University of Groningen

Universal risk factors for multifactorial diseases - LifeLines: a three-generation population-based study

Stolk, Ronald P.; Rosmalen, Judith G. M.; Postma, Dirkje S.; de Boer, Rudolf A.; Navis, Gerjan; Slaets, Joris P. J.; Ormel, Johan; Wolffenbuttel, Bruce H. R.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2008

Link to publication in University of Groningen/UMCG research database

NEW STUDY

Universal risk factors for multifactorial diseases

LifeLines: a three-generation population-based study

Ronald P. Stolk · Judith G. M. Rosmalen · Dirkje S. Postma ·
Rudolf A. de Boer · Gerjan Navis · Joris P. J. Slaets · Johan Ormel ·
Bruce H. R. Wolffenbuttel

Abstract The risk for multifactorial diseases is deter-
mined by risk factors that frequently apply across disorders
(universal risk factors). To investigate unresolved issues on
etiology of and individual's susceptibility to multifactorial
diseases, research focus should shift from single determi-
nant-outcome relations to effect modification of universal
risk factors. We present a model to investigate universal risk
factors of multifactorial diseases, based on a single risk
factor, a single outcome measure, and several effect modi-
fiers. Outcome measures can be disease overriding, such as
clustering of disease, frailty and quality of life. "Life course
epidemiology" can be considered as a specific application of
the proposed model, since risk factors and effect modifiers of
multifactorial diseases typically have a chronic aspect. Risk
factors are categorized into genetic, environmental, or
complex factors, the latter resulting from interactions
between (multiple) genetic and environmental factors (an
example of a complex factor is overweight). The proposed
research model of multifactorial diseases assumes that
determinant-outcome relations differ between individuals
because of modifiers, which can be divided into three cate-
gories. First, risk-factor modifiers that determine the effect
of the determinant (such as factors that modify gene-
expression in case of a genetic determinant). Second, out-
come modifiers that determine the expression of the studied
outcome (such as medication use). Third, generic modifiers
that determine the susceptibility for multifactorial diseases
(such as age). A study to assess disease risk during life
requires phenotype and outcome measurements in multiple
generations with a long-term follow up. Multiple genera-
tions will also enable to separate genetic and environmental
factors. Traditionally, representative individuals (probands)
and their first-degree relatives have been included in this
type of research. We put forward that a three-generation
design is the optimal approach to investigate multifactorial
diseases. This design has statistical advantages (precision,
multiple-informants, separation of non-genetic and genetic
familial transmission, direct haplotype assessment, quantify
genetic effects), enables unique possibilities to study social
characteristics (socioeconomic mobility, partner prefer-
ences, between-generation similarities), and offers practical

R. P. Stolk (✉)
Department of Epidemiology, University Medical Center
Groningen, University of Groningen, P.O. BOX 30.001, 9700
RB Groningen, The Netherlands
e-mail: R.P.Stolk@epi.umcg.nl

J. G. M. Rosmalen · J. Ormel
Department of Psychiatry, University Medical Center
Groningen, University of Groningen, Groningen,
The Netherlands

D. S. Postma
Department of Pulmonology, University Medical Center
Groningen, University of Groningen, Groningen,
The Netherlands

R. A. de Boer
Department of Cardiology, University Medical Center
Groningen, University of Groningen, Groningen,
The Netherlands

G. Navis
Department of Nephrology, University Medical Center
Groningen, University of Groningen, Groningen,
The Netherlands

J. P. J. Slaets
Department of Geriatrics, University Medical Center Groningen,
University of Groningen, Groningen, The Netherlands

B. H. R. Wolffenbuttel
Department of Endocrinology, University Medical Center
Groningen, University of Groningen, Groningen,
The Netherlands

benefits (efficiency, lower non-response). LifeLines is a study based on these concepts. It will be carried out in a representative sample of 165,000 participants from the northern provinces of the Netherlands. LifeLines will contribute to the understanding of how universal risk factors are modified to influence the individual susceptibility to multifactorial diseases, not only at one stage of life but cumulatively over time: the lifeline.

Keywords Multifactorial disease · Effect modification · Gene–environment · Biobank

Introduction

Multifactorial diseases are by definition the result of multiple risk factors that are both genetically and environmentally determined. Examples of multifactorial diseases are depression, COPD, cancer, cardiovascular and endocrine diseases. Together they comprise the most common disorders in adulthood and are responsible for the use of the majority of health care resources [1]. Biomedical and epidemiological research on the etiology of multifactorial diseases frequently focuses on single determinant—outcome relations, without taking into account other risk factors, other diseases and time dependent effects. This has been recognized over the last years and resulted in new study designs sometimes referred to as "life course epidemiology" [2].

Multifactorial diseases may have more in common than generally recognized since similar risk factors are associated with multiple diseases, as has been shown for example by low birth weight [3]. A risk factor like smoking results in lung cancer in some individuals and myocardial infarction in others, whereas it has a protective effect on Parkinson's disease, suggesting an individual susceptibility for specific risk factors [4, 5]. The individual differences that determine which disease occurs in the presence of a given universal risk factor are called modifiers [6].

Since different diseases share identical risk factors, it is clear that a continuing exclusive focus on associations between single risk factors and single outcomes will not unravel the unresolved issues of etiology and individual prognosis of multifactorial diseases. Instead, research has to focus on the underlying mechanisms why individuals with similar (established) risk factors develop different diseases, i.e. the modification of the universal risk factors for multiple disorders [7].

In this paper we present a model based on effect modification to investigate universal risk factors and their modifiers of multifactorial diseases. We also discuss the three-generation design, which provides optimal methods to study the interplay of genetic and environmental risk factors.

Research model for multifactorial diseases

The proposed model to investigate universal risk factors of multifactorial diseases is based on a single (universal) risk factor, a single outcome measure (which may refer to multiple diseases, see below), and several effect modifiers. This research model is summarized in Fig. 1. Risk factors can be categorized into genetic, environmental or complex factors, the latter resulting from interactions between (multiple) genetic and environmental factors. Examples of complex factors are overweight and personality traits. In the same way, we distinguish effect modifiers as genetic, environmental and complex modifiers.

It is important to realize that the same factor can be risk factor in one research question but modifier or outcome in another research question. The research question determines the role of the included factors. Like all models, the proposed research model of risk factor—modifier—outcome (Fig. 1) is a simplification to understand the underlying pathophysiology. Especially the time dependent effect of modification needs careful attention: the susceptibility of an individual for the effects of a risk factor may differ during the lifeline. The research model is based on risk and not on causation, like for example Rothman's pie model of sufficient cause [6]. Within Rothman's scheme the proposed model explains why individuals with a similar component cause (risk factor) may develop different diseases.

Research models for multifactorial diseases can be distinguished into disorder-specific, latency and pathway models. *The disorder-specific models* for morbidity accumulation constitute the basis of most biomedical research and of medical superspecialisation [8–10]. They focus on single outcomes. For example, low birth weight raises the risk of COPD and, in turn, COPD (for example due to the stress associated with being a COPD patient) raises the risk of depression. Low birth weight needs only to be linked with the first problem in this chain to fully explain the association between birth weight and depression. In these chains-of-risk models, risk factors are disorder specific. In contrast, *latency models* propose a disorder-generic liability to whichever morbidity as a stable programmed individual feature, related to early events and genetic influences [3, 11, 12]. The link of low birth weight with both COPD and depression is, in these models, taken to indicate that it raises the risk of each of them irrespective
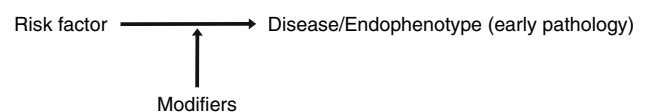

Fig. 1 The role of modifiers in the research model for multifactorial diseases

of whether the other is experienced. From this point of view, not only morbidity but also an individual's lifespan is determined early on. Finally, *pathway models* like the proposed research model of risk factor-modifier-outcome, are models that allow for the fact that disease susceptibility needs not be stable but may change over time in response to illness, life styles or advancing age [13]. In fact, "life course epidemiology" can be regarded as a specific application of our proposed research model [2]. Outcome at a certain age can constitute a risk factor or a modifier at a later age.

Genetic risk factors

Genetic risk factors are defined as changes in the base pair sequence of the human genome, which do not change during life. In the previous decade, genetic association studies have generated many data concerning the genetic basis of multifactorial diseases. Several genetic polymorphisms have been linked to more than one disease, examples are polymorphisms in the *ACE*, *TGF-β* and *TNF-α* genes [14–16]. Even if specific polymorphisms have only been linked to a single disease, often different mutations in the same gene are related to other diseases, supporting the idea of common pathways in different multifactorial diseases.

Environmental risk factors

Environmental risk factors are experienced throughout life. They vary from life events to air pollution and medical interventions. The rapid change of morbidity patterns within one or two generations clearly illustrates the importance of environmental factors in the development, progression and remission of multifactorial diseases. Exposure to "western lifestyle" is frequently blamed for the increase in prevalence of many of these disorders over the past decades [17]. Some diseases have a well-established environmental risk factor, which explains a large proportion of disease risk. For instance cigarette smoking is associated with COPD (90% of COPD patients have significant smoking history) [18], and major life events increase the risk for depression [19, 20]. Though these associations have been clearly established, still details of the pathophysiologic pathways have been elucidated insufficiently.

As mentioned above, environmental risk factors include medical interventions (drugs, surgery, psychological consultations, etc.) and exposure to life style factors (diet, smoking, physical activity). These environmental factors are (thought to be) modifiable and often used in clinical practice and intervention studies. In contrast, other environmental risk factors are more or less 'fixed', like past environmental experiences (intra-uterine environment; exposures at day care center, school and occupation) and macro-environmental exposures (air pollution).

Socioeconomic status (SES) is an intriguing complex risk factor for multivariate diseases. It has a strong environmental component, but genetic factors may be involved as well [21]. Several studies have indicated that SES affects the onset of diseases through a higher prevalence of risk factors like hypertension and obesity among people with lower SES. An alternative hypothesis proposes that lower socioeconomic status is (at least partly) caused by ill health instead of the other way around [22].

Complex risk factors

The origin of many risk factors is not entirely genetic or environmental but the result of interactions between genes and environmental factors. The interaction between multiple genetic and/or environmental factors can result in (endo)phenotypic characteristics, which in the proposed model are referred to as complex risk factors. Examples are body weight, personality traits, and plasma cortisol level. This also include epigenetic changes: environmental factors that result in phenotypic changes in gene expression without altering the genotype [23]. Often the interaction between genetic and environmental risk factors resulting in a complex risk factor is not completely elucidated, even though the role of this complex factor as risk factor for disease development has been established, for example by proven benefit from intervention on this factor. Selected multifactorial diseases are largely explained by a single complex risk factor, for example obesity and type 2 diabetes mellitus [24]. Because of their well-defined pathophysiologic role, numerous biomedical and epidemiological studies focus on determinants of and interventions on complex risk factors.

Effect modifiers

The risk of a genetic, environmental or complex factor on the occurrence of disease often differs between individuals as well as between different stages of an individual's life. Similar risk factors result in separate multifactorial diseases in different individuals, which can be explained by modifying factors. This notion forms the basis for the proposed research model (Fig. 1). Modifiers explain the difference in effect of a risk factor between individuals or even within an individual over time. It is the challenge of future biomedical and epidemiological research to unravel the role of modifiers in multifactorial diseases. As

mentioned above, the same factor can be risk factor in one research question but modifier or outcome in another research question.

Effect modifiers can be divided into three categories, by referring to three different parts of the pathophysiologic mechanisms of a disease (Fig. 2). First, *risk-factor modifiers* that change the effect of the determinant (such as renal function and factors that modify gene-expression in case of a genetic determinant), which is sometimes referred to as homeostatic switch. Second, *outcome modifiers* that determine the expression of the studied outcome (such as medication use, immune-status). Third, *generic modifiers* that determine the susceptibility for multifactorial diseases (such as age, socioeconomic status). These categories are based on pathophysiologic mechanisms; the statistical approach is similar for all modifiers.

Apart from their direct effects on different parts of the pathophysiologic pathway of a disease, modifiers may have a time-dependent effect. The same factor that occurs during childhood will change susceptibility differently than when occurring at older ages.

Statistical methods to model different effects of a modifier over time have been developed for "life course epidemiology" [2].

Outcomes

Research on multifactorial disease tends to focus on single outcomes, typically clinically defined diseases like diabetes, myocardial infarction, depression, dementia and COPD. Although these concepts are relevant in clinical care in order to make decisions on treatment, the distinction between presence and absence of a multifactorial clinical disease is often not clearly defined. Thus a diagnosis may be insufficiently clear-cut for genetic purposes (e.g. in case the disease becomes only clinically apparent at older age, but is present already at earlier age) and as a consequence obscures the true underlying mechanism. In that case, subphenotypes are often used to overcome the bias of a doctor's diagnosis. In the same way, hospitalization and mortality are attractive measures, but their cause does not exclusively depend on the disease under study alone. On the other hand, many diseases are defined by an arbitrary cut-off in a continuous disease marker or endophenotype like plasma glucose level, serum IgE or score on a depression scale.

More importantly, like the principal shortcoming of using single determinants in research on multifactorial disease one should apply multiple outcome measures. Comorbidity (when a certain index disorder is accompanied by one or more other disorders [25] or multimorbidity (concurrence of two or more medical conditions within a person) [26] is ubiquitous in clinical practice of multifactorial diseases. It has been well established that the proportion of people with multiple diseases may vary from 30% in the general population [27] to over 50% in people aged 60 years and older [28]. The presented research model (Fig. 1) includes a single outcome measure. There are two approaches to investigate co-morbidity in this model, depending on the research question. A first approach is to define clustering of disease as a separate endpoint (either related disorders like diabetes and amputation, or presumably non-related like arthritis and cancer). This assumes that a specific combination of risk factors and modifiers results in a specific combination of disorders.

A second approach to quantify multifactorial disease in the proposed model is to use a generic pathophysiologic

Fig. 2 Effect modifiers classified according to their role in the pathophysiologic mechanisms of multifactorial diseases
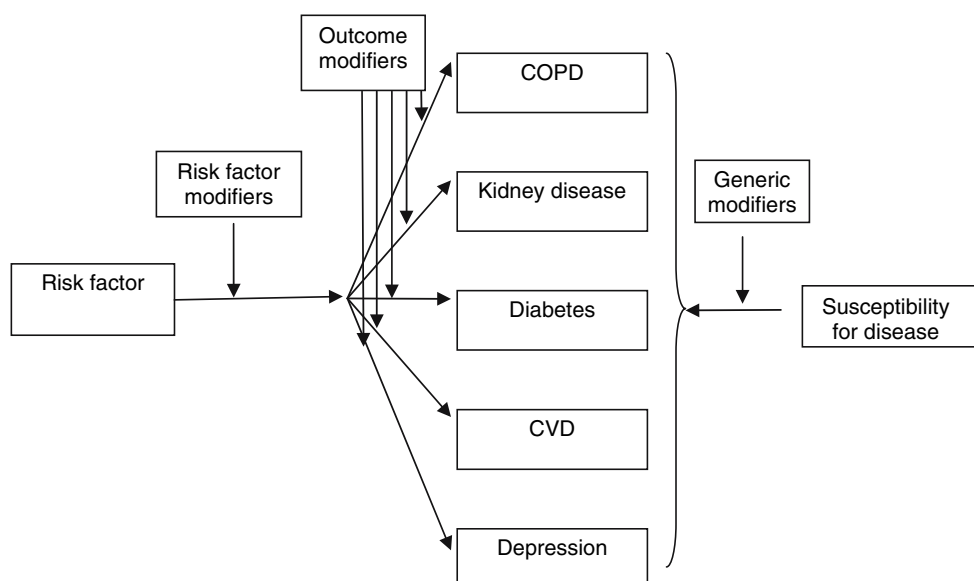
process (for example inflammation) or a marker of burden of disease, such as the Chronic Disease Score [29]. Other terms applied in the literature for the burden of disease include disability, quality of life, and frailty. Although these are used interchangeably, they refer to different entities with each a distinct content [30]. Disability is defined as the difficulty to carry out activities of daily living, but can also be viewed as social phenomenon [31] Quality of life can be defined both from societal (including income, employment, housing, education) and individual perspective (including personal experiences and values, happiness, life satisfaction) [32]. Frailty is defined as a state of increased physiologic vulnerability for adverse health outcomes, characterized by wasting, loss of endurance, decreased mobility and potentially decreased cognitive function [33]. Frailty is associated with chronological ageing and is equated to an increased risk of death. It can be regarded as biological, as opposed to chronological age [34]. Therefore, frailty might be the preferred generic measure of (multifactorial) disease when considering modification of risk factors. Also longevity might be a suitable general outcome measure [35].

Study design

The investigation of effect modification on development of disease requires dedicated study design solutions, principally a large sample size to allow for stratified analyses. In addition, assessing risk factors during life requires a long-term follow up, with measurements in multiple generations. The advantage of including more generations is the possibility to separate genetic and environmental factors. Traditionally, representative individuals (probands) and their first-degree relatives have been included in this type of research. A three-generation study design goes a step further by including also the partner of the proband and his/her parents (if present) as well as the children of the proband (if any). This design has statistical advantages with respect to multiple-level information, separation of non-genetic and genetic familial transmission and direct haplotype assessment. In addition, because of the inclusion of step-family members it the design enables to quantify genetic effects [36]. Furthermore it opens unique possibilities to study social characteristics (socioeconomic mobility, partner preferences, between-generation similarities), and offers practical benefits (lower non-response). An overview of these aspects is given in Table 1.

Another phenomenon that can be studied within this design is "assortative mating" (partner preference), the selection of partners based on (patho)physiological characteristics. This contributes to the concentration within particular families of genetic and environmental risk factors. There is substantial evidence that assortative mating affects height, physical attractiveness, SES, ethnicity, religion, social attitudes, and particular behaviors like antisocial behavior [37]. Estimating the size of this phenomenon enables better interpretation of the relevance of (absence of) effect modification.

An additional advantage of a three-generation design is the wide age range of the participants. This allows ascertaining (pre-clinical) events at an early age thereby providing insight into time-dependent effects. Furthermore, a variety of exposures that affect disease development at a different age can be examined, an important aspect since exposures often vary by age. Also, genotype-exposure interaction can be examined stratified by age.

Outline of the LifeLines project

Based on the reviewed concepts of modifiers and three-generation design we have developed a cohort study to investigate universal risk factors and their modifiers for multifactorial diseases: LifeLines. This study may result in better understanding of the causes and prognosis of burden of disease over lifetime and may ultimately result in optimal tailored treatment of individual diseases and disease overriding preventive strategies. Specific research questions will focus on risk factors and modifiers (genetic, environmental and complex factors) for single and multiple diseases. Rather than co-morbidity, LifeLines focuses on co-determinants. Compared to large scale genetic epidemiological studies like Biobank UK and deCODE Iceland, LifeLines includes more detailed measurements of environmental factors, as well as changes of risk factors, assessment of endophenotypes and incidence of disease.

Secondary aims include the assessment of the prevalence and incidence of multifactorial diseases and their risk factors in individuals as well as in families. The burden of disease for the society will be quantified in terms of care needed.

LifeLines is an observational follow-up study in a large representative sample of the population of the northern provinces of the Netherlands covering three generations. Firstly, a random sample of persons aged between 25 and 50 years are invited to participate. Subsequently their family members if present are invited to also take part (parents, partner, parents in law, children), resulting in a three-generation study.

The core of the LifeLines project consists of dedicated data collection and biological sample storage, including genetic samples ("biobank"). All participants receive a number of questionnaires and a basic medical examination and are followed for many years with extensive standardized measurements.

Table 1 Advantages of a three-generation study design

Statistical advantages	
Haplotype assessment	The inclusion of both parents enables the direct assessment of haplotypes, which is usually not possible in population-based studies.
Multi-informants	Family members can provide data on individual characteristics and on shared environmental exposures. Multi-informant data permits to estimate and reduce information source bias and to increase the reliability in assessing phenotypes.
Family-wide effects	Family-wide effects can be examined at three levels by means of multilevel statistical methods: the marital relationship (proband + partner), the primary family (proband, partner, and offspring living at home) and the extended family (proband, partner, their parents and offspring).
Quantitative genetics	Possibility to perform quantitative genetics.
Separation of non-genetic (cultural) and genetic familial transmission	The difference between the familial and genetic loadings helps to disentangle the familial similarity in genetic and non-genetic transmitted components.
Unique possibilities to study social patterns	
Between-generation (dis)similarities	The transmission of severity and specificity of a particular trait or disorder from (grand)mother and/or (grand)father can be examined. It also enables to distinguish individual behavioral changes versus changes due to external effects like cultural changes or legislation.
Socioeconomic mobility	The three-generation design offers the unique possibility to examine socioeconomic mobility.
Assortative mating	The effects of assortative mating on a variety of individual characteristics can be investigated, ranging from the (patho)physiological to the psychosocial domains.
Practical benefits	
Reduction of non-response and attrition	It seems likely that it is easier to maintain the cohort if it consists of family members as compared to non-related individuals. Participation in the study could become sort of family activity.

A cohort study, in contrast to a case-control study, enables the prospective investigation of risk factors, which is crucial in the study of environmental and other time-varying exposures, as well as interactions between environmental risk factors [38]. For genetic studies a case-control design is often more appropriate, but in such a design it is virtually impossible to investigate gene–environment interactions.

Sample size

The LifeLines project will include 165,000 participants: anticipated to consist of approximately 45,000 probands, 30,000 partners, 55,000 parents (in law) and 35,000 children. This number is based on balancing costs and practical limitations with sufficient number of incident diseases. When estimating the number of events one has to realize that sick individuals are less likely to participate. Following the adjustment for this "Hawthorne effect" as suggested by the UK Biobank [39], after 5 years of follow-up the expected incident cases of some common

multifactorial diseases in this cohort are 1,000 individuals with myocardial infarctions, 500 with stroke, 2,000 with depression. Based on a prevalence of 20–40% of the risk factors of interest, and estimated relative risk around 1.2, these numbers are sufficient to identify statistically significant associations. Newly developed statistical methods to analyze combinations of (genetic) risk factors will improve the effectiveness of these databases [40].

However, the main objective for LifeLines is to investigate effect modification, or interaction in statistical terms. By performing stratified analyses or introducing more additional terms into a regression model, each with its own variation, the required number of participants increases substantially. This is only partly compensated by the fact that these analyses will typically use continuous measures as outcome (endophenotype). The power calculations for these analyses are strongly influenced by the interaction ratio, which is often not known.

Methods of data collection are matched with other ongoing biobank studies (P3G consortium) [41], which enables combining datasets to construct large study populations.

Conclusion

Large-scale biobank studies have recently been started in the United Kingdom (the UK BioBank), Iceland (deCODE), Estonia, Germany, Canada and Japan. There are serious plans to start a biobank study in the United States [42]. Most of these projects are focused on DNA and collect only limited data on environmental factors. As recognized by the NIH, population based studies on multifactorial diseases should include both genetic and environmental factors on a population basis, and focus on identifying genetic and environmental modifiers of this risk (gene-gene and gene–environment interactions) [42].

For the assessment of multifactorial diseases it is important to understand how the interaction between universal risk factors and specific modifiers, e.g. risk-factor modifiers, outcome modifiers, and generic modifiers, accounts for the development of multifactorial disease in individuals. The explicit aim of Lifelines, to investigate risk factors that apply across disorders, is at odds with common practice of biomedical research that strongly focuses on single disease entities. LifeLines constitutes to our opinion a large step forward and a challenge to better understanding of the origins of health and disease cumulatively over time: the lifeline.

References

1. World Health Organisation. World Health Report 2002; Reducing risk, promoting healthy life. Geneva: World Health Organisation; 2002.
2. De Stavola BL, Nitsch D, dos Santos Silva I, et al. Statistical issues in life course epidemiology. Am J Epidemiol. 2006;163:84–96.
3. Barker DJ, Eriksson JG, Forsen T, et al. Fetal origins of adult disease: strength of effects and biological basis. Int J Epidemiol. 2002;31:1235–39.
4. Zaher C, Halbert R, Dubois R, et al. Smoking-related diseases: the importance of COPD. Int J Tuberc Lung Dis. 2004;8:1423–8.
5. Baron JA. Beneficial effects of nicotine and cigarette smoking: the real, the possible and the spurious. Br Med Bull. 1996;52:58–73.
6. Rothman KJ, Greenland S. Causation and causal inference in epidemiology. Am J Public Health 2005;95 Suppl 1:S144–50.
7. Kelada SN, Eaton DL, Wang SS, et al. The role of genetic polymorphisms in environmental health. Environ Health Perspect. 2003;111:1055–64.
8. Aneshensel CS, Frerichs RR, Huba GJ. Depression and physical illness: a multiwave, nonrecursive causal model. J Health Soc Behav. 1984;25:350–71.
9. Hotopf M, Mayou R, Wadsworth M, et al. Temporal relationships between physical symptoms and psychiatric disorder - Results from a national birth cohort. Br J Psychiatry 1998;173:255–61.
10. Stansfeld SA, Smith GD, Marmot M. Association between physical and psychological morbidity in the Whitehall II Study. J Psychosom Res. 1993;37:227–38.
11. Cynader MS. Mechanisms of brain-development and their role in health and well-being. Daedalus 1994;123:155–65.
12. Plagemann A. Fetal programming and functional teratogenesis: on epigenetic mechanisms and prevention of perinatally acquired lasting health risks. J Perinat Med. 2004;32:297–305.
13. Power C, Hertzman C. Social and biological pathways linking early life and adult disease. Br Med Bull. 1997;53:210–21.
14. Sayed-Tabatabaei FA, Oostra BA, Isaacs A, et al. ACE polymorphisms. Circ Res. 2006;98:1123–33.
15. Neil JR, Galliher AJ, Schiemann WP. TGF-beta in cancer and other diseases. Future Oncol. 2006;2:185–9.
16. Russo C, Polosa R. TNF-alpha as a promising therapeutic target in chronic asthma: a lesson from rheumatoid arthritis. Clin Sci (Lond). 2005;109:135–42.
17. Pereira MA, Kartashov AI, Ebbeling CB, et al. Fast-food habits, weight gain, and insulin resistance (the CARDIA study): 15-year prospective analysis. Lancet 2005;365:36–42.
18. Rabe KF, Hurd S, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med. 2007;176:532–55.
19. Tennant C. Life events, stress and depression: a review of recent findings. Aust N Z J Psychiatry 2002;36:173–82.
20. Ormel J, Oldehinkel AJ, Brilman EI. The interplay and etiological continuity of neuroticism, difficulties, and life events in the etiology of major and subsyndromal, first and recurrent depressive episodes in later life. Am J Psychiatry 2001;158:885–91.
21. Lichtenstein P, Pedersen NL. Does genetic variance for cognitive abilities account for genetic variance in educational achievement and occupational status? A study of twins reared apart and twins reared together. Soc Biol. 1997;44:77–90.
22. Kessler RC, Foster CL, Saunders WB, et al. Social consequences of psychiatric disorders, I: educational attainment. Am J Psychiatry 1995;152:1026–32.
23. Fraga MF, Ballestar E, Paz MF, et al. Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci USA 2005;102:10604–9.
24. Bray GA. Medical consequences of obesity. J Clin Endocrinol Metab. 2004;89:2583–9.
25. Feinstein AR. The pre-therapeutic classification of comorbidity in chronic disease. J Chronic Dis. 1970;23:455–68.
26. van den Akker M, Buntinx F, Knottnerus JA. Comorbidity or multimorbidity; what's in a name? A review of the literature. Eur J Gen Pract. 1996;2:65–70.
27. van den Akker M, Buntinx F, Metsemakers JFM, et al. Multimorbidity in general practice: prevalence, incidence, and determinants of co-occurring chronic and recurrent diseases. J Clin Epidemiol. 1998;51:367–75.

28. Guralnik JM, Kaplan GA. Predictors of healthy aging: prospective evidence from the Alameda County study. Am J Public Health 1989;79:703–8.
29. Von KM, Wagner EH, Saunders K. A chronic disease score from automated pharmacy data. J Clin Epidemiol. 1992;45:197–203.
30. Fried LP, Ferrucci L, Darer J, et al. Untangling the concepts of disability, frailty, and comorbidity: implications for improved targeting and care. J Gerontol A Biol Sci Med Sci. 2004;59:255–63.
31. Ormel J, Vonkorff M, Ustun TB, et al. Common mental disorders and disability across cultures. Results from the WHO Collaborative Study on Psychological Problems in General Health Care. JAMA 1994;272:1741–8.
32. Bowling A. Mode of questionnaire administration can have serious effects on data quality. J Public Health (Oxf) 2005;27:281–91.
33. Ferrucci L, Guralnik JM, Studenski S, et al. Designing randomized, controlled trials aimed at preventing or delaying functional decline and disability in frail, older persons: a consensus report. J Am Geriatr Soc. 2004;52:625–34.
34. Rockwood K, Howlett SE, MacKnight C, et al. Prevalence, attributes, and outcomes of fitness and frailty in community-dwelling older adults: report from the Canadian study of health and aging. J Gerontol A Biol Sci Med Sci. 2004;59:1310–7.
35. Boxenbaum H. Gompertz mortality analysis: aging, longevity hormesis and toxicity. Arch Gerontol Geriatr. 1991;13:125–37.
36. Quantitative Genetic Theory. In: Plomin R, De Fries JC, McClearn GE, et al., editors. Behavioral genetics. A primer.4th ed. New York: Freeman & co; 2000. p. 207–254.
37. Luo S, Klohnen EC. Assortative mating and marital quality in newlyweds: a couple-centered approach. J Pers Soc Psychol. 2005;88:304–26.
38. Collins FS. The case for a US prospective cohort study of genes and environment. Nature 2004;429:475–7.
39. Barbour V. UK Biobank: a project in search of a protocol? Lancet 2003;361:1734–8.
40. Ritchie MD, Hahn LW, Roodi N, et al. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet. 2001;69:138–47.
41. http://www.p3gconsortium.org. 2006. Accessed Dec 2007.
42. http://www.genome.gov/13014436. 2006. Accessed Dec 2007.