are the product of a balance between constitutional genetic mutations and gene expression modulation through life due to environmental factor exposure while complex genetic traits are frequent with an association of low penetrance gene (3,4), making it difficult to identify.

Using an individual genetic approach combined with an original transgenic mouse model, our research team has demonstrated a genetic link between malignancies and auto-immune diseases (5). Yet, the biological demonstration of the functional value of most genetic variants remains a considerable challenge, particularly when polygeny is suspected.

Traditional statistical approaches often struggle to capture the complex, interdependent nature of these relationships, where biological pathways intersect and influence one another through both direct and indirect effects. This underscores the need for advanced analytical approaches,

especially in multimorbidity to understand latent factors. Structural Equation Modeling (SEM) is a powerful multivariate statistical tool that allows researchers to examine complex relationships among observed and latent variables (6–9).

In the context of aging and to understand the links between cancers and chronic diseases, we hypothesized that the balance between constitutional and acquired genetic abnormalities could be model into two latent variables called “constitutional genetics” and “inflammaging” (Fig.1). Inflammaging refers to low-level chronic inflammation linked to ageing and leading to gene mutations, cell senescence, and immunity dysfunction associated with most aging-related diseases (10). It is also favored by age-independent exposition factors such as cigarette smoke or pollution (11–13), and by other factors like obesity (14). Since there is data suggesting a genetic determinism of age-related biological hallmarks (15,16), we assumed that constitutional genetics not only influence all diseases but the inflammaging itself.

It is unknown whether inflammaging can be validly operationalized as a latent clinical construct using routinely collected variables rather than biological biomarkers, and whether such a construct meaningfully contributes to cancer severity and mortality. Moreover, familial cancer risk and familial vascular risk are generally studied independently, and their potential interdependence in shaping cancer outcomes has not been formally tested. Early-onset autoimmune and cardiovascular diseases are typically treated as confounders rather than as indicators of constitutional vulnerability, and it remains unclear whether these early chronic diseases constitute a distinct latent domain. Finally, whether the structural relationships linking familial risk, chronic disease, and inflammaging to cancer outcomes differ by sex has not been systematically evaluated.

Using data from a unique and rare cohort of cancer patients displaying severe and/or early onset diseases, we used SEM’s ability to test theoretical models and validate causal hypotheses to decipher the interconnected pathophysiological mechanisms driving these

conditions. Indeed, we took advantage of the expression of severe and chronic diseases in other members of the family of each index-case cancer patient to estimate the impact of chronic diseases expressed in relatives of next generations.

Methods

As this is a cohort of patients with all types of cancers and other pathologies, both men and women were included. Insofar as there are known differences in the prevalence of diseases between the two sexes, particularly cardiovascular and autoimmune diseases, we chose to use sex as a biological variable in our analysis and in one of the SEM models.

A monocentric prospective cohort

In this monocentric study, patients were prospectively and consecutively included in registry from January 2016 to December 2022 in Avicenne hospital (APHP, France). All patients were referred to a medical oncologist (GB) for a cancer diagnosis, and underwent comprehensive clinical evaluation. At the time of the first consultation, personal and family history data intended to be as exhaustive as possible. Genealogical trees were built for relevant cases (see examples in Supplementary Fig.1).

For cancer diagnosis, we considered the age cut-off less than 50 years for individual or familial genetic predisposition risk (17–19). Cancer types were defined as individual categories when they represented more than 5% of the total cohort.

Data collected

The following cardiovascular diseases (CVD) were considered: stroke, myocardial infarction/acute coronary syndrome, venous thromboembolism, arrhythmia, peripheral artery disease, cardiomyopathy of any origin. We retained the cut-off of 50 years to define early events and consider a genetic predisposition risk (20,21), this cut-off representing usually less than 10% of events in a global population (22). Cardiovascular risk factors (CVFR) included

arterial hypertension, obesity, smoking, dyslipidemia, and diabetes. Smoking, determined by multiplying the number of cigarettes smoked per day by the number of years of smoking, was categorized into no smoking (smoking index = 0), moderate smoking ($0 < \text{smoking} < 10$), and severe smoking (smoking ≥ 10). Obesity was assessed by the body mass index (BMI). Here, we considered BMI as a binary variable: ≤ 40 and $> 40 \text{ kg/m}^2$ which corresponds to people with very high BMI where excess adiposity is assumed (23).

Auto-immune disease (AID) was defined as severe when at least one complication occurred, and as an early form when it occurred before 30 years for Hashimoto's thyroiditis (24), before 20 years for Crohn disease(25), before 40 years for rheumatoid arthritis (26,27), before 25 years for ankylosing spondylitis (28), and before 25 years for systemic lupus erythematosus (29).

For Familial Vascular Risk, a first-degree relative includes a parent, a sibling, or a child while the second-degree relative includes an aunt, an uncle, a grandparent, a grandchild, a niece, a nephew or a half-sibling of the patient.

According to French law, all patients were informed of the use of their clinical data collected during routine care and did not oppose to it.

Data analysis

A conceptual model was developed from literature and personal analysis before data analysis. We used Python for data cleaning, recodes, and descriptive analysis, and STATA for all model construction and testing.

Presentation of the dataset

The dataset consists of 677 entries and 28 variables, including *age*, *sex*, *BMI*, *cancer site*, as well as self-reported and transgenerational information on cancer, autoimmune diseases, cardiovascular diseases, and mortality.

Data cleaning and processing

Before beginning the analysis, we performed several data preprocessing steps: i) Handling missing values, observed in columns such as *smoking*, *age at first cancer diagnosis* and *early or multiple cancer* (Suppl.Fig.2). We imputed the missing values using iterative imputer; ii) data transformation: to enhance model performance, we standardized all continuous variables, including *age*, *smoking status*, *BMI*, and *age at first cancer diagnosis*. No outlier was observed in the data set.

Exploratory Data Analysis

Categorical variables were summarized using counts and percentages, while continuous variables were summarized either as the mean \pm standard deviation (SD) or as the median with the interquartile range (IQR).

Correlations: We created a correlation matrix with observed variables using Cramer's V test.

Confirmatory factor analysis (CFA): latent variables were included in the CFA. To obtain estimates, standard errors, and standardized coefficients, we used the maximum likelihood.

Model modification strategies

Throughout the model building process, we evaluated various models and modifications to ensure both statistical and conceptual fit. We performed a confirmatory factor analysis to affirm the observed variables for each latent variable. In our model analysis, we chose the maximum likelihood chi-square estimation method. Model goodness of fit was assessed using chi-square, comparative fit index (CFI), the root mean square error of approximation (RMSEA), standardized root mean square residual (SRMSR), and the maximum likelihood. The cutoff criteria for the model goodness of fit were as follows: chi-square score with *P*-value>0.05, CFI>0.90, RMSEA<0.06, SRMR<0.08 (30).

For more clarity, observed variables were written in italic while latent variables were written in bold italic and capitalized first letter.

Results

Prospective cohort and patients' characteristics

From January 2016 to December 2022, 677 patients have been prospectively included in our monocentric cohort. Their characteristics are detailed in Table 1 and Supp.Table 1. Briefly, 81% were women with a mean age of 63 years. *Smoking* was low, and only 13% of patients smoked more than 10 packet-years at the time of inclusion in this register. The mean *BMI* was 27 ± 6 kg/m² for the whole cohort, 26 patients (4%) with extreme class III obesity (*BMI* > 40). Regarding constitutional genetics of high penetrance, 3% of patients (19/677) had an identified mutation.

The mean age at first cancer diagnosis was 60 ± 17 years, and 18% of patients had at least 2 different cancers. Considering all 711 *cancers*, 30 different types have been included in this cohort, mainly women cancers (74%) (Supp.Table 2). Thirty-three percent of patients had at least two cancers or a first diagnosis before the age of 50 years, namely *early or multiple cancers*. In addition, 53% of patients had a *metastatic disease*, either at initial diagnosis or during follow-up, which is the main characteristic of our cohort of severe cancers.

Sixteen percent of patients also had an AID (autoimmune disease), 23% had a CVD (cardiovascular disease), and 38% had an associated CVRF (cardiovascular risk factor). These pathologies occurred early for 10%, 3%, and 17% of patients, respectively. In addition, 2.5 % of the cohort had a severe autoimmune disease.

Considering familial history of diseases, 14% of patients had a family history of cancer and 12% with *at least two first-degree cancers*. Family histories of other chronic disease were sparser, including 5% of patients with at least one first-degree AID or CVD, and 8% of patients with one first-degree CVRF (Table 1).

When analyses were performed after 2 years or 3 years of median follow-up, 70% and 65% of patients were alive, respectively.

Correlation matrix

To understand the relationship between variables as shown in Fig.2A, we used the Crammer's V test. In particular, the observed variable *early or multiple cancer* was positively correlated with the variable *familial cancer*, so that we used them to feed the same latent variable. In the same way, *early AID* was positively correlated with *severe AID*, and also with *early CVD* and *early CVRF*. Finally, *early familial CVRF* was positively correlated with *familial CVRF*, *early familial CVD* and *familial CVD*.

Latent variable construction

To improve goodness of fit of our initial model (Fig.1), we decided to divide “constitutional genetics” into two variables linked to a familial risk of disease: ***Familial Cancer Risk*** and ***Familial Vascular Risk***. Conversely, early auto-immune and cardiovascular diseases were fused into a unique variable called ***Early Chronic Disease***. We finally kept four latent variables to test their statistical reliability: ***Inflammaging***, ***Familial Cancer Risk***, ***Familial Vascular Risk***, and ***Early Chronic Disease*** (Fig.2A to 2D, Suppl.Table 3).

For each latent variable, a minimum of 5 models has been tested to optimize statistical parameters. Typically, we hypothesized that *sex*, *smoking* and *BMI* should be strongly related to ***Inflammaging*** (Fig.2B). The model was identified as just, meaning that our model perfectly fitted. The CFI and RMSEA were 1 and 0 respectively. In particular, *smoking* was strongly related to the latent variable ***Inflammaging***, with a loading R of 0.81, implying that 65% (R^2) of the variance in smoking was explained by ***Inflammaging***. For *sex*, this link was moderate since 25% of the variance explained ***Inflammaging***. For *BMI*, the link was weak with a loading of 0.04. Overall, 69% of the variance in *smoking*, *BMI* and *sex* was captured by ***Inflammaging***.

Familial Vascular Risk was fed by the observed variables *familial CVD*, *early familial CVD*, *familial CVFR* and *early CVFR*. ***Early Chronic Disease*** was fed using *early AID*, *severe AID*,

early CVD and *early CVFR* as observed variables (Fig.2D), and the observed variables *familial cancer* and *early familial cancer* were used to define **Familial Cancer Risk** (Fig.2E). All four selected latent variables were thus plausible.

Description of the links between cancers, inflammaging and chronic diseases

After a median follow-up of 2.14 years and 30.5% deaths, the first model retained was over-identified with a degree of freedom of 121. The chi-square relative to the degree of freedom statistic was significant, $\chi^2 = 281$, $P < 0.001$ (Supp.Fig.3) and 4 for examples of preliminary models). The model showed a goodness of fit with a RMSEA of 0.044 and a CFI of 0.918. Figure 3A illustrates the corresponding recursive SEM diagram, where we explain the endogenous outcome variable *death* and the relevant pathways leading to death in the context of cancer disease. **Familial Vascular Risk** was significantly associated with **Familial Cancer Risk** which itself was significantly associated with **Inflammaging**. Furthermore, **Inflammaging** was strongly associated with *cancer* (loading factor of 0.8) and notably associated with *metastatic disease* (0.14, $P < 0.035$) and death (0.54, $P < 0.01$).

Interestingly, the latent variable **Early Chronic Disease** was negatively associated with *early AID*, *severe AID*, *early CVD* and *early CVRF* (Supp.Table5). **Inflammaging**, *early or multiple cancers* and **Familial Vascular Risk** were also negatively associated to **Early Chronic Disease**.

With a longer median follow-up of 3.03 years and 5.5% additional dead patients, the performance was slightly better with a CFI of 0.926 and RMSEA of 0.043 (Fig.3B). The latent variable **Inflammaging** was not anymore directly associated with *death* ($P = 0.08$).

We wanted to test if adding a fifth latent variable, **Late Chronic Disease**, loaded with two other observed variables – *at least 2 CVRF* and *at least 1 AID* – corresponding to personal but not familial disease, would improve the model. We separated early diseases (**Early Chronic Disease**) from late diseases (**Late Chronic Disease**) to test the hypothesis of constitutional

genetics influence for the first variable and of environmental factors like smoking for the second one. This model was also plausible (RMSEA of 0.042, CFI of 0.91), but did not add any information since *Inflammaging* was not associated with *Late Chronic disease* (Fig.4A, Supp.Table 4 and 5 for model fit and comparisons).

Women with severe cancer history are prone to develop multiple diseases with marked genetic predisposition

When we considered men and women separately, respectively 127 and 550 patients, we evidenced several differences between the two populations: smoking was 2.5 more frequent in men whereas obesity (BMI >30 kg/m²) was more frequent in women (28% vs. 18%, respectively) (Table 2 and Supp.Table 6). Metastatic diseases were particularly frequent in men compared to women (71% vs. 43%, mostly explained by cancer types, $P<0.01$). As expected, AID were 2.6-times more frequent in women ($P=0.02$) and CVD 1.7-times more frequent in men ($P<0.01$), while early CVRF were unexpectedly more frequent in women (18% vs. 11%). Finally, all familial histories of diseases were more frequent in women, including early familial cancers that occurred 2.7-times more often than in men ($P=0.01$) (Table 2).

When we applied our three SEM models for the woman population, they did not fit. With some ameliorations of the second model, we obtained a goodness of fit with a RMSEA of 0.045 and a CFI of 0.901 (Fig.4B). In particular, the variables *smoking* and *BMI* were then considered as continuous and not categorial variables as it was previously the case. The latent variable *Familial Vascular Risk* was still associated with *Familial Cancer Risk* that was now negatively associated with *Inflammaging*. Conversely, *Familial Vascular Risk* was from now positively associated with *Early Chronic Disease* but also with *early or multiple cancers* in a pathway directly leading to *metastatic disease* and thus to *death*. The latent variable *Inflammaging* was no longer a mediator of death, but directly and significantly associated

with death. In addition, *Inflammaging* was then significantly associated with the very old *age* and negatively with *BMI*, but no longer with *smoking*.

Discussion

This study represents the first successful application of SEM to elucidate the interconnected pathways linking familial vascular risk, inflammaging, cancer, and chronic inflammatory diseases. Biology alone would not permit to define all the existing links between biological processes, highlighting the added-value of complementary *in silico* approaches. SEM provides a powerful tool to navigate into the "black box" of disease interactions. By formalizing latent constructs like *Inflammaging*, *Familial Vascular Risk* and *Early Chronic Disease*, we moved beyond correlation to infer causal links, accounting for measurement error and bidirectional effects that traditional regression models overlook (31–33). Recently, SEM has been applied to genetics data to better decipher the links between complex genetic traits (34). In our study, regardless of genetic data, we showed that familial early and/or severe cardiovascular diseases and cancers, which represent some part of the constitutional genetics, are strongly interconnected. Our results based on SEM reinforce the hypothesis that when severe polypathological associations occur at a young age, there are most likely common constitutional genetic factors that contribute to the development of the different diseases leading to metastatic cancer and to cancer death. To our knowledge, this is the first mathematical demonstration of such links.

The construction of our latent variables allows us to demonstrate their links but also their direct role in the patients' death. We show interactions beyond common pathophysiological knowledge, in particular in the close links between cancers and chronic inflammatory diseases, reinforcing the need for more systematic constitution genetic analyses, in routine care, to identify new therapeutic targets in a personalized medicine framework (35).

Interestingly, our different models illustrate both population heterogeneity and ageing heterogeneity, with singular pathways leading to death among women. For the whole population, the path to death is through *Inflammaging*, while in women it is the polypathological association of severe diseases that promotes metastatic evolution and leads to death, except probably for very old women with low BMI. An explanation why the variable *Inflammaging* is no longer a main mediator towards cancer death in our female population could be the fact that immunosenescence, meaning immune system aging, is less marked in women than in men (36). In very old women with metastatic cancer, the low BMI usually corresponds to cachexia and sarcopenia that are associated with inflammation and an increased risk of specific death (37). In older women with cancer, but not men, obesity has been reported as a protective factor in the absence of significant weight loss (38), which supports the use of continuous variable for obesity (14–23).

The gender disequilibrium partly explains the high frequency of AIDs, but not the frequency of CVDs and CVRFs. The higher incidence of AID in women than in men is well-known, with the possible implication of sex hormones or X chromosome-linked genetics determinants (39). In contrast, premenopausal women are theoretically protected from an increased vascular risk compared to men (20), but also from the physiological processes of aging (40). This vascular risk advantage disappears in case of obesity (20)}, which may be an explanation of the high incidence of CVD and CVRF in our women population since obesity affects 26% of them, and more often than men as usually observed in modern countries (20). It is also acknowledged that there are some disparities in diagnosis and treatment of cardiovascular diseases in women (41). All these risk factors combined, in addition to a certain genetic determinism common to advanced cancers, autoimmune diseases and severe cardiovascular diseases, explains why the path to death found for our woman cohort passes through these polypathological associations favoring the metastatic evolution of cancer. Despite the fact that

biological pathways are still to be fully deciphered, there are growing preclinical evidences that severe cardiovascular diseases favor tumor growth, metastatic spread and specific death (42). To this point of view, our model contributes to demonstrate the preponderant impact of gene in polypathology associations.

One major strength of our study is the fact that patients were prospectively and consecutively included in registry with very few missing data. The systematic collection of the medical family histories, which are hardly exhaustive in individual medical reports when patients are not specifically referred for genetic counselling, offers mathematical power. Patients of our cohort were initially referred to an oncologist, mainly with a severe form since 53% of patients had cancer metastases and were 60 of aged at cancer diagnosis, that is lower than the average age of cancer disease in France. Metastatic disease is the main factor of death in our study, but in general, the proportion of metastatic patients is much lower so that less than 30% of patients die because of their cancer, and even less than 15% when only breast and genital cancers are considered (43). Interestingly, in our study, we observed that early or severe autoimmune diseases and cardiovascular risk factors were also directly associated with mortality, multiple severe diseases altogether contributing to the patient's death.

A limitation of our study is the fact that familial histories were self-reported (as usual in genealogy description) and may be partial or even subjected to bias. We could also improve the definition of the latent variable inflammaging by including, for example, biological markers. Moreover, some associations such as negative relationships involving chronic disease, were unexpected and should be interpreted cautiously. This is possible, because with SEM, results depend on model specification, and alternative plausible model can not be excluded. There was also a lack of biological data for constitutional genetics, since less than 3% of patients in our series had an identified mutation in high penetrance genes, much lower than the prevalence found in high-

risk populations (44) or when using whole-genome sequencing on constitutional DNA (45). In fact, the proportion of patients who were referred to a genetic council was small due to cost issues, with a limited number of genes tested. Considering the weight of familial factors in our SEM approach, such research should be considered much more often and with larger gene panels.

In conclusion, in women with an association of severe cancer and chronic inflammatory disease, regardless of age, our results support the hypothesis of a significant cancer progression and that is link to genetic progression and we think it is associated with the weight of low-to moderate penetrance constitutional genetics, and the interest of more systematic genetic analyses for the understanding of common pathophysiology's, the identification of new therapeutic targets to improve their survival.

Data availability

De-identified data supporting this study are available from the corresponding author upon reasonable request, pending ethical approvals.

Authors' contributions

Conception and design: GB, GF, IM, GG

Development of methodology: KLVD, FP

Acquisition of data: GB, EA, BG

Analyses and interpretation of data: KLVD, GB, GF, FP

Drafting, review: KLVD, GB, GF, IM, FP

Revision of the manuscript: All authors

Material support: GB, GG

Administrative and technical: GB, GG

Study supervision: GB, GF, IM

Funding support

No external funding. Kelly Larissa Vomo Donfack received a Ph.D. Grant from Sorbonne Paris Nord University.

References

1. Molnár AÁ, Birgés K, Surman A, Merkely B. The Complex Connection Between Myocardial Dysfunction and Cancer Beyond Cardiotoxicity: Shared Risk Factors and Common Molecular Pathways. *Int J Mol Sci*. 2024 Dec 8;25(23):13185.
2. eBioMedicine. Multimorbidity: a complex challenge to support and optimise individual health. *eBioMedicine*. 2024 Jan;99:104973.
3. Hamdan D, Nguyen VT, Paris J, Leboeuf C, El Bouchtaoui M, Espié M, et al. The germline genetics of mild-to-moderate penetrance: An intriguing role of PRAME in multiple carcinogenesis. *Genes Dis*. 2024 May 1;11(3):101003.
4. Weale ME, Riveros-Mckay F, Selzam S, Seth P, Moore R, Tarran WA, et al. Validation of an Integrated Risk Tool, Including Polygenic Risk Score, for Atherosclerotic Cardiovascular Disease in Multiple Ethnicities and Ancestries. *Am J Cardiol*. 2021 Jun 1;148:157–64.
5. El Bouchtaoui M, Do Cruzeiro M, Leboeuf C, Loisel-Ferreira I, Fedronie C, Ferreira C, et al. A Constitutional Activating MET Mutation Makes the Genetic Link between Malignancies and Chronic Inflammatory Diseases. *Clin Cancer Res*. 2019 Jul 15;25(14):4504–15.
6. Allen C, Waters EA, Hamilton JG, Vu M, Gabriel J, Roberts MC. Multifactorial causal beliefs and colorectal cancer screening: A structural equation modeling investigation. *J Health Psychol*. 2022 Sep;27(11):2463–77.
7. Bountress KE, Cusack SE, Hawn SE, Grotzinger A, Bustamante D, Kirkpatrick RM, et al. Genetic associations between alcohol phenotypes and life satisfaction: a genomic structural equation modelling approach. *Sci Rep*. 2023 Aug 18;13:13443.
8. Pamoukdjian F, Laurent M, Martinez-Tapia C, Rolland Y, Paillaud E, Canoui-Poitaine F. Frailty Parameters, Morbidity and Mortality in Older Adults with Cancer: A Structural Equation Modelling Approach Based on the Fried Phenotype. *J Clin Med*. 2020 Jun 11;9(6):1826.
9. Schreiber JB. Issues and recommendations for exploratory factor analysis and principal component analysis. *Res Soc Adm Pharm*. 2021 May 1;17(5):1004–11.
10. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. *Cell*. 2013 Jun 6;153(6):1194–217.
11. Babizhayev MA, Savel'yeva EL, Moskvina SN, Yegorov YE. Telomere Length is a Biomarker of Cumulative Oxidative Stress, Biologic Age, and an Independent Predictor of Survival and Therapeutic Treatment Requirement Associated With Smoking Behavior. *Am J Ther*. 2011 Nov;18(6):e209–26.
12. MacNee W. Aging, inflammation, and emphysema. *Am J Respir Crit Care Med*. 2011 Dec 15;184(12):1327–9.
13. Zhao B, Vo HQ, Johnston FH, Negishi K. Air pollution and telomere length: a systematic review of 12,058 subjects. *Cardiovasc Diagn Ther*. 2018 Aug;8(4):480–92.
14. Franceschi C, Garagnani P, Parini P, Giuliani C, Santoro A. Inflammaging: a new immune–metabolic viewpoint for age-related diseases. *Nat Rev Endocrinol*. 2018 Oct;14(10):576–90.
15. Clark J, Reddy S, Zheng K, Betensky RA, Simon DK. Association of PGC-1α polymorphisms with age of onset and risk of Parkinson's disease. *BMC Med Genet*. 2011 May 19;12:69.
16. Codd V, Nelson CP, Albrecht E, Mangino M, Deelen J, Buxton JL, et al. Identification of seven loci affecting mean telomere length and their association with disease. *Nat Genet*. 2013 Apr;45(4):422–427e2.
17. Kiser D, Elhanan G, Bolze A, Neveux I, Schlauch KA, Metcalf WJ, et al. Screening Familial Risk for Hereditary Breast and Ovarian Cancer. *JAMA Netw Open*. 2024 Sep 25;7(9):e2435901.
18. Mauri G, Sartore-Bianchi A, Russo A, Marsoni S, Bardelli A, Siena S. Early-onset colorectal cancer in young individuals. *Mol Oncol*. 2019 Feb;13(2):109–31.
19. Porter N, Laheru D, Lau B, He J, Zheng L, Narang A, et al. Risk of Pancreatic Cancer in the Long-Term Prospective Follow-Up of Familial Pancreatic Cancer Kindreds. *JNCI J Natl Cancer Inst*. 2022 Aug 27;114(12):1681–8.
20. Colafella KMM, Denton KM. Sex-specific differences in hypertension and associated cardiovascular disease. *Nat Rev Nephrol*. 2018 Mar;14(3):185–201.

- 21.Vasan RS, Song RJ, van den Heuvel ER. Temporal Trends in Incidence of Premature Cardiovascular Disease Over the Past 7 Decades: The Framingham Heart Study. *J Am Heart Assoc Cardiovasc Cerebrovasc Dis.* 2022 Sep 29;11(19):e026497.
- 22.Conrad N, Molenberghs G, Verbeke G, Zaccardi F, Lawson C, Friday JM, et al. Trends in cardiovascular disease incidence among 22 million people in the UK over 20 years: population based study. *The BMJ.* 2024 Jun 26;385:e078523.
- 23.Rubino F, Cummings DE, Eckel RH, Cohen RV, Wilding JPH, Brown WA, et al. Definition and diagnostic criteria of clinical obesity. *Lancet Diabetes Endocrinol.* 2025 Mar 1;13(3):221–62.
- 24.Kaur J, Vadakekut ES. Hashimoto Thyroiditis. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 [cited 2026 Jan 14]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK459262/>
- 25.Shivashankar R, Tremaine WJ, Harmsen WS, Loftus EV. Incidence and Prevalence of Crohn’s Disease and Ulcerative Colitis in Olmsted County, Minnesota From 1970 Through 2010. *Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc.* 2017 Jun;15(6):857–63.
- 26.Eriksson JK, Neovius M, Ernestam S, Lindblad S, Simard JF, Askling J. Incidence of rheumatoid arthritis in Sweden: a nationwide population-based assessment of incidence, its determinants, and treatment penetration. *Arthritis Care Res.* 2013 Jun;65(6):870–8.
- 27.Myasoedova E, Crowson CS, Kremers HM, Thernau TM, Gabriel SE. Is the incidence of rheumatoid arthritis rising? Results from Olmsted County, Minnesota, 1955–2007. *Arthritis Rheum.* 2010 Jun;62(6):1576–82.
- 28.Boel A, López-Medina C, van der Heijde DMFM, van Gaalen FA. Age at onset in axial spondyloarthritis around the world: data from the Assessment in SpondyloArthritis international Society Peripheral Involvement in Spondyloarthritis study. *Rheumatol Oxf Engl.* 2022 Apr 11;61(4):1468–75.
- 29.Fatoye F, Gebrye T, Mbada C. Global and regional prevalence and incidence of systemic lupus erythematosus in low-and-middle income countries: a systematic review and meta-analysis. *Rheumatol Int.* 2022;42(12):2097–107.
- 30.Alan C. Acock. Stata Bookstore: Discovering Structural Equation Modeling Using Stata, Revised Edition [Internet]. Stata Press; 2013 [cited 2026 Jan 14]. Available from: <https://www.stata.com/bookstore/discovering-structural-equation-modeling-using-stata/>
- 31.Minoura A, Sugiyama T, Koyama T, Yoshioka T, Tabuchi T. Structural equation modeling of the effects of psychological distress and a fear of coronavirus disease 2019 on diabetes care in Japan: a cross-sectional study. *Sci Rep.* 2022 Sep 27;12(1):16142.
- 32.Schreiber JB. Core reporting practices in structural equation modeling. *Res Soc Adm Pharm.* 2008 Jun;4(2):83–97.
- 33.EGG Consortium, Warrington NM, Beaumont RN, Horikoshi M, Day FR, Helgeland Ø, et al. Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat Genet.* 2019 May;51(5):804–14.
- 34.Grotzinger AD, Rhemtulla M, De Vlaming R, Ritchie SJ, Mallard TT, Hill WD, et al. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat Hum Behav.* 2019 Apr 8;3(5):513–25.
- 35.Hamdan D, Marisa L, Tlemsani C, Angeli E, Soussan M, Derive N, et al. Olaparib in the Setting of Radiotherapy-Associated Sarcoma: What Can Precision Medicine Offer For Rare Cancers? *JCO Precis Oncol.* 2023 Feb;7:e2200582.
- 36.Calabrò A, Accardi G, Aiello A, Caruso C, Candore G. Sex and gender affect immune aging. *Front Aging.* 2023 Nov 28;4:1272118.
- 37.Shachar SS, Williams GR, Muss HB, Nishijima TF. Prognostic value of sarcopenia in adults with solid tumours: A meta-analysis and systematic review. *Eur J Cancer.* 2016 Apr;57:58–67.
- 38.Martinez-Tapia C, Diot T, Oubaya N, Paillaud E, Poisson J, Gisselbrecht M, et al. The obesity paradox for mid- and long-term mortality in older cancer patients: a prospective multicenter cohort study. *Am J Clin Nutr.* 2021 Jan;113(1):129–41.
- 39.E O, M P, A M, C V, F A, Y S. Sex-based differences in autoimmune diseases. *Ann Ist Super Sanita* [Internet]. 2016 Jun [cited 2026 Jan 14];52(2). Available from: <https://pubmed.ncbi.nlm.nih.gov/27364395/>

- 40.Lopez-Lee C, Torres ERS, Carling G, Gan L. Mechanisms of sex differences in Alzheimer's disease. *Neuron*. 2024 Apr;112(8):1208–21.
- 41.Vogel B, Acevedo M, Appelman Y, Bairey Merz CN, Chieffo A, Figtree GA, et al. The Lancet women and cardiovascular disease Commission: reducing the global burden by 2030. *The Lancet*. 2021 Jun;397(10292):2385–438.
- 42.Newman AAC, Dalman JM, Moore KJ. Cardiovascular Disease and Cancer: A Dangerous Liaison. *Arterioscler Thromb Vasc Biol*. 2025 Mar;45(3):359–71.
- 43.Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. *CA Cancer J Clin*. 2024 Jan;74(1):12–49.
- 44.Pearlman R, Frankel WL, Swanson B, Zhao W, Yilmaz A, Miller K, et al. Prevalence and Spectrum of Germline Cancer Susceptibility Gene Mutations Among Patients With Early-Onset Colorectal Cancer. *JAMA Oncol*. 2017 Apr 1;3(4):464.
- 45.Mandelker D, Donoghue M, Talukdar S, Bandlamudi C, Srinivasan P, Vivek M, et al. Germline-focussed analysis of tumour-only sequencing: recommendations from the ESMO Precision Medicine Working Group. *Ann Oncol*. 2019 Aug;30(8):1221–31.

Figure 1: Theoretical model with latent variables (orange disks) and observed variables (green rectangles).

Figure 2: Construction of latent variables. A panel shows the correlation matrix between the different variables using Crammer's V test. B, C, D and E panels shows the four latent variables (blue disks) selected for the final models, with the respective observed variables (green rectangles) and with the corresponding global statistical parameters. β coefficients have been transformed as odd-ratios.

BMI: body mass index; AID: autoimmune disease; CVRF: cardiovascular risk factors; CVD: cardiovascular disease; RMSEA: root mean square error of approximation; CFI: comparative fit index; TLI: Tucker-Lewis indice; SRMR: standardized root mean square residual.

* $P < 0.05$, ** $P < 0.01$.

Figure 3: Schematic representation of models 1 (A) and 2 (B), after a median follow-up of 2 years and 3 years, respectively. Age is an exogenous variable feeding the latent inflammaging variable. The significant variations between the two models are highlighted in red.

* $P < 0.05$, ** $P < 0.01$.

Figure 4: Schematic representation of models 3 (A) and 4 (B), both performed using data obtained after a median follow-up of 3 years. Model 3 includes a fifth latent variable, namely **Late Chronic Disease**. Model 4 includes only women. The loading factors that are significantly modified when compared with model 1 are written in red.

Table 1: Baseline characteristics of the study population (n = 677)

Variables	Mean \pm SD or N (%)
Age (y)	62.5 \pm 17.0
Sex (women)	550 (81.2)
Smoking (pack /year)	6.2 \pm 14.0
No smoking	479 (71)
< 10 pack/year	106 (16)
\geq 10 pack/year	92 (13)
Body Mass Index (BMI)	27.2 \pm 5.7
< 19	25 (4)
19 – 25	232 (34)
25 – 30	256 (38)
30 – 35	97 (14)
35 - 40	41 (6)
\geq 40	26 (4)
Age at first cancer diagnosis	60.3 \pm 17.0
Woman cancers	503 (74)
Breast	381 (56)
Cervix	45 (7)
Ovary	43 (6)
Endometrium	34 (5)
Other cancers	174 (26)
Metastatic disease	
At 2 years of follow-up	317 (47)
At 3 years of follow-up	324 (48)
AID	108 (16)
Severe AID	17 (2.5)

Early AID	73 (10)
CVD	156 (23)
Early CVD	22 (3)
CVRF	258 (38)
Early CVRF	114 (17)
Early or multiple cancers	222 (33)
Early familial cancer	96 (14.5)
At least 2 cancers at 1 st degree	79 (11.5)
Familial AID	34 (5)
Early familial AID	20 (3)
Familial CVD	34 (5)
Early familial CVD	20 (3)
Familial CVRF	51 (7.5)
Early familial CVRF	24 (3.5)
Number of deaths at the time of analysis	
At 2 years of follow-up	207 (30.5)
At 3 years of follow-up	244 (36)

severe AID = at least 1 auto immune disease personal severe

early AID = at least 1 auto immune disease personal early

early CVD = at least 1 personal cardiovascular disease before 50 years

early CVRF = at least 1 personal cardiovascular risk factor before 50 years

early or multiple cancer = first cancer before 50 years and or at least 2 personal cancers

early familial cancer = cancer before 50 years 1st or 2nd degree family

familial AID = at least 2 AID 1st and 2nd degree

early familial AID = at least 1 AID 1st degree before 50 years

familial cardiovascular disease = at least 2 CVD 1st or 2nd degree family

early familial cardiovascular disease = at least 1 CVD 1st degree before 50 years

familial vascular risk factor = at least 2 CVRF 1st or 2nd degree family

early familial risk factor = at least 1 CVRF before 50 years 1st degree family

Table 2: Comparison of baseline characteristics between men and women

Variables	Mean \pm SD or N (%)		<i>P</i> -value
	Men (N = 127)	Women (N = 550)	
Age (y)	63.37 \pm 18.43	62.5 \pm 17.0	0.52
Smoking (pack /year)	16 \pm 18.49	3.9 \pm 11.5	0.0
No smoking	41 (32)	438 (80)	
< 10 pack/year	35 (28)	57 (10)	
\geq 10 pack/year	51 (40)	55 (10)	
Body Mass Index (BMI)	26.7 \pm 5.0	26.6 \pm 5.0	0.20
< 19	5 (4)	20 (4)	
19 – 25	44 (35)	188 (34)	
25 – 30	55 (43)	201 (36)	
30 – 35	15 (12)	82 (15)	
35 - 40	4 (3)	37 (7)	
\geq 40	4 (3)	22 (4)	
Age at first cancer diagnosis	60.4 \pm 17.6	60.3 \pm 16.3	0.96
Metastatic disease			
At 2 years of follow-up	87 (69)	230 (41)	0.00
At 3 years of follow-up	90 (71)	234 (43)	0.00
AID	9 (7)	99 (18)	0.02
Severe AID	1 (1)	14 (3)	0.37
Early AID	1 (1)	38 (7)	0.01
CVD	45 (34)	111 (20)	0.00
Early CVD	2 (2)	20 (4)	0.36
CVRF	45 (35)	213 (39)	0.56
Early CVRF	14 (11)	100 (18)	0.07

Early or multiple cancers	39 (31)	183 (33)	0.65
Early familial cancer	8 (6)	88 (16.0)	0.01
At least 2 cancers at 1 st degree	11 (9)	68 (12)	0.31
Familial AID	0 (0)	24 (4)	0.03
Early familial AID	-	-	-
Familial CVD	2 (2)	32 (6)	0.08
Early familial CVD	3 (2)	17 (3)	0.88
Familial CVRF	5 (4)	46 (8)	0.109
Early familial CVRF	1 (1)	23 (3)	0.11
<hr/>			
Number of deaths at the time			
of analysis			
At 2 years of follow-up	61 (48)	146 (25)	0.00
At 3 years of follow-up	74 (58)	170 (31)	0.00

P-values derived from Student's t-test for continuous variables and Chi-square test for categorical variables.