Depression and diabetes

Methodological issues in etiologic research

Mirjam Judith Knol

Depression and diabetes
Methodological issues in etiologic research

Utrecht, Universiteit Utrecht, Faculteit Geneeskunde
Thesis, with a summary in Dutch
Proefschrift, met een samenvatting in het Nederlands

ISBN	978-90-393-4802-4
Author	M.J. Knol
Lay-out	M.J. Knol
Cover design	H. te Walvaart
Print	Gildeprint BV, Enschede

The research described in this thesis was funded by an unrestricted grant from Novo Nordisk and the Scientific Institute of Dutch Pharmacists (WINAp). A research visit to the Institute of Social and Preventive Medicine in Bern, Switzerland, was funded by the Prince Bernhard Cultural Foundation and the University Medical Center Utrecht.

Financial support by the Dutch Diabetes Research Foundation, the J.E. Jurriaanse Stichting, the Julius Center for Health Sciences and Primary Care and Utrecht Institute for Pharmaceutical Sciences for the publication of this thesis is gratefully acknowledged.

Depression and diabetes
Methodological issues in etiologic research

Depressie en diabetes

Methodologische aspecten van etiologisch onderzoek

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht

op gezag van de rector magnificus, prof.dr. J.C. Stoof,

ingevolge het besluit van het college voor promoties in het openbaar te verdedigen

op donderdag 29 mei 2008 des ochtends te 10.30 uur

door

Mirjam Judith Knol

geboren op 22 juni 1981

te Amersfoort

Promotoren:	Prof.dr. D.E. Grobbee
	Prof.dr. A.C.G. Egberts
Co-promotoren:	Dr. M.I. Geerlings
	Dr. E.R. Heerdink

Manuscripts based on the studies presented in this thesis

Chapter 2.1

MJ Knol, JWR Twisk, ATF Beekman, RJ Heine, FJ Snoek, F Pouwer. Depression as a risk factor for the onset of type 2 diabetes. *Diabetologia 2006; 49: 837-845.*

Chapter 2.2

MJ Knol, MI Geerlings, ACG Egberts, KJ Gorter, DE Grobbee, ER Heerdink. No increased incidence of diabetes in antidepressant users. *International Clinical Psychopharmacology 2007; 22: 382-386.*

Chapter 2.3

MJ Knol, HJ Derijks, MI Geerlings, ER Heerdink, PC Souverein, KJ Gorter, DE Grobbee, ACG Egberts. Influence of antidepressants on glycemic control in patients with diabetes mellitus. *Pharmacoepidemiology and Drug Safety (accepted for publication).*

Chapter 2.4

MJ Knol, ER Heerdink, ACG Egberts, MI Geerlings, KJ Gorter, ME Numans, DE Grobbee, OH Klungel, H Burger. Depressive symptoms in subjects with diagnosed and undiagnosed type 2 diabetes. *Psychosomatic Medicine 2007; 69: 300-305*.

Chapter 2.5

MJ Knol, MI Geerlings, DE Grobbee, ACG Egberts, ER Heerdink. Antidepressant use before and after the start of diabetes mellitus treatment. *Submitted.*

Chapter 3.1

MJ Knol, KJM Janssen, ART Donders, ACG Egberts, ER Heerdink, DE Grobbee, KGM Moons, MI Geerlings. Unpredictable bias when using the missing indicator method or complete case analysis for missing confounder values: an empirical example. *Submitted.*

Chapter 3.2

MJ Knol, I van der Tweel, DE Grobbee, ME Numans, MI Geerlings. Estimating interaction on an additive scale between continuous determinants in a logistic regression model. *International Journal of Epidemiology 2007; 36: 1111-1118.*

Chapter 3.3

MJ Knol, M Egger, P Scott, MI Geerlings, JP Vandenbroucke. When one depends on the other: reporting of interaction in case-control and cohort studies published in leading journals. *Epidemiology (in revision).*

Chapter 3.4

MJ Knol, ER Heerdink, ACG Egberts, DE Grobbee, MI Geerlings. Unexpected lower prevalence of depression in patients with diabetes: potential for selection bias in a waiting room population. *Epidemiology (in revision).*

Chapter 3.5

MJ Knol, JP Vandenbroucke, P Scott, M Egger. What do case-control studies estimate? Survey of methods and assumptions in published case-control research. *American Journal of Epidemiology (in revision).*

Contents

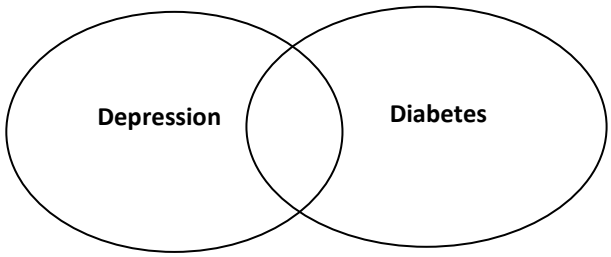
Chapter 1

General introduction

Depression and diabetes

Depression and diabetes are both common conditions in today's society. The World Health Organisation estimated that 154 million people suffered from depressive disorder worldwide in 2002 [1] and that there are more than 180 million people with diabetes worldwide [2]. In the Netherlands in 2003, 850,000 persons had a depressive disorder [3] and more than 600,000 patients had diabetes [4]. Diabetes and depressive disorder often co-occur (Figure 1). Two meta-analyses of cross-sectional studies showed that the prevalence of depression is increased in diabetes patients [5,6]. In addition, diabetes patients with depression may have worse glycemic control and more diabetes related complications than non-depressed diabetes patients [7,8].

Figure 1



The reason why depression and diabetes co-occur is not clear. There are several possible explanations. First, a common factor, such as a genetic factor, may increase the risk for depression as well as for diabetes. Second, depression may be a causal risk factor for diabetes. Possible pathophysiological mechanisms underlying this association include disturbance of the hypothalamic-pituitary-adrenocortical axis or sympathetic nervous system [9,10], and dysregulation of the immune system [11]. Third, diabetes may be a causal risk factor for depression. Biochemical changes associated with diabetes could account for the increased risk of depression, or depression may be a result of the burden of disease [12,13]. Finally, the co-occurrence of depression and diabetes can be due to chance. As both diseases are common, the chance that they are present in the same individual is higher than if the diseases were rare. This thesis focuses on the second and third explanations: whether depression is a risk factor for diabetes and whether diabetes is a risk factor for depression.

The first longitudinal study investigating whether depression is a risk factor for diabetes was performed in 1996 [14] and since then more studies examined this association [15-24]. The results of these studies are inconsistent: some studies found a significantly increased incidence of diabetes among depressed subjects, while other studies did not, and the strength of the association observed in these studies ranged from 0.98 to 2.50. The methods

of depression and diabetes assessment, the study designs and the study populations differed considerably across these studies, and might explain the large heterogeneity in effect estimates. Whether diabetes is a risk factor for depression, which is intuitively a more likely explanation for the co-occurrence of depression and diabetes, has only been investigated in two studies [23,25]. Both found no increased risk of depression in diabetes patients compared with subjects without diabetes after a period of 8 respectively 6 years of follow-up.

Methodological issues in etiologic research

Etiologic research focuses on finding causal risk factors for disease. Methodological issues, such as confounding, interaction, selection bias and information bias, are important topics commonly encountered in etiologic research. Potential biases play a substantial role in the estimation and interpretation of a relative risk, especially when the strength of the 'true' effect estimate is small (relative risk between 0.5 and 2), which is probably the case for both directions of the association between depression and diabetes. In this thesis we will discuss several of these methodological issues.

To establish that a certain risk factor is an independent and causal risk factor, external factors that can explain the association, i.e. confounders, should be ruled out. One of the challenges of etiologic research is to control for all factors that confound a certain association. Data on one or more confounders are often missing, for example because questionnaires are not filled in completely. Simple methods to handle these missing data are complete case analysis and the missing indicator method [26], while more sophisticated methods apply regression analysis to predict a missing value [27]. Simple methods, although often biased, are still widely used [28].

The association between a risk factor and disease can be modified by certain factors, such as sex, age or education level, which is called interaction or effect modification. In epidemiology, a distinction is made between interaction on an additive and a multiplicative scale, irrespective of the underlying statistical model [29]. Several epidemiologists have written about additive interaction and its calculation when a multiplicative statistical model is used [29-31]. Articles on methodology of additive interaction [32,33] and articles in which additive interaction is applied [34,35] are increasing. Despite this increase, there is still debate among epidemiologists in what circumstances additive or multiplicative scales should be used to examine interaction. Therefore, it has been suggested to provide sufficient information so that the reader can interpret interaction on an additive as well as on a multiplicative scale [36,37]. However, it is unclear whether researchers really apply this method of presenting interaction.

Typical study designs in etiologic research are cohort studies and case-control studies. In cohort studies, subjects with and without the exposure of interest are selected and followed up until occurrence of the outcome. In case-control studies, subjects with and without the outcome of interest are selected and exposure is assessed in both cases and controls. Selection bias will occur in cohort studies if the selection of persons with and without the exposure depends on the outcome of these persons, and in case-control studies if the selection of persons with and without the outcome depends on exposure in these persons. Selection bias should be dealt with in the design phase of the study and cannot be adjusted for in data-analysis.

In cohort studies both absolute and relative risks of disease can be calculated. In contrast, only relative risks, in the form of an odds ratio, can be calculated in case-control studies. Depending on certain choices in the design of a case-control study, an odds ratio can be interpreted as a risk or rate ratio [38].

Objectives

The first objective of this thesis is to investigate whether depression is a risk factor for diabetes and whether diabetes is a risk factor for depression by using different study designs and different settings. The second objective is to illustrate and discuss several methodological issues in etiologic research, in order to improve the conduct and reporting of this type of research.

Outline of this thesis

Chapter 2 describes five studies with different designs and different settings to investigate the association between depression and diabetes. **Chapter 2.1** provides a quantitative summary of nine longitudinal studies that investigated whether depression was a risk factor for the onset of type 2 diabetes. In **chapter 2.2** we used the pharmacy registration database PHARMO to investigate the association between depression and incidence of diabetes, where we used antidepressant and benzodiazepine use as proxies for depression and psychosocial complaints. The association between antidepressant use and glycemic control within diabetes patients was examined in **chapter 2.3**. **Chapter 2.4** describes the prevalence of depressive symptoms in subjects with normal fasting glucose, impaired fasting glucose, undiagnosed type 2 diabetes and diagnosed type 2 diabetes, using data from the Utrecht Health Project (UHP). In **chapter 2.5** we assessed the use of antidepressants before and after the initiation of diabetes treatment to investigate the time-specific association between diabetes and depression.

In **chapter 3** we discuss several methodological issues in etiologic research. In **chapter 3.1** we assessed, using empirical data, the degree of bias that results from using the missing indicator method or complete case analysis to handle missing data in a confounder. In **chapter 3.2** we illustrated with data from the UHP how interaction on an additive scale can be estimated when considering continuous determinants. **Chapter 3.3** shows if and how interaction is examined and reported in published cohort and case-control studies. The possibility of selection bias in a waiting room population is illustrated with empirical data from the PREDICT-NL study in **chapter 3.4**. In **chapter 3.5** we conducted a survey of published case-control studies to assess what the odds ratio estimated.

Chapter 4 provides a general discussion of the current literature including the studies presented in this thesis on whether depression is a causal risk factor for diabetes and on whether diabetes is a causal risk factor for depression. The available evidence was summarized and methodological issues of the studies were discussed.

References

1. World Health Organization: http://www.who.int/mental_health/management/depression/definition/en/.
2. World Health Organization: http://www.who.int/mediacentre/factsheets/fs312/en/index.html.
3. Rijksinstituut voor Volksgezondheid en Milieu: http://www.rivm.nl/vtv/object_document/o1275n17537.html.
4. Rijksinstituut voor Volksgezondheid en Milieu: http://www.rivm.nl/vtv/object_document/o1259n17502.html.
5. Ali S, Stone MA, Peters JL, Davies MJ, Khunti K. The Prevalence of Co-Morbid Depression in Adults With Type 2 Diabetes: a Systematic Review and Meta-Analysis. *Diabet Med* 2006;23:1165-73.
6. Anderson RJ, Freedland KE, Clouse RE, Lustman PJ. The Prevalence of Comorbid Depression in Adults With Diabetes: a Meta-Analysis. *Diabetes Care* 2001;24:1069-78.
7. de Groot M, Anderson R, Freedland KE, Clouse RE, Lustman PJ. Association of Depression and Diabetes Complications: a Meta-Analysis. *Psychosom Med* 2001;63:619-30.
8. Lustman PJ, Anderson RJ, Freedland KE, de Groot M, Carney RM, Clouse RE. Depression and Poor Glycemic Control: a Meta-Analytic Review of the Literature. *Diabetes Care* 2000;23:934-42.
9. Bjorntorp P, Holm G, Rosmond R. Hypothalamic Arousal, Insulin Resistance and Type 2 Diabetes Mellitus. *Diabet Med* 1999;16:373-83.
10. Bjorntorp P. Do Stress Reactions Cause Abdominal Obesity and Comorbidities? *Obes Rev* 2001;2:73-86.
11. Kiecolt-Glaser JK and Glaser R. Depression and Immune Function: Central Pathways to Morbidity and Mortality. *J Psychosom Res* 2002;53:873-6.
12. Peyrot M and Rubin RR. Persistence of Depressive Symptoms in Diabetic Adults. *Diabetes Care* 1999;22:448-52.
13. Talbot F and Nouwen A. A Review of the Relationship Between Depression and Diabetes in

Adults: Is There a Link? *Diabetes Care* 2000;23:1556-62.

14. Eaton WW, Armenian H, Gallo J, Pratt L, Ford DE. Depression and Risk for Onset of Type II Diabetes. A Prospective Population-Based Study. *Diabetes Care* 1996;19:1097-102.
15. Akker van den M, Schuurman A, Metsemakers J, Buntinx F. Is Depression Related to Subsequent Diabetes Mellitus? *Acta Psychiatr Scand* 2004;110:178-83.
16. Arroyo C, Hu FB, Ryan LM, Kawachi I, Colditz GA, Speizer FE et al. Depressive Symptoms and Risk of Type 2 Diabetes in Women. *Diabetes Care* 2004;27:129-33.
17. Carnethon MR, Kinder LS, Fair JM, Stafford RS, Fortmann SP. Symptoms of Depression As a Risk Factor for Incident Diabetes: Findings From the National Health and Nutrition Examination Epidemiologic Follow-Up Study, 1971-1992. *Am J Epidemiol* 2003;158:416-23.
18. Everson-Rose SA, Meyer PM, Powell LH, Pandey D, Torrens JI, Kravitz HM et al. Depressive Symptoms, Insulin Resistance, and Risk of Diabetes in Women at Midlife. *Diabetes Care* 2004;27:2856-62.
19. Golden SH, Williams JE, Ford DE, Yeh HC, Paton SC, Nieto FJ et al. Depressive Symptoms and the Risk of Type 2 Diabetes: the Atherosclerosis Risk in Communities Study. *Diabetes Care* 2004;27:429-35.
20. Kawakami N, Takatsuka N, Shimizu H, Ishibashi H. Depressive Symptoms and Occurrence of Type 2 Diabetes Among Japanese Men. *Diabetes Care* 1999;22:1071-6.
21. Kessing LV, Nilsson FM, Siersma V, Andersen PK. Increased Risk of Developing Diabetes in Depressive and Bipolar Disorders? *J Psychiatr Res* 2004;38:395-402.
22. Kumari M, Head J, Marmot M. Prospective Study of Social and Other Risk Factors for Incidence of Type 2 Diabetes in the Whitehall II Study. *Arch Intern Med* 2004;164:1873-80.
23. Palinkas LA, Lee PP, Barrett-Connor E. A Prospective Study of Type 2 Diabetes and Depressive Symptoms in the Elderly: the Rancho Bernardo Study. *Diabet Med* 2004;21:1185-91.
24. Saydah SH, Brancati FL, Golden SH, Fradkin J, Harris MI. Depressive Symptoms and the Risk of Type 2 Diabetes Mellitus in a US Sample. *Diabetes Metab Res Rev* 2003;19:202-8.
25. Bisschop MI, Kriegsman DM, Deeg DJ, Beekman AT, van Tilburg W. The Longitudinal Relation Between Chronic Diseases and Depression in Older Persons in the Community: the Longitudinal Aging Study Amsterdam. *J Clin Epidemiol* 2004;57:187-94.
26. Schafer JL and Graham JW. Missing Data: Our View of the State of the Art. *Psychol Methods* 2002;7:147-77.
27. Rubin DB. Multiple Imputation After 18+ Years. *J Am Stat Assoc* 1996;91:473-89.
28. Burton A and Altman DG. Missing Covariate Data Within Cancer Prognostic Studies: a Review of Current Reporting and Proposed Guidelines. *Br J Cancer* 2004;91:4-8.
29. Rothman KJ and Greenland S. *Modern Epidemiology*. Lippincott-Raven Publishers, Philadelphia, PA, 1998.
30. Ahlbom A and Alfredsson L. Interaction: A Word With Two Meanings Creates Confusion. *Eur J Epidemiol* 2005;20:563-4.
31. Hallqvist J, Ahlbom A, Diderichsen F, Reuterwall C. How to Evaluate Interaction Between Causes: a Review of Practices in Cardiovascular Epidemiology. *J Intern Med* 1996;239:377-82.
32. Assmann SF, Hosmer DW, Lemeshow S, Mundt KA. Confidence Intervals for Measures of

Interaction. *Epidemiology* 1996;7:286-90.

33. Hosmer DW and Lemeshow S. Confidence Interval Estimation of Interaction. *Epidemiology* 1992;3:452-6.
34. Hallan S, de MR, Carlsen S, Dekker FW, Aasarod K, Holmen J. Obesity, Smoking, and Physical Inactivity As Risk Factors for CKD: Are Men More Vulnerable? *Am J Kidney Dis* 2006;47:396-405.
35. Kamphuis MH, Geerlings MI, Tijhuis MA, Giampaoli S, Nissinen A, Grobbee DE et al. Physical Inactivity, Depression, and Risk of Cardiovascular Mortality. *Med Sci Sports Exerc* 2007;39:1693-9.
36. Vandenbroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *Epidemiology* 2007;18:805-35.
37. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. *Ann Intern Med* 2007;147:573-7.
38. Greenland S and Thomas DC. On the Need for the Rare Disease Assumption in Case-Control Studies. *Am J Epidemiol* 1982;116:547-53.

Chapter 2

Depression and diabetes

Chapter 2.1

Depression as a risk factor for the onset of type 2 diabetes mellitus. A meta-analysis

Abstract

Background

Evidence strongly suggests that depression and type 2 diabetes are associated, but the direction of the association is still unclear. Depression may occur as a consequence of having diabetes, but may also be a risk factor for the onset of type 2 diabetes. This study examined the latter association by reviewing the literature and conducting a meta-analysis of longitudinal studies on this topic.

Methods

Medline and PsycInfo were searched for articles published up to January 2005. All studies that examined the relationship between depression and the onset of type 2 diabetes were included. Pooled relative risks were calculated using fixed and random effects models. To explore sources of heterogeneity between studies, subgroup analyses and meta-regression analyses were performed.

Results

Nine studies met our inclusion criteria for this meta-analysis. The pooled relative risk was 1.26 (1.13-1.39) using the fixed effects model and 1.37 (1.14-1.63) using the random effects model. Heterogeneity between studies could not be explained by 1) whether studies controlled for undetected diabetes at baseline; 2) the method of diabetes assessment at follow-up; 3) the baseline overall risk of diabetes in the study population; and 4) follow-up duration.

Conclusions

Depressed adults have a 37% increased risk of developing type 2 diabetes mellitus. The pathophysiological mechanisms underlying this relationship are still unclear and warrant further research. A randomized controlled study is needed to test whether effective prevention or treatment of depression can reduce the incidence of type 2 diabetes and its health consequences.

Introduction

Diabetes and depression are both common conditions in today's society. There are currently about 200 million people with diabetes worldwide. If nothing is done to slow down the epidemic, the number will exceed 333 million by the year 2025 [1]. Moreover, an estimated 121 million people currently suffer from depression: 6% of men and 10% of women will experience a depressive episode in any given year [2].

There is ample evidence that diabetes and depression are associated. According to a recent meta-analysis, the prevalence of depression is doubled in individuals with type 2 diabetes compared with those without diabetes [3]. However, the temporal or causal relationship between depression and type 2 diabetes remains unclear. Depression is often regarded as a comorbid condition that results from the daily burden of having diabetes and/or its complications. Interestingly, there are also indications that depression in turn is an independent risk factor for the development of type 2 diabetes [4,5]. This is an observation that dates back to 1684, when the English physician Thomas Willis noted that emotional factors such as grief or sadness could bring on diabetes [6,7]. About 10 years ago, Eaton and colleagues were the first to report the results of an epidemiological study that confirmed Willis' hypothesis. Since then, a number of studies have investigated the relation between depression and onset of type 2 diabetes longitudinally, with inconsistent findings. Some report that depression is associated with an increased risk of developing type 2 diabetes, while other studies do not find a significant association.

The aim of this study was to examine the relationship between depression and risk of onset of type 2 diabetes by conducting a meta-analysis of longitudinal studies published on this subject in the peer-reviewed literature.

Methods

Retrieval of studies

To identify the studies of interest, two authors (F. Pouwer and M.J. Knol) independently searched Medline (1966 to January 2005) and PsycInfo (1872 to January 2005) using the search terms 'depression or depressive' and 'diabetes', limited to studies written in English and availability of an abstract. Titles and abstracts of the retrieved studies were scanned to exclude studies that were clearly irrelevant. The full texts of the remaining studies were then read to determine whether the studies met our inclusion criteria. Furthermore, the reference lists of articles that indeed studied our topic of interest were scanned to check for additional publications. Finally, in order to minimise publication bias, all members of the Psychosocial Aspects of Diabetes (PSAD) study group of the European Association for the

Study of Diabetes (EASD) and the Behavioral Research In Diabetes Group Exchange (BRIDGE) from the USA were asked by e-mail whether they had any unpublished/rejected results of studies investigating the relation between depression and onset of type 2 diabetes.

Inclusion and exclusion criteria

In this meta-analysis we included all studies that longitudinally examined the relationship between depression and onset of type 2 diabetes, irrespective of their study design. Studies were excluded if the authors did not explicitly exclude subjects with prevalent diabetes at baseline and if there were insufficient data to estimate a relative risk, either an odds ratio, risk ratio or hazard ratio. When multiple publications from the same study population were available, we included the most recent publication, regarding this study as an improvement of the older publication, representing both studies.

Data extraction

The two authors who conducted the literature search (F.P. and M.J.K.) also independently extracted data from the studies, in particular regarding: 1) name of first author; 2) publication year; 3) study design; 4) follow-up time in years; 5) number of subjects in the analysis; 6) sex of subjects; 7) age of subjects; 8) method of depression assessment; 9) method of type 2 diabetes assessment; 10) relative risk and 95% CI (the one adjusted for the largest number of confounders); 11) adjustment for confounders; 12) method of exclusion diabetes patients at baseline; 13) overall incidence per year.

The method of depression assessment was a diagnosis of depression assessed by a diagnostic psychiatric interview, the assessment of depressive symptoms by a self-reported questionnaire, or a diagnosis by a general practitioner (with an unknown method of diagnosis). The method of assessment of type 2 diabetes was either self-report or screening, i.e. measuring blood glucose of all subjects. The method of exclusion of diabetes patients at baseline was also either self-report or screening; the latter correcting for undetected diabetes as well.

If depressive symptoms were categorized in more than two groups, the relative risk of the highest versus the lowest depressive symptoms group was used. Overall incidence per year was extracted as the crude incidence of type 2 diabetes of the whole study population, divided by follow-up duration.

Statistical analysis

For each study, the relative risk of the most adjusted model was used to estimate a pooled

relative risk. Both the fixed effects model and the random effects model were used. The fixed effects model assumes that variability between studies is exclusively due to random variation and individual studies are simply weighted by their precision. The random effects model assumes a different underlying effect for each study and takes this into consideration as an additional source of variation. A random effects meta-analysis is more conservative than a fixed effects meta-analysis, as it gives wider CIs around the point estimate, and is recommended for use when heterogeneity between studies exists [8]. In the fixed effects model the weight of each study is equal to the inverse variance of the natural logarithm of the relative risk. In the random effects model an extra term is added to the variance according to the DerSimonian and Laird method [9]. A forest plot was made to show the relative risk and 95% CI of each study and the pooled relative risk and 95% CI. To provide visual assessment of publication bias a funnel plot was drawn. In this funnel plot, the relative risk for each study was plotted on the vertical axis (logarithmic scale) against the corresponding standard error on the horizontal axis. Asymmetry of the funnel plot is an indicator of publication bias. Publication bias was also assessed by means of the Begg-adjusted rank correlation test. To check the influence of publication bias, a pooled relative risk was calculated when excluding the three smallest studies.

The homogeneity between the studies was assessed visually with the forest plot and tested by means of the Cochran's Q test, of which the null hypothesis assumes homogeneity. In trying to explain heterogeneity between studies we identified certain study characteristics and assessed whether there was an association between these characteristics and the relative risks of the included studies. Four study characteristics were identified: 1) whether studies controlled for undetected diabetes at baseline; 2) the method of diabetes assessment at follow-up; 3) the overall risk of diabetes of the particular study population; and 4) follow-up duration. The first two characteristics are categorical and therefore a stratified meta-analysis was performed. The last two characteristics are continuous and meta regression analysis was performed. In these subgroup analyses only the random effects model was used.

Finally, the influence of adjusting for certain confounders was investigated qualitatively.

All statistical analyses were performed using STATA, version 7.0 (Stata, College Station, TX, USA).

Results

Search results

The literature search in Medline using 'depression or depressive' and 'diabetes', limited to

Table 1 Characteristics of studies included in the meta-analysis

Study	Study design	F-up (years)	N (% male)	Mean age ± SD	Assessment depression	Assessment diabetes	Relative risk (95% CI)	Adjustment for confounders[h]
Eaton et al. 1996 [13][a]	Cohort	13	1,715 (37.8)	18+[c]	Interview: DIS	Self report	2.23 (0.90-5.55)[d]	1,2,3,4,5,8,26
Kawakami et al. 1999 [16]	Cohort	8	2,380 (100.0)	18-53[c]	Self rating: Zung SDS	Screening WHO 1981	2.31 (1.03-5.20)[e]	1,5,6,7,8,12, 13,14,24,29
Carnethon et al. 2003 [12]	Cohort	16	6,190 (45.7)	48 ± 14	Self rating: GWB-DS	Self or doctor report	1.86 (1.27-2.71)[e]	1,2,3,8,12,13,14
Arroyo et al. 2004 [11]	Cohort	4	72,178 (0.0)	58 ± 7	Self rating: MHI-5	Self report	1.22 (1.00-1.50)[d]	1,8,12,13,14, 16,23,29
Golden et al. 2004 [15]	Cohort	6	11,615 (44.6)	56 ± 6	Self rating: VES	Screening ADA, 1998	1.31 (1.04-1.64)[e]	1,2,3,5,8,10,12,13, 15,17,18,19,20,21
Kumari et al. 2004 [18]	Cohort	11	8,320 (69.4)	35-55[c]	Self rating: GHQ-D	Screening ADA, 1998	1.17 (0.8-1.7)[d, f] 1.03 (0.6-1.8)[d, g]	1,3,4,8,11,12,13, 15,22,25,28,29
Palinkas et al. 2004 [19]	Cohort	8	971 (43.0)	66 ± 9	Self rating: BDI	Screening ADA,1998	2.50 (1.29-4.87)[d]	1,2,8,12
Akker et al. 2004 [10]	Cohort	16[b]	68,004 (48.8)	38 ± 14	Diagnosis: ICHPPC-2, '83	Diagnosis DCGP, 1999	0.98 (0.79-1.21)[e]	1,2,4,8
Everson-Rose et al. 2004 [14]	Cohort	3	2,662 (0.0)	46 ± 3	Self Rating: CES-D	Screening ADA, 1998	1.46 (0.90-2.36)[d]	1,3,5,9,12,26,27

ADA, American Diabetes Association; BDI, Beck Depression Inventory; CES-D, Center for Epidemiological Studies Depression Scale; DCGP, Dutch College of General Practitioners; DIS, Diagnostic Interview Schedule; GHQ-D, General Health Questionnaire Depression; GWB-DS; General Well-Being Depressive Symptoms; ICHPPC, International Classification of Health Problems in Primary Care; MHI, Mental Health Index; SDS, Self-rating Depression Scale; VES, Vital Exhaustion Scale; WHO, World Health Organization.

[a] numbers in square brackets: reference citations. [b] mean duration of follow-up. [c] mean age not described. [d] odds ratio. [e] hazard ratio. [f] subgroup analysis of male subjects. [g] subgroup analysis of female subjects. [h] 1, age; 2, sex; 3, race; 4, socioeconomic status; 5, education; 6, occupation; 7, shift work; 8, body mass index; 9, waist circumference; 10, waist-to-hip ratio; 11, height; 12, physical activity; 13, smoking; 14, alcohol consumption; 15, systolic blood pressure; 16, history of hypertension; 17, HDL cholesterol; 18, triglycerides; 19, fasting insulin; 20, fasting glucose; 21, caloric intake; 22, ECG abnormalities; 23, menopausal status; 24, chronic medical conditions; 25, life events; 26, use of medication for depression; 27, study site; 28, length of follow-up; 29 family history of diabetes.

items with an abstract and written in English, resulted in 1722 articles. After careful selection, eleven studies appeared to have studied the relation between depression and onset of type 2 diabetes longitudinally [10-20]. Searching the on-line PsycInfo database yielded no additional studies. One of the eleven studies was excluded [17] because there was insufficient information in the article to calculate a relative risk. Two studies used data from the same cohort [12,20], the National Health and Nutrition Examination Survey (NHANES), and the most recent publication was included [12]. All studies excluded prevalent diabetes at baseline.

The search for unpublished work among members of the PSAD and BRIDGE resulted in 20 responses to our e-mail but this yielded no additional studies that met our inclusion criteria.

The extracted data of the nine studies included in the meta-analysis are presented in Table 1. One study reported relative risks separately for men and women [18]. These two relative risks were pooled using the random effects model and this pooled relative risk (1.12) and its 95% CI (0.82-1.53) were used in further analyses.

Meta-analysis

The forest plot shows the relative risk and 95% CI of each study and the pooled relative risk of both the fixed effects model and the random effects model (Figure 1). The pooled relative risk (95% CI) was 1.26 (1.13-1.39) using the fixed effects model and 1.37 (1.14-1.63) using the random effects model.

The funnel plot to detect publication bias showed some asymmetry, as six studies lay above and three studies below the line representing the pooled relative risk (Figure 2). Studies with a large standard error (small sample size) and small relative risk are missing in the graph. This could indicate publication bias as studies showing small (or no) associations and large CIs are probably less often submitted by authors and less often published by editors. The Begg-adjusted rank correlation test for publication bias resulted in a p-value of 0.10.

The forest plot showed heterogeneity between the studies, as the CIs of the different studies appeared to have no or only partial overlap (Figure 1). The Cochran's Q test was statistically significant (Q=18.264; p=0.02), indicating heterogeneity. Because of this heterogeneity the pooled relative risk resulting from the fixed effects model should be disregarded. In further analyses only random effects modelling was performed.

Figure 1 Forest plot showing the relative risk and 95% CI of each study and the pooled relative risk (RR) and 95% CI using both the fixed effects model and random effects model

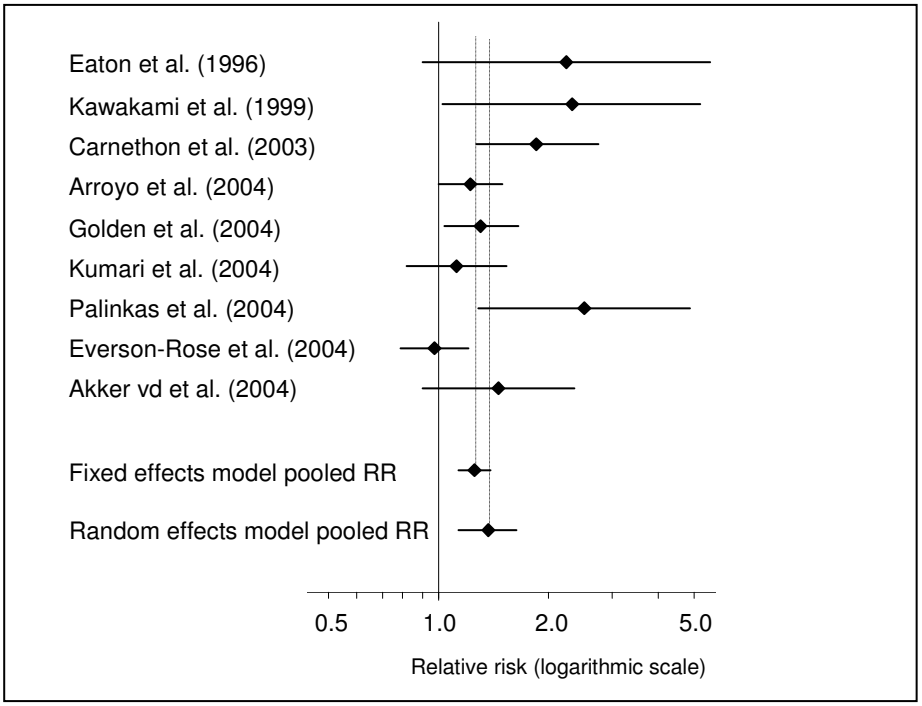
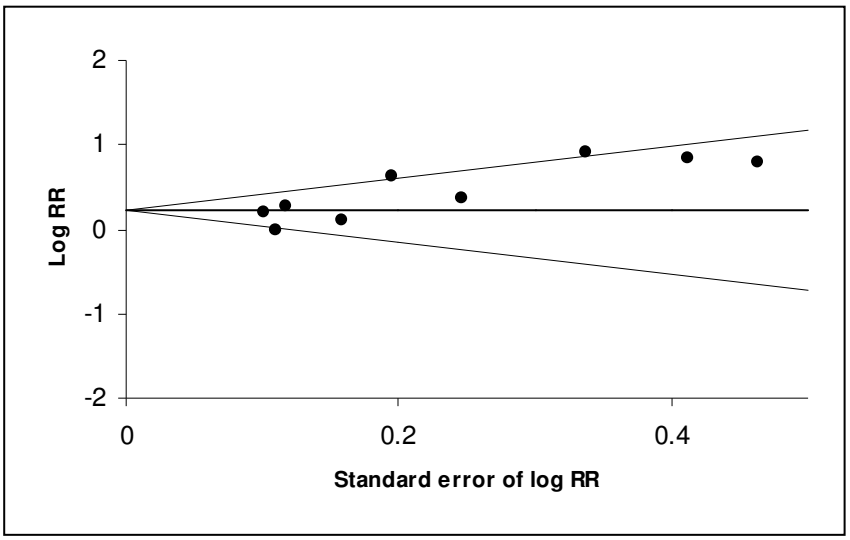


Figure 2 Funnel plot showing that studies with a large standard error (small sample size) and low relative risk (RR) are missing, which indicates publication bias



Subgroup and sensitivity analyses

The pooled relative risk (95% CI) of studies that relied on self-reported diabetes to exclude prevalent diabetes at baseline, and thus did not control for undetected diabetes at baseline

[10-13,16,18], was 1.32 (1.04-1.66) (Table 2). The pooled relative risk (95% CI) of studies that did control for undetected diabetes by screening all subjects for high blood glucose [14,15,19] was slightly higher, namely 1.54 (1.07-2.22).

Table 2 Pooled relative risks (using random effects meta-analysis) stratified by exclusion of undetected diabetes at baseline and method of diabetes assessment at follow-up

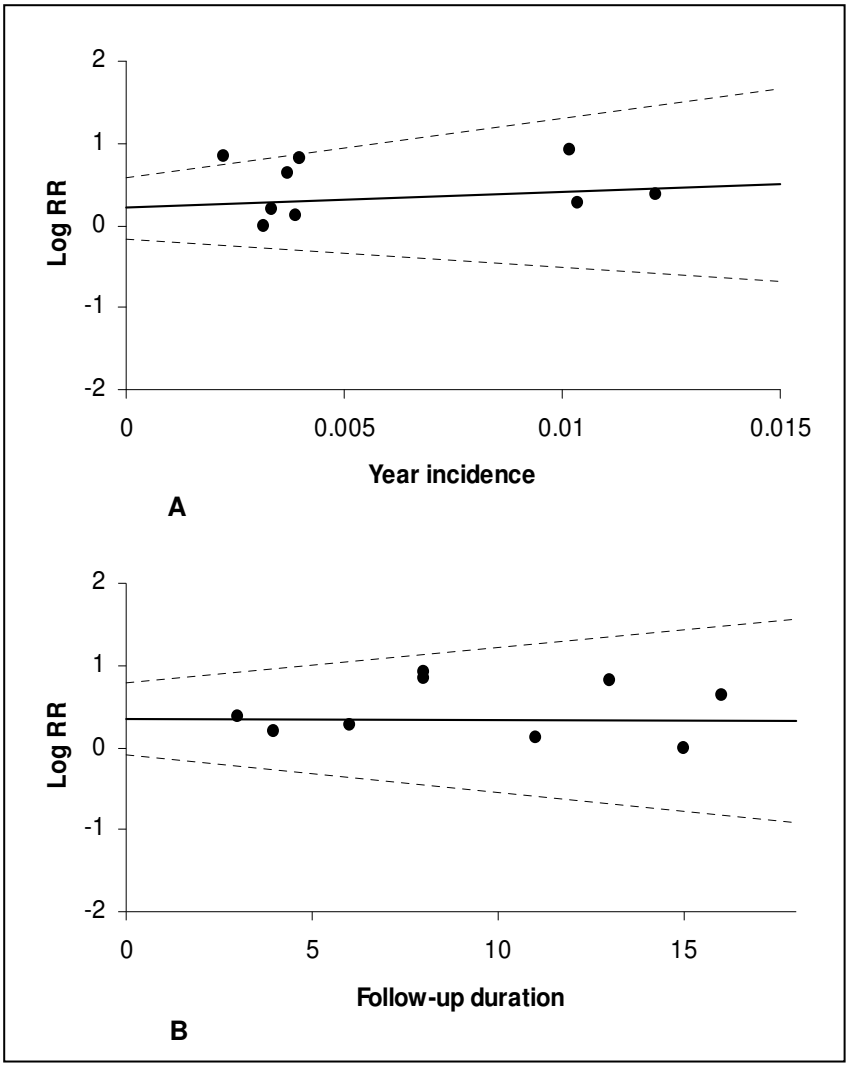
	N	Pooled relative risk (95% CI)
Overall	9	1.37 (1.14-1.63)
Exclusion of undetected diabetes at baseline		
No	6	1.32 (1.04-1.66)
Yes	3	1.51 (1.11-2.06)
Method of diabetes assessment at follow-up		
Self-report	4	1.32 (0.98-1.78)
Screening	5	1.43 (1.12-1.81)

The four studies [10-13] that determined type 2 diabetes at follow-up by means of self-report had a pooled relative risk (95% CI) of 1.32 (0.98-1.78) (Table 2). The studies that assessed diabetes onset by measuring glucose levels [14-16,18,19] instead of self-report, had a pooled relative risk (95% CI) of 1.43 (1.12-1.81).

The overall risk of diabetes in each study, i.e. the overall incidence per year, was plotted against the natural logarithm of the relative risk of each study (Figure 3A). The regression coefficient (95% CI) was 18.6 (-35.9-73.1) but not significantly different from zero, which means there is no relation between overall risk of diabetes and relative risk. Also, no relation was found between follow-up duration and relative risk, as can be seen in the plot (Figure 3B). The regression coefficient (95% CI) was -0.0018 (-0.045-0.042).

A sensitivity analysis was performed to determine whether excluding the Dutch study [10] influenced the pooled relative risk, as this particular study used routine care data and doctor's diagnosis to measure depression instead of screening by means of a diagnostic interview or questionnaire. When this study [10] was excluded the pooled relative risk (95% CI) was 1.44 (1.21-1.71), which is slightly larger than the pooled relative risk if that study was included in the analysis. When the three studies with the smallest sample size [13,16,19] were excluded the pooled relative risk (95% CI) was 1.24 (1.06-1.46). To check the influence of our decision to include the most recent publication when multiple publications from the same study population were available, we excluded the study of Carnethon et al. and included the study of Saydah et al. This resulted in a pooled relative risk (95% CI) of 1.26 (1.08-1.47).

Figure 3 Relation between natural logarithm of relative risk (log RR) and overall incidence per year (A) and follow-up duration (B) with regression line (solid line) and upper and lower limits of 95% CI (dashed lines)



Influence of adjusting for confounders

Table 3 presents an overview of unadjusted and adjusted relative risks (95% CI) which were reported by each study, with a detailed description of the different sets of confounders that were used. Unfortunately, as the studies adjusted for many different sets of confounders, it is impossible to quantify the exact influence of adjusting for certain confounders. Qualitatively, no association was seen between adjustment for certain confounders and the magnitudes of the relative risks, as some relative risks decreased after adjusting for certain confounders and others increased.

Table 3 Overview of unadjusted and adjusted relative risks reported in each study, with a description of all confounders that were used

Study	Relative risk (95% CI)		Adjustment for confounders[c]
Eaton et al.	1.58 (0.71-3.51)		-
1996 [13]	2.05 (0.85-4.94)		1,2,3
	2.23 (0.90-5.55)		1,2,3,8
Kawakami et al.	2.04 (0.93-4.45)		-
1999 [16]	2.32 (1.06-5.08)		1
	2.31 (1.03-5.20)		1,5,6,7,8,12,13,14,24,29
Carnethon et al.	2.52 (1.73-3.67)		1,2,3
2003 [12]	1.86 (1.27-2.71)		1,2,3,8,12,13,14
Arroyo et al.	1.47 (1.20-1.79)		-
2004 [11]	1.55 (1.27-1.90)		1
	1.36 (1.11-1.67)		1,8
	1.22 (1.00-1.50)		1,8,12,13,14,16,23,29
Golden et al.	1.63 (1.31-2.02)		1,2,3,5
2004 [15]	1.38 (1.10-1.73)		1,2,3,5,8,10,15,17,18,19,20
	1.51 (1.21-1.89)		1,2,3,5,12,13,21
	1.28 (1.02-1.60)		1,2,3,5,8,10,12,13,21
	1.31 (1.04-1.64)		1,2,3,5,8,10,12,13,15,17,18,19,20,21
Kumari et al.	1.17 (0.8-1.7)[a]	1.08 (0.6-1.9)[b]	1,3,4,22,28
2004 [18]	1.17 (0.8-1.7)[a]	1.03 (0.6-1.8)[b]	1,3,4,8,11,12,13,15,22,25,28,29
Palinkas et al.	2.50 (1.29-4.87)		1,2,8,12
2004 [19]			
Akker et al.	1.41 (1.15-1.73)		-
2004 [10]	1.04 (0.84-1.28)		1,2,4
	0.98 (0.79-1.21)		1,2,4,8
Everson-Rose et al.	1.66 (1.05-2.61)		1,3,5,26,27
2004 [14]	1.46 (0.90-2.36)		1,3,5,9,12,26,27

[a] subgroup analysis of male subjects. [b] subgroup analysis of female subjects. [c] 1, age; 2, sex; 3, race; 4, socioeconomic status; 5, education; 6, occupation; 7, shift work; 8, body mass index; 9, waist circumference; 10, waist-to-hip ratio; 11, height; 12, physical activity; 13, smoking; 14, alcohol consumption; 15, systolic blood pressure; 16, history of hypertension; 17, HDL cholesterol; 18, triglycerides; 19, fasting insulin; 20, fasting glucose; 21, caloric intake; 22, ECG abnormalities; 23, menopausal status; 24, chronic medical conditions; 25, life events; 26, use of medication for depression; 27, study site; 28, length of follow-up; 29, family history of diabetes.

Discussion

To our knowledge, this study represents the first application of meta-analysis of literature regarding depression as a risk factor for the onset of type 2 diabetes mellitus.

Results of this meta-analysis of nine longitudinal studies suggest that adults with depression or high depressive symptoms have a 37% increased risk of developing type 2 diabetes compared with those who are not depressed or have low depressive symptoms.

Heterogeneity between studies regarding relative risks could not be explained by: 1) whether studies controlled for undetected diabetes at baseline; 2) the method of diabetes assessment at follow-up; 3) the baseline overall risk of diabetes in the study population; or 4) follow-up duration. Also, adjustment for several confounders did not explain differences in effect sizes between studies.

Before drawing a conclusion on the findings of this meta-analysis, we will discuss several biases which may have confounded our results. First, reversed causality could be an issue. In reversed causality, presymptomatic persons with diabetes develop depression. These subjects are more likely to develop symptomatic diabetes and this will overestimate the effect size. However, we believe that this hypothesis is less probable, as we have found that studies that exclude cases with undetected diabetes at baseline showed a pooled relative risk similar to the overall pooled relative risk. Second, ascertainment or diagnostic bias could play a role in explaining the results of the present meta-analysis. Subjects with depression tend to visit their doctor more often and may thus be more likely to be recognized as having diabetes [21]. This bias could have occurred particularly in studies that relied on self-reported diabetes at follow-up. However, the pooled relative risk of studies that assessed diabetes by measuring glucose levels (as opposed to self-report or doctor's diagnosis) appeared to be similar to the overall pooled relative risk. These findings do not support the notion that ascertainment bias explains the results of our study. Third, although all studies adjusted for multiple potential confounders, residual confounding may have influenced our findings. Given the fact that most of the studies adjusted for a considerable amount of confounders (median = 7; range = 4-14), we consider this as less likely. In contrast, overcorrection may have occurred. It may be true that some studies adjusted for intermediate rather than confounding factors, resulting in an underestimation of the pooled relative risk. Fourth, another potential bias of the present meta-analysis is publication bias, which is a threat to the validity of every systematic review. We tried to minimise publication bias by asking members of relevant study groups whether they had any unpublished/rejected results of studies investigating the relation between depression and the onset of type 2 diabetes. Still, the funnel plot did show some asymmetry, as studies with a small sample size and low relative risk were missing, which indicates publication bias. However, even after excluding the three smallest studies [13,16,19] a significant, pooled relative risk was found. In sum, reversed causality, ascertainment bias, confounding factors and publication bias do not seem to explain the relationship found in this meta-analysis.

Another potential problem in this meta-analysis is that each study used a different method to assess depression. These different methods can be categorized into four

hierarchical groups. First, only one study used a diagnostic interview schedule [13], which is the gold standard for diagnosis of major depression. Second, four studies used validated depression severity scales, the Zung Depression Scale, the General Health Questionnaire, the Beck Depression Inventory and the CES-Depression scale [14,16,18,19]. In these studies validated cut off scores were used to define levels of depressive affect. The sensitivity and specificity of these measures proved to be acceptable [22]. As a third measure of depression, three studies used semi-depression severity scales: the General Well Being Depression Scale, the Mental Health Index of the SF-36 and the Vital Exhaustion Scale [11,12,15]. These measures were not designed to measure depression severity but are commonly used as a proximal measure of negative affect. The relative risks found in the studies that used these semi-depression scales were similar to the relative risks of the studies that used the more sophisticated scales and showed the same direction of effect. Finally, one study used general practitioner's diagnosis of depression [10]. It has been reported that depressive symptoms are not recognized in about half of attending patients with depressive disorders in UK general practice [23] and this under recognition of depression would result in underestimation of the effect size between depression and onset of diabetes. Therefore, we performed a sensitivity analysis excluding this study which resulted in a slightly higher pooled relative risk: 1.44 (1.21-1.71).

A second issue in the method of assessment of depression is that two studies made three categories of depression level [12,16] and one study divided the depression scores into quartiles [15]. This might have influenced our results, as we used the relative risk of the highest versus the lowest group of depressive symptoms. In these three studies the relative risk of the highest versus the lowest group is larger than the relative risks for the other categories. An additional stratified analysis also showed that the pooled relative risk for these three studies is somewhat higher (1.59 [1.16-2.17]) than the pooled relative risk of the other studies (1.26 [1.02-1.56]). However, we still believe that the relative risk of the lowest versus the highest category of depression represents best the difference between 'absence/presence' depression.

In the literature, several hypotheses have been described regarding the pathophysiological mechanisms that could explain the increased risk of type 2 diabetes in depressed subjects. First, the hypothesis of increased activity of the hypothalamic pituitary adrenocortical (HPA) axis and sympathetic nervous system will be discussed. Depression is associated with increased activity of the HPA axis and sympathetic nervous system [24], resulting in increased cortisol release and increased release of the catecholamines epinephrine and norepinephrine. Cortisol is a stress hormone, which stimulates glucose

production, increases lipolysis and circulating free fatty acids, decreases insulin secretion from beta cells and decreases the sensitivity to insulin [24-27]. It is postulated that a chronically high cortisol level, which is a feature of about 50% of depressed patients, results in obesity, insulin resistance and type 2 diabetes [24,28,29]. Some studies found evidence for this hypothesis [27,28]. Epinephrine generates responses on glucose and fat metabolism similar to those of cortisol [26], also possibly resulting in insulin resistance and type 2 diabetes.

The credibility of this hypothesis is further strengthened by findings on other medical problems that are accompanied by hypercortisolemia. For example, Cushing's syndrome, sleeping disorders, work stress and schizophrenia [30-33] appeared to be associated with an increased level of cortisol and also with an increased risk of type 2 diabetes and insulin resistance, although studies on sleep disorders showed inconsistent results regarding risk of diabetes [33-37].

A second hypothesis is that dysregulation of the immune system plays a role in the relationship between depression and increased risk of type 2 diabetes. Both depression and type 2 diabetes are found to be associated with increased C-reactive protein, TNF-alpha and proinflammatory cytokines including IL-6 [38-42]. A contradiction between this hypothesis and the first hypothesis is that cortisol inhibits inflammation and immune response, whereas depression is associated with both elevated cortisol and increased inflammatory markers. A recent finding possibly explains this contradiction by showing that melancholic depressed patients had increased HPA axis activity and no signs of inflammation while non-melancholic depressed patients did show signs of inflammation and normal HPA axis activity [43].

Finally, a low intake or impaired metabolism of omega-3 polyunsaturated fatty acids (PUFA) could contribute to both depression and type 2 diabetes. Omega-3 PUFA have direct and indirect actions on cerebral function and depletion of these fatty acids is clearly associated with psychiatric illness, including depression [44,45]. In addition, there is evidence that a low intake of omega-3 PUFA is associated with an increased risk of type 2 diabetes, but these results were less convincing [44].

It is known that the most important risk factor for type 2 diabetes is obesity [46,47], and that physical inactivity further increases the risk, independently of obesity [48]. In view of our findings, depression could be regarded as an additional risk factor for type 2 diabetes, comparable in size to smoking and physical activity [47,49]. Clinicians should be made aware of the fact that depressive affect might be an additional risk factor for type 2 diabetes as this makes adequate detection and treatment of depression even more important than it already is. Assessing fasting glucose and advising exercise in depressed patients might also prevent type 2 diabetes.

In conclusion, this meta-analysis suggests that depression is a risk factor for the onset of type 2 diabetes mellitus, comparable in size to smoking and physical activity. However, further well-designed research with adequate control for confounding factors is needed to establish the exact size of the relationship. The influence of length of depression or change in depression over time on risk of type 2 diabetes should be studied especially. Furthermore, research is warranted to elucidate the pathophysiological mechanisms underlying the association. With the expectation of more than 100 million new cases of type 2 diabetes in the coming two decades, prevention becomes more important every day. Whether prevention of depression or the treatment of depressed people can truly prevent or delay the onset of type 2 diabetes mellitus remains to be tested in long-term intervention studies.

References

1. Wild S, Roglic G, Green A, Sicree R, King H. Global Prevalence of Diabetes: Estimates for the Year 2000 and Projections for 2030. *Diabetes Care* 2004;27:1047-53.
2. Murray CJ and Lopez AD. Alternative Projections of Mortality and Disability by Cause 1990-2020: Global Burden of Disease Study. *Lancet* 1997;349:1498-504.
3. Anderson RJ, Freedland KE, Clouse RE, Lustman PJ. The Prevalence of Comorbid Depression in Adults With Diabetes: a Meta-Analysis. *Diabetes Care* 2001;24:1069-78.
4. Musselman DL, Betan E, Larsen H, Phillips LS. Relationship of Depression to Diabetes Types 1 and 2: Epidemiology, Biology, and Treatment. *Biol Psychiatry* 2003;54:317-29.
5. Talbot F and Nouwen A. A Review of the Relationship Between Depression and Diabetes in Adults: Is There a Link? *Diabetes Care* 2000;23:1556-62.
6. Rubin RR and Peyrot M. Was Willis Right? Thoughts on the Interaction of Depression and Diabetes. *Diabetes Metab Res Rev* 2002;18:173-5.
7. Willis T. *Pharmaceutice Rationalis Sive Diatriba de Medicamentorum Operationibus in Humano Corpore*. E Theatro Sheldoniano, M.DC.LXXV, Oxford, 1674.
8. Fleiss JL and Gross AJ. Meta-Analysis in Epidemiology, With Special Reference to Studies of the Association Between Exposure to Environmental Tobacco Smoke and Lung Cancer: a Critique. *J Clin Epidemiol* 1991;44:127-39.
9. DerSimonian R and Laird N. Meta-Analysis in Clinical Trials. *Control Clin Trials* 1986;7:177-88.
10. Akker van den M, Schuurman A, Metsemakers J, Buntinx F. Is Depression Related to Subsequent Diabetes Mellitus? *Acta Psychiatr Scand* 2004;110:178-83.
11. Arroyo C, Hu FB, Ryan LM, Kawachi I, Colditz GA, Speizer FE et al. Depressive Symptoms and Risk of Type 2 Diabetes in Women. *Diabetes Care* 2004;27:129-33.
12. Carnethon MR, Kinder LS, Fair JM, Stafford RS, Fortmann SP. Symptoms of Depression As a Risk Factor for Incident Diabetes: Findings From the National Health and Nutrition Examination Epidemiologic Follow-Up Study, 1971-1992. *Am J Epidemiol* 2003;158:416-23.
13. Eaton WW, Armenian H, Gallo J, Pratt L, Ford DE. Depression and Risk for Onset of Type II Diabetes. A Prospective Population-Based Study. *Diabetes Care* 1996;19:1097-102.

14. Everson-Rose SA, Meyer PM, Powell LH, Pandey D, Torrens JI, Kravitz HM et al. Depressive Symptoms, Insulin Resistance, and Risk of Diabetes in Women at Midlife. *Diabetes Care* 2004;27:2856-62.
15. Golden SH, Williams JE, Ford DE, Yeh HC, Paton SC, Nieto FJ et al. Depressive Symptoms and the Risk of Type 2 Diabetes: the Atherosclerosis Risk in Communities Study. *Diabetes Care* 2004;27:429-35.
16. Kawakami N, Takatsuka N, Shimizu H, Ishibashi H. Depressive Symptoms and Occurrence of Type 2 Diabetes Among Japanese Men. *Diabetes Care* 1999;22:1071-6.
17. Kessing LV, Nilsson FM, Siersma V, Andersen PK. Increased Risk of Developing Diabetes in Depressive and Bipolar Disorders? *J Psychiatr Res* 2004;38:395-402.
18. Kumari M, Head J, Marmot M. Prospective Study of Social and Other Risk Factors for Incidence of Type 2 Diabetes in the Whitehall II Study. *Arch Intern Med* 2004;164:1873-80.
19. Palinkas LA, Lee PP, Barrett-Connor E. A Prospective Study of Type 2 Diabetes and Depressive Symptoms in the Elderly: the Rancho Bernardo Study. *Diabet Med* 2004;21:1185-91.
20. Saydah SH, Brancati FL, Golden SH, Fradkin J, Harris MI. Depressive Symptoms and the Risk of Type 2 Diabetes Mellitus in a US Sample. *Diabetes Metab Res Rev* 2003;19:202-8.
21. Ciechanowski PS, Katon WJ, Russo JE. Depression and Diabetes: Impact of Depressive Symptoms on Adherence, Function, and Costs. *Arch Intern Med* 2000;160:3278-85.
22. Mulrow CD, Williams JW, Jr., Gerety MB, Ramirez G, Montiel OM, Kerber C. Case-Finding Instruments for Depression in Primary Care Settings. *Ann Intern Med* 1995;122:913-21.
23. Gilbody SM, Whitty PM, Grimshaw JM, Thomas RE. Improving the Detection and Management of Depression in Primary Care. *Qual Saf Health Care* 2003;12:149-55.
24. Bjorntorp P. Do Stress Reactions Cause Abdominal Obesity and Comorbidities? *Obes Rev* 2001;2:73-86.
25. Ramasubbu R. Insulin Resistance: a Metabolic Link Between Depressive Disorder and Atherosclerotic Vascular Diseases. *Med Hypotheses* 2002;59:537-51.
26. Genuth SM. *The adrenal glands.* In *Physiology*. (Ed. R. M. Berne and M. N. Levy) pp. 930-964, Mosby, St. Louis, Missouri, 1998.
27. Weber-Hamann B, Hentschel F, Kniest A, Deuschle M, Colla M, Lederbogen F et al. Hypercortisolemic Depression Is Associated With Increased Intra-Abdominal Fat. *Psychosom Med* 2002;64:274-7.
28. Bjorntorp P, Holm G, Rosmond R. Hypothalamic Arousal, Insulin Resistance and Type 2 Diabetes Mellitus. *Diabet Med* 1999;16:373-83.
29. Maes M, Minner B, Suy E, Vandervorst C, Raus J. Coexisting Dysregulations of Both the Sympathoadrenal System and Hypothalamic-Pituitary-Adrenal-Axis in Melancholia. *J Neural Transm Gen Sect* 1991;85:195-210.
30. Buckley TM and Schatzberg AF. On the Interactions of the Hypothalamic-Pituitary-Adrenal (HPA) Axis and Sleep: Normal HPA Axis Activity and Circadian Rhythm, Exemplary Sleep Disorders. *J Clin Endocrinol Metab* 2005;90:3106-14.
31. Lundberg U. Stress Hormones in Health and Illness: The Roles of Work and Gender. *Psychoneuroendocrinology* 2005;30:1017-21.

32. Morris D and Grossman A. The Medical Management of Cushing's Syndrome. *Ann N Y Acad Sci* 2002;970:119-33.
33. Ryan MC, Collins P, Thakore JH. Impaired Fasting Glucose Tolerance in First-Episode, Drug-Naive Patients With Schizophrenia. *Am J Psychiatry* 2003;160:284-9.
34. Ayas NT, White DP, Al Delaimy WK, Manson JE, Stampfer MJ, Speizer FE et al. A Prospective Study of Self-Reported Sleep Duration and Incident Diabetes in Women. *Diabetes Care* 2003;26:380-4.
35. Kawakami N, Araki S, Takatsuka N, Shimizu H, Ishibashi H. Overtime, Psychosocial Working Conditions, and Occurrence of Non-Insulin Dependent Diabetes Mellitus in Japanese Men. *J Epidemiol Community Health* 1999;53:359-63.
36. Meisinger C, Heier M, Loewel H. Sleep Disturbance As a Predictor of Type 2 Diabetes Mellitus in Men and Women From the General Population. *Diabetologia* 2005;48:235-41.
37. Nilsson PM, Roost M, Engstrom G, Hedblad B, Berglund G. Incidence of Diabetes in Middle-Aged Men Is Related to Sleep Disturbances. *Diabetes Care* 2004;27:2464-9.
38. Dentino AN, Pieper CF, Rao MK, Currie MS, Harris T, Blazer DG et al. Association of Interleukin-6 and Other Biologic Variables With Depression in Older People Living in the Community. *J Am Geriatr Soc* 1999;47:6-11.
39. Kiecolt-Glaser JK and Glaser R. Depression and Immune Function: Central Pathways to Morbidity and Mortality. *J Psychosom Res* 2002;53:873-6.
40. Maes M, Bosmans E, De Jongh R, Kenis G, Vandoolaeghe E, Neels H. Increased Serum IL-6 and IL-1 Receptor Antagonist Concentrations in Major Depression and Treatment Resistant Depression. *Cytokine* 1997;9:853-8.
41. Pradhan AD, Manson JE, Rifai N, Buring JE, Ridker PM. C-Reactive Protein, Interleukin 6, and Risk of Developing Type 2 Diabetes Mellitus. *JAMA* 2001;286:327-34.
42. Schmidt MI, Duncan BB, Sharrett AR, Lindberg G, Savage PJ, Offenbacher S et al. Markers of Inflammation and Prediction of Diabetes Mellitus in Adults (Atherosclerosis Risk in Communities Study): a Cohort Study. *Lancet* 1999;353:1649-52.
43. Kaestner F, Hettich M, Peters M, Sibrowski W, Hetzel G, Ponath G et al. Different Activation Patterns of Proinflammatory Cytokines in Melancholic and Non-Melancholic Major Depression Are Associated With HPA Axis Activity. *J Affect Disord* 2005;87:305-11.
44. Pouwer F, Nijpels G, Beekman AT, Dekker JM, Dam RM, Heine RJ et al. Fat Food for a Bad Mood. Could We Treat and Prevent Depression in Type 2 Diabetes by Means of Omega-3 Polyunsaturated Fatty Acids? A Review of the Evidence. *Diabet Med* 2005;22:1465-75.
45. Hallahan B and Garland MR. Essential Fatty Acids and Mental Health. *Br J Psychiatry* 2005;186:275-7.
46. Colditz GA, Willett WC, Rotnitzky A, Manson JE. Weight Gain As a Risk Factor for Clinical Diabetes Mellitus in Women. *Ann Intern Med* 1995;122:481-6.
47. Hu FB, Manson JE, Stampfer MJ, Colditz G, Liu S, Solomon CG et al. Diet, Lifestyle, and the Risk of Type 2 Diabetes Mellitus in Women. *N Engl J Med* 2001;345:790-7.
48. Manson JE, Rimm EB, Stampfer MJ, Colditz GA, Willett WC, Krolewski AS et al. Physical Activity and Incidence of Non-Insulin-Dependent Diabetes Mellitus in Women. *Lancet* 1991;338:774-8.
49. Eliasson B. Cigarette Smoking and Diabetes. *Prog Cardiovasc Dis* 2003;45:405-13.

Chapter 2.2

No increased incidence of diabetes in antidepressant users

Abstract

Background

This study investigated whether the association between depression and diabetes was influenced by the presence of chronic somatic disease. To distinguish between depression and other psychosocial complaints, we studied the onset of diabetes in antidepressant (AD) users and benzodiazepine (BD) users, respectively.

Methods

From the PHARMO database, which includes complete drug prescription data, we identified subjects using: (1) no ADs and no BDs; (2) AD but no BD; (3) BD but no AD; and (4) AD and BD. A total of 60,516 individuals (age: 45.5 +/- 17 years; 42.1% men) were followed from their first prescription for AD or BD until end of registration or a first prescription for antidiabetic drugs.

Results

The crude incidence rate in AD but no BD users was not increased compared with no AD and no BD users. After adjustment for age, sex and chronic diseases, the hazard ratios (95% CI) were 1.05 (0.88-1.26) for AD but no BD users, 1.21 (1.02-1.43) for BD but no AD users and 1.37 (1.12-1.68) for AD and BD users compared with no AD and no BD users.

Conclusions

We did not find an increased risk of diabetes in individuals using ADs. The association between BD use and diabetes was partly explained by chronic somatic comorbidity.

Introduction

Depression and diabetes are both common conditions in today's society and have a large impact on the well-being and functioning of patients [1]. A meta-analysis of 20 cross-sectional studies showed that the prevalence of depression is doubled in patients with diabetes compared with subjects without diabetes [2], indicating that depression and diabetes often co-occur. The temporal direction of the association between depression and diabetes, however, is not clear.

Depression may be a risk factor for diabetes. A recent meta-analysis of nine longitudinal studies showed that depressed individuals have a 35% increased risk of developing diabetes compared with non-depressed individuals [3]. The reason for this increased risk is not known. Hypotheses about pathophysiological mechanisms linking depression and diabetes include disturbance of the hypothalamic-pituitary-adrenocortical axis, disturbance of the sympathetic nervous system and dysregulation of the immune system, but none of these hypotheses have been confirmed by literature yet [4,5].

An alternative may be that the presence of chronic somatic diseases explains (part of) the association between depression and diabetes. Most prospective studies that assessed depression as a risk factor for diabetes did not evaluate whether chronic diseases influenced the association. Only one prospective study, which still found an increased risk of diabetes in depressed individuals, adjusted for hypertension, coronary heart disease, cerebrovascular disease and metabolic disease [6]. In addition, two cross-sectional studies found that diabetes and depression were associated in individuals with chronic diseases but not in those without chronic diseases [7,8].

The aim of this study was to investigate to what extent the association between depression and onset of diabetes was influenced by the presence of chronic somatic disease. A large pharmacy database was used to study this research question. We studied the onset of diabetes in antidepressant (AD) users and benzodiazepine (BD) users to distinguish between depression and other psychosocial complaints, respectively.

Methods

Design

A historical cohort study was performed in which we compared users of ADs, users of BDs, and users of both AD and BD with users of neither AD nor BD with regard to starting any glucose lowering drug (oral hypoglycemic agents and/or insulin) during follow-up.

Data source

The PHARMO database was used to identify our cohort. This database is described in detail elsewhere [9]. In short, the PHARMO database is an anonymous pharmacy registry database that comprises all pharmacy-dispensing records of all residents of about fifty Dutch municipalities, counting for about two million patient histories. As virtually all patients in the Netherlands are registered with a single community pharmacy, independent of prescriber, pharmacy records are almost complete with regard to prescription drugs. In the Netherlands, ADs, BDs, oral hypoglycemic agents and insulin are only available as prescription drugs. Therefore, pharmacy data will cover all use of these drugs. Available variables in the PHARMO database include sex, date of birth, dispensed drugs (coded according to the Anatomical Therapeutic Chemical (ATC) classification), drug dispensing date, amount of drug dispensed, prescribed dosage regimen and prescriber. Pharmacy data from 1996 until 2003 were used in this study.

Study population

We first selected the source population from the PHARMO database. This source population comprised all individuals with at least two subsequent prescriptions of any AD (ATC code = N06A*) and/or any BD (ATC code = N05BA*, N05CD* or N05CF*) in the period from 1 January 1996, through 31 December 2003. From this source population, three exposure groups were formed: participants using AD but no BD, participants using BD but no AD, and those using both AD and BD. The nonexposed group was formed by taking a random sample from the PHARMO database of participants without any prescription for AD and BD in the period from 1 January 1996, through 31 December 2003. The index date was defined as the date of the first prescription for either AD or BD. In the group with no prescriptions for AD and BD, the index date was randomly assigned.

Individuals were included in the study population if they were 18 years or older at the index date; were new users of AD or BD (does not apply to users of neither AD nor BD); had at least two prescriptions of AD or BD in the year after the index date (does not apply to users of neither AD nor BD); and had follow-up data for at least 90 days after the index date. New use of AD or BD was defined as a first prescription for AD or BD in the study period and no prescription of AD or BD in the preceding year. We wanted to include 'regular' users and not individuals with just one prescription for a single occasion and therefore we included individuals only if they had at least two prescriptions of AD or BD in the year after the index date. Participants were excluded if information about date of birth or sex was not available. Moreover, prevalent cases of diabetes were excluded. A prevalent case of diabetes was

defined as an individual with a prescription for any glucose lowering drug (ATC code = A10A* and A10B*) at or before the index date.

The following four exposure groups were used in the analyses: (1) participants using no AD and no BD (n=23,919); (2) participants using AD but no BD (n=18,507); (3) participants using BD but no AD (n=12,117); and (4) participants using AD and BD (n=5973).

Outcome

The outcome of interest was the initiation of diabetes medication, defined as the first prescription for any glucose-lowering drug, either oral hypoglycemic agents and/or insulin, after the index date.

Covariates

Age at index date, sex and the Chronic Disease Score (CDS) were used as covariates. The CDS is a measure of the chronic disease status among drug users, and can be considered as an indicator of an individual's morbidity and overall health status. Exposure to various prescription drugs has been shown to be a valid measure of chronic diseases [10]. The CDS includes the major chronic diseases such as heart disease, respiratory illness, cancer, ulcers and high cholesterol. The CDS was calculated over the period of one year before the index date.

Data-analysis

In each exposure group, the crude incidence rate of initiation of diabetes treatment was calculated by dividing the number of diabetes cases by person-years. For cases, person-years were calculated as the time between index date and start of a glucose lowering drug. For non-cases, person-years were calculated as the time between index date and end of database registration or end of study (31 December 2003). By means of Cox regression analysis, hazard ratios (HR) and 95% confidence intervals (CI) of initiation of diabetes treatment were calculated for AD but no BD users, BD but no AD users and users of both AD and BD, compared with users of neither AD nor BD. First, we included age and sex into the model. Second, we additionally included the CDS into the model to investigate whether (part of) the associations under study could be explained by chronic diseases.

To investigate the association between chronic diseases and initiation of diabetes treatment, a stratified analysis for the Chronic Disease Score was performed. In addition, the age-adjusted incidence rates in the four exposure groups across strata of the CDS were calculated with the direct method by using the total study population as a standard

population.

All analyses were performed with SPSS version 12.0.1 for Windows (SPSS Inc., Chigago, Illinois, USA).

Results

Baseline characteristics of the four exposure groups are presented in Table 1. A total of 60,516 participants were included in the analysis, of whom 42.1% were men and mean age (SD) was 45.5 (17) years. Participants using BD but no AD were somewhat older and participants using AD and/or BD were more likely to be women. All chronic diseases were more prevalent in the participants using BD but no AD. Especially, heart disease was highly prevalent in these participants. As expected, the percentage of participants with a CDS of 0 was highest in those using neither AD nor BD. The number of participants excluded because of prevalent diabetes were 761 (3.1%), 940 (4.8%), 811 (6.3%) and 242 (3.9%) in those using neither AD nor BD, AD but no BD, BD but no AD and both AD and BD, respectively. These prevalences are in concordance with the expected prevalence of diabetes in a Dutch population with a mean age of 46 years. [11]

Table 1 Baseline characteristics of no antidepressant (AD) and no benzodiazepine (BD) users, AD but no BD users, BD but no AD users, both AD and BD users

	No AD no BD	**AD but no BD**	**BD but no AD**	**AD and BD**
N	23,919	18,507	12,117	5973
Age, mean (SD)	43.4 (17)	43.3 (17)	53.4 (18)	45.0 (16)
Male, %	48.3	36.4	41.9	35.8
Chronic diseases				
Heart disease, %	7.5	9.8	20.5	11.3
Respiratory illness, %	6.3	8.8	12.1	10.5
Cancer, %	0.8	1.0	2.2	1.1
Ulcers, %	6.6	13.0	17.1	16.1
High cholesterol, %	4.2	3.8	7.2	4.6
CDS score, %				
0	75.6	64.9	52.5	60.3
1-3	17.7	25.1	27.7	27.4
>=4	6.7	9.9	19.8	12.3

The crude incidence rate of initiation of diabetes treatment was highest in participants using BD but no AD (9.6/1000 person-years) and lowest in participants using neither AD nor

Table 2 Crude and adjusted hazard ratios (HR) of initiation of diabetes mellitus (DM) treatment among antidepressant (AD) but no benzodiazepine (BD) users, BD but no AD users and both AD and BD users, compared with no AD and no BD users

Exposure	N	DM cases (person-years)	Incidence rate	Crude HR (95% CI)	Adjusted* HR (95% CI)	Adjusted** HR (95% CI)
No AD no BD	23,919	252 (49,666)	5.1/1000 py	1.00	1.00	1.00
AD but no BD	18,507	247 (47,185)	5.2/1000 py	1.03 (0.86-1.22)	1.11 (0.93-1.33)	1.06 (0.89-1.26)
BD but no AD	12,117	329 (34,383)	9.6/1000 py	1.86 (1.58-2.20)	1.31 (1.11-1.55)	1.21 (1.02-1.44)
AD and BD	5973	152 (20,142)	7.5/1000 py	1.46 (1.19-1.78)	1.47 (1.20-1.80)	1.37 (1.12-1.68)

* adjusted for age and sex; ** adjusted for age, sex and Chronic Disease Score; AD: antidepressant; BD: benzodiazepine; DM: diabetes mellitus; HR: hazard ratio; CI: confidence interval

Table 3 Crude and adjusted hazard ratios (HR) of initiation of diabetes mellitus (DM) treatment among antidepressant (AD) but no benzodiazepine (BD) users, BD but no AD users and both AD and BD users compared with no AD and no BD users, stratified for Chronic Disease Score (CDS)

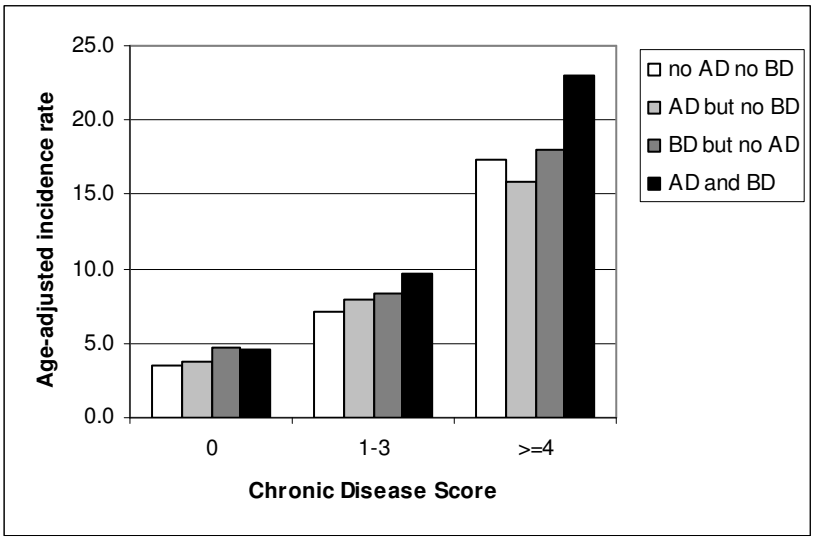
CDS	Exposure	N	DM cases (person-years)	Incidence rate	Age-adjusted incidence rate	Crude HR (95% CI)	Adjusted* HR (95% CI)
0	No AD no BD	18,075	127 (36,922)	3.4/1000 py	3.5/1000 py	1.00	1.00
	AD but no BD	12,020	101 (30,766)	3.3/1000 py	3.7/1000 py	0.94 (0.73-1.22)	1.07 (0.82-1.39)
	BD but no AD	6363	116 (18,499)	6.3/1000 py	4.7/1000 py	1.77 (1.38-2.29)	1.32 (1.02-1.71)
	AD and BD	3603	52 (12,102)	4.3/1000 py	4.6/1000 py	1.20 (0.87-1.66)	1.27 (0.92-1.76)
1-3	No AD no BD	4244	67 (9351)	7.2/1000 py	7.1/1000 py	1.00	1.00
	AD but no BD	4647	83 (12,103)	6.9/1000 py	8.0/1000 py	0.95 (0.69-1.31)	1.13 (0.81-1.56)
	BD but no AD	3358	97 (9724)	10.0/1000 py	8.3/1000 py	1.37 (1.00-1.87)	1.20 (0.87-1.64)
	AD and BD	1638	50 (5625)	8.9/1000 py	9.7/1000 py	1.20 (0.83-1.74)	1.39 (0.96-2.02)
>=4	No AD no BD	1600	58 (3393)	17.1/1000 py	17.3/1000 py	1.00	1.00
	AD but no BD	1840	63 (4316)	14.6/1000 py	15.8/1000 py	0.86 (0.60-1.23)	0.90 (0.63-1.30)
	BD but no AD	2396	116 (6161)	18.8/1000 py	18.0/1000 py	1.11 (0.81-1.52)	1.05 (0.77-1.45)
	AD and BD	732	50 (2415)	20.7/1000 py	23.0/1000 py	1.23 (0.84-1.80)	1.31 (0.89-1.92)

* adjusted for age and sex; CDS: Chronic Disease Score; AD: antidepressant; BD: benzodiazepine; DM: diabetes mellitus; HR: hazard ratio

BD (5.1/1000 person-years) (Table 2). Participants using AD but no BD did not have a higher incidence rate than those using neither AD nor BD. After adjustment for age and sex, the risk in BD but no AD users decreased (HR (95% CI) = 1.30 (1.10-1.54)). Additional adjustment for the CDS lowered the HR (95% CI) to 1.21 (1.02-1.43) in participants using BD but no AD and to 1.37 (1.12-1.68) in participants using both AD and BD, indicating that part of the association between BD use and initiation of diabetes treatment was explained by the presence of chronic disease.

The presence of chronic diseases did not modify the association between AD or BD use and initiation of diabetes treatment because the adjusted HRs were fairly similar across the three strata of the CDS (Table 3). The crude and age-adjusted incidence rate, however, increased substantially with increasing CDS (Table 3 and Figure 1), suggesting that the presence of chronic disease was associated with initiation of diabetes treatment.

Figure 1 Age-adjusted incidence rate of initiation of diabetes treatment in no antidepressant (AD) and no benzodiazepine (BD) users, AD but no BD users, BD but no AD users, both AD and BD users, stratified for Chronic Disease Score (CDS)



Discussion

In this study, we did not find an increased risk of diabetes in adults taking ADs and the association between ADs and diabetes could not be explained by the presence of chronic somatic disease. We did find that BD use was associated with an increased risk of diabetes. This increased risk was, in part, explained by chronic morbidity. Moreover, there was a clear association between the presence of chronic diseases and initiation of diabetes treatment.

A recent meta-analysis of nine prospective studies found that depression increased the risk of diabetes with approximately 35% [3]. An explanation for our contradictory findings is that the majority of the studies included in the meta-analysis used self-report questionnaires to assess depressive symptoms, whereas we used AD use as an indicator for depression. Although the sensitivity of these self-report questionnaires is generally rather high, the specificity can be low. A review on several case-finding questionnaires to identify depression in primary care found a median sensitivity for major depression of 85% and a median specificity of 74% [12]. Our finding that BD use but not AD use increased the risk of diabetes might indicate an effect of psychosocial complaints and not a specific effect of depression.

Another explanation for our findings involves the potential influence of chronic diseases on the association between depression and diabetes. We observed that the association between BD use and diabetes was partly explained by chronic diseases. As none of the studies that found an increased risk of diabetes in depressed participants adjusted or stratified for chronic diseases other than cardiovascular disease, it might be that previously observed associations could have been explained by chronic diseases.

Finally, a possible explanation for not finding an association between AD use and diabetes is that treatment of depression prevented the onset of diabetes. We, however, consider this not very likely because it assumes that treatment was successful, and ADs are only effective in about 50-60% of depressed participants [13]. Furthermore, clinical trials on the efficacy of ADs among diabetes patients did not find improved diabetes outcomes [14,15].

An advantage of this study was that we used a large pharmacy database to address our research question. This database consists of a representative sample of about 200 pharmacies in more than fifty regions scattered over the Netherlands. Currently, it covers data of more than two million residents. The health care system in The Netherlands secures that individuals with different social economical status have equal access to health care. As virtually all patients in the Netherlands are registered with a single community pharmacy, pharmacy records are almost complete with regard to prescription drugs, independent of the prescriber. Owing to this large database, many individuals could be included in this study, which resulted in a considerable number of incident cases of diabetes and a large amount of patient years. This led to precise estimates of the associations under study. Moreover, we had enough participants to make four exposure groups, which enabled us to look at AD and BD use separately. Likewise, we had enough power to analyse the associations under study within the strata of the CDS. Information about these chronic diseases was widely available through medication use registered in the database. Finally, the cohort design and the exclusion of prevalent diabetes cases at baseline made it possible to

look at the temporal associations between AD and BD use and the onset of diabetes.

A potential disadvantage of this study is that we had no information on lifestyle factors, such as smoking, body mass index and physical activity, which could be confounders of the associations under study. Not adjusting for these factors could give an overestimation of the effect. For the association between AD use and diabetes treatment, this appeared to be no problem because we did not find an association at all. It is possible that part of the observed association between BD use and diabetes treatment is explained by lifestyle factors.

In this study, we used ADs, BDs and antidiabetic medication as proxies for depression, other psychosocial complaints and diabetes, respectively. By taking AD use as a proxy for depression, we will have missed mild and unrecognized cases. The results of our study might therefore not apply to subjects with mild or untreated depression. Especially in the group with participants using neither AD nor BD, some unrecognized depressed participants will have been included. As this would have been a relatively small proportion, it could not have caused complete dilution of the effect. We chose to include participants with at least two prescriptions of ADs or BDs. A sensitivity analysis including only participants with more than four prescriptions of AD or BD, that is participants with more severe depression or psychosocial complaints, gave similar results. It is possible that in some patients, ADs were prescribed for other indications than depression, such as anxiety and perhaps eating disorders and neuropathy. These other indications than depression could also be related to chronic diseases and might explain part of the association between AD use and chronic diseases. Insomnia and anxiety are the main indications for BD use but in practice they are prescribed more broadly for psychosocial related indications [16], and therefore we think BD use is a good proxy for having general psychosocial complaints. The onset of diabetes was defined as the initiation of diabetes treatment in this study. As the actual date of diabetes diagnosis for some patients may have occurred before the date of initiation of diabetes treatment we cannot completely exclude the possibility of reversed causality, meaning that diabetes could have led to BD use.

In conclusion, we did not find an increased incidence of diabetes in depression when using ADs as an indicator of depression and therefore could not confirm previous studies on this topic. We did find an increased risk of diabetes in BD users, which was partly explained by the presence of chronic diseases. Future studies on the relationship between depression and diabetes should at least include the presence of chronic diseases to study its influence on this relationship.

References

1. Goldney RD, Phillips PJ, Fisher LJ, Wilson DH. Diabetes, Depression, and Quality of Life: a Population Study. *Diabetes Care* 2004;27:1066-70.
2. Anderson RJ, Freedland KE, Clouse RE, Lustman PJ. The Prevalence of Comorbid Depression in Adults With Diabetes: a Meta-Analysis. *Diabetes Care* 2001;24:1069-78.
3. Knol MJ, Twisk JW, Beekman AT, Heine RJ, Snoek FJ, Pouwer F. Depression As a Risk Factor for the Onset of Type 2 Diabetes Mellitus. A Meta-Analysis. *Diabetologia* 2006;49:837-45.
4. Bjorntorp P, Holm G, Rosmond R. Hypothalamic Arousal, Insulin Resistance and Type 2 Diabetes Mellitus. *Diabet Med* 1999;16:373-83.
5. Musselman DL, Betan E, Larsen H, Phillips LS. Relationship of Depression to Diabetes Types 1 and 2: Epidemiology, Biology, and Treatment. *Biol Psychiatry* 2003;54:317-29.
6. Kawakami N, Takatsuka N, Shimizu H, Ishibashi H. Depressive Symptoms and Occurrence of Type 2 Diabetes Among Japanese Men. *Diabetes Care* 1999;22:1071-6.
7. Engum A, Mykletun A, Midthjell K, Holen A, Dahl AA. Depression and Diabetes: A Large Population-Based Study of Sociodemographic, Lifestyle, and Clinical Factors Associated With Depression in Type 1 and Type 2 Diabetes. *Diabetes Care* 2005;28:1904-9.
8. Pouwer F, Beekman AT, Nijpels G, Dekker JM, Snoek FJ, Kostense PJ et al. Rates and Risks for Co-Morbid Depression in Patients With Type 2 Diabetes Mellitus: Results From a Community-Based Study. *Diabetologia* 2003;46:892-8.
9. Herings RM, Bakker A, Stricker BH, Nap G. Pharmaco-Morbidity Linkage: a Feasibility Study Comparing Morbidity in Two Pharmacy Based Exposure Cohorts. *J Epidemiol Community Health* 1992;46:136-40.
10. Von Korff M, Wagner EH, Saunders K. A Chronic Disease Score From Automated Pharmacy Data. *J Clin Epidemiol* 1992;45:197-203.
11. Baan CA and Poos MJ. *Prevalence and Mortality of Diabetes Mellitus in The Netherlands (in Dutch).* In *Volksgezondheid Toekomst Verkenning, Nationaal Kompas Volksgezondheid*. RIVM, Bilthoven, 2005.
12. Williams JW, Jr., Pignone M, Ramirez G, Perez SC. Identifying Depression in Primary Care: a Literature Synthesis of Case-Finding Instruments. *Gen Hosp Psychiatry* 2002;24:225-37.
13. Lustman PJ and Clouse RE. Treatment of Depression in Diabetes: Impact on Mood and Medical Outcome. *J Psychosom Res* 2002;53:917-24.
14. Lustman PJ, Griffith LS, Clouse RE, Freedland KE, Eisen SA, Rubin EH et al. Effects of Nortriptyline on Depression and Glycemic Control in Diabetes: Results of a Double-Blind, Placebo-Controlled Trial. *Psychosom Med* 1997;59:241-50.
15. Paile-Hyvarinen M, Wahlbeck K, Eriksson JG. Quality of Life and Metabolic Status in Mildly Depressed Women With Type 2 Diabetes Treated With Paroxetine: a Single-Blind Randomised Placebo Controlled Trial. *BMC Fam Pract* 2003;4:7.
16. Olfson M and Pincus HA. Use of Benzodiazepines in the Community. *Arch Intern Med* 1994;154:1235-40.

Chapter 2.3

Influence of antidepressants on glycemic control in patients with diabetes mellitus

Abstract

Background

Anecdotal evidence suggests that antidepressants may complicate glycemic control. The objective of this longitudinal study was to investigate the influence of antidepressants (AD) on glycemic control within diabetes patients.

Methods

From the pharmacy registry database PHARMO, we selected insulin users who did not use oral antidiabetics. The study population comprised: 133 patients with at least 12 months insulin use before and 6 months during an AD episode, including 56 patients with an additional 6 months of insulin use after the AD episode; 180 patients with 24 months insulin use without an AD episode. Glycemic control was measured as the amount of insulin used, which was calculated intra-individually in 3-month periods. We stratified for selective serotonin reuptake inhibitors (SSRI) and tricyclic antidepressants (TCA).

Results

Mean age (SD) of the subjects was 53.9 (19) years; 46.9% were men. Overall, the amount of insulin used did not change during or after AD use. No-AD users showed an increase of 16% in amount of insulin used over a period of 2 years ($p<0.001$). SSRI users showed a decrease of 13% in amount of insulin used during the AD episode ($p=0.029$), while no change was seen in TCA users. Notable was the large intra- and interindividual variation in amount of insulin used across all groups.

Conclusions

Overall, antidepressant use did not influence glycemic control in diabetes patients. The tendency for a difference between SSRIs and TCAs is suggestive for a pharmacologic effect of antidepressants rather than a general effect of depression on glycemic control.

Introduction

Depression is a common co-morbidity in patients with diabetes mellitus. The risk of depression is doubled in patients with diabetes compared with those without diabetes [1]. In addition, among diabetes patients depression is associated with poor glycemic control [2]. In turn, poor glycemic control is a risk factor for macro vascular complications, such as cardiovascular disease, and micro vascular complications, such as retinopathy and nephropathy. HbA1c, which is an aggregate measure of glycemic control over the 120-day period before testing, is an important indicator for diabetes regulation. Efficacy studies have demonstrated that achieving and maintaining HbA1c levels below 7% substantially decreased diabetes-related complications in individuals [3,4].

Both depression and antidepressant use can influence glycemic control in diabetes patients in several ways [5,6]. Depression can worsen glycemic control by life style changes such as altered food intake, decreased physical activity, smoking and decreased medication adherence. Although one study showed that self-care behaviour could not fully explain the association between depression and glycemic control [7]. From a physiological perspective, depression can lead to increased cortisol secretion of the hypothalamic pituitary adrenal (HPA) axis, which can cause hyperglycemia and thereby worsen glycemic control. Evidence on the effect of antidepressants on glucose and insulin levels mainly comes from animal studies, case reports and short term trials with selected and small groups of patients. Antidepressant use can disturb glycemic control by its hyperglycemic effect, which is thought to be more pronounced in some tricyclic antidepressants such as nortriptyline [8-10]. In contrast, some specific serotonin reuptake inhibitors (SSRIs), such as fluoxetine and paroxetine, may decrease blood glucose levels and HbA1c and reduce insulin requirements [11-13]. Antidepressant use could also improve glycemic control because of successful treatment of depression.

In diabetes patients, changes in the degree of glycemic control can be deduced from changes in the amount of insulin that these patients need. Patients using insulin monitor their own blood glucose levels and they will use more insulin when their blood glucose is high and less when it is low.

The objective of the present study was to investigate the influence of antidepressant treatment on glycemic control, measured as the amount of insulin used, within diabetes patients with an episode of antidepressant use. To study this we described the variability in amount of insulin used in diabetes patients before, during and after an episode of antidepressant use. To study the natural course of the amount of insulin used over time we also included diabetes patients without an episode of antidepressant use.

Methods

Data source

Our cohort was selected from the PHARMO database. This database is described in detail elsewhere [14]. In short, the PHARMO database comprises all pharmacy dispensing records of all residents of about fifty Dutch municipalities, counting for approximately two million patient histories. Since virtually all patients in the Netherlands are registered with a single community pharmacy, pharmacy records are almost complete with regard to prescription drugs independent of the prescriber. In the Netherlands, antidepressants (AD) and insulin are only available as prescription drugs. Therefore, pharmacy data will cover all use of these drugs. Available variables in this database include sex, date of birth, dispensed drugs (coded according to the Anatomical Therapeutic Chemical (ATC) classification), drug dispensing date, amount of drug dispensed, prescribed dosage regimen and prescriber.

Study population

For this study, pharmacy data from 1991 until 2003 were used. From the PHARMO database, all subjects with a prescription of any antidepressant (ATC code = N06A*), a prescription of insulin (ATC code = A10A*) and no prescription of any oral antidiabetic drug (ATC code = A10B*) were identified (n=840). From this sample, patients were included in the study population if they met the following criteria: 1) 18 years or older at the start of AD use; 2) used insulin for at least 12 months before the start of the AD; and 3) the episode of AD use (defined in detail later) lasted for at least 6 months. This resulted in a study sample of 133 patients, of whom 56 also had at least 6 months of insulin use after they stopped with the AD. The duration of the periods before, during and after an antidepressant episode were arbitrarily chosen so that we would have both a considerable amount of follow-up time and a considerable amount of patients to study.

To study the natural course of the amount of insulin used over a certain time period, a random sample of insulin users without a prescription of any antidepressant and without a prescription of any oral antidiabetic drug was selected from the PHARMO database. These patients were assigned a random index date in such a way that the distribution of the index date relative to the total period of insulin use was similar to the distribution of the start of the AD episode relative to the total period of insulin use in AD users. Subsequently, patients were included in the study population if they were 18 years or older at the index date and if they used insulin for at least 12 months before the index date and for at least 12 months after the index date (n=180).

Thus, the study population comprised: 133 patients with at least 12 months of insulin use

before and at least 6 months of insulin use during the AD episode of whom 56 patients had also at least 6 months of insulin use after the end of the AD episode (AD users), and 180 patients without an AD episode and 12 months of insulin use before and after the index date (no-AD users) (Figure 1).

Antidepressant use

The start of an episode of antidepressant use was defined as the first prescription for any AD within the study period. The end of an episode of AD use was defined as no new prescription of any AD within 6 months after the end date of the last AD prescription. The end date of a prescription was calculated by adding the duration of AD use (amount dispensed divided by daily dose) to the start date of the prescription. If a subject had more than one episode of AD use in the study period, only the first episode was used. Switching of type of AD and changes of the defined daily dose per day were allowed during one episode.

Amount of insulin used

The amount of insulin used was calculated in international units per day, by dividing the units of insulin dispensed by the days between two subsequent prescriptions. Per subject the mean number of units of insulin per day was calculated over periods of 3 months. In The Netherlands, insulin is usually prescribed for periods of 3 months.

Data-analysis

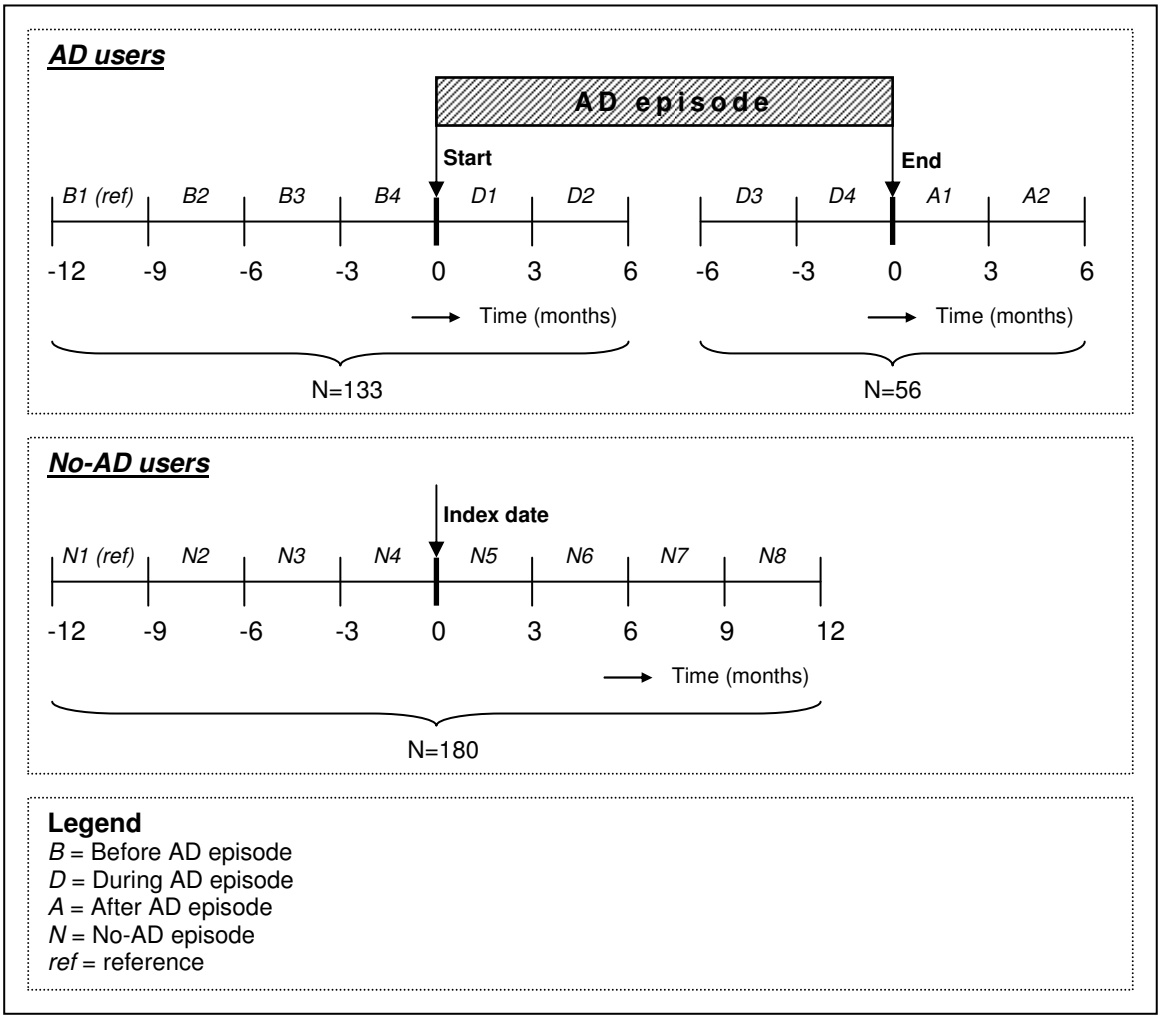
Mean age (+/- SD) at the start of insulin use and the percentage of male patients was calculated in AD users and no-AD users. In AD users the number of AD episodes and the duration of the first AD episode were calculated. Also, the percentage of patients that used SSRIs (selective serotonin reuptake inhibitors), TCAs (tricyclic antidepressants) and other antidepressants and the percentage of patients that had a change of defined daily dose of AD or a switch of antidepressant during the first episode were calculated.

The distribution of the amount of insulin used at the different time points in AD and no-AD users was skewed to the left. The tests for normality (Kolmogorov-Smirnov) were all significant, indicating non-normality. Log transformation solved this non-normality for some time points but not for all. For this reason we chose to perform non-parametric tests.

Differences in the amount of insulin used over time were tested with the non-parametric Friedman test for repeated measurements. Differences between two time points were tested with the non-parametric Wilcoxon test for paired observations. To show the intra-individual differences over time we calculated the relative amount of insulin used in each

period relative to the reference period. The reference period in the AD users and no-AD users was the 3-month period between 12 and 9 months before the start of the AD episode (Figure 1; period B1) and the 3-month period between 12 and 9 months before the index date (Figure 1; period N1), respectively. For each subject the relative amount of insulin used was calculated by dividing the units of insulin used in each 3-month period by the units of insulin used in the reference period. For example, a subject used 50 units of insulin per day in reference period B1 and used 55 units per day in period B2, meaning that the relative amount of insulin used in this subject was 1.10 (55/50) in period B2. Subsequently, for each 3-month period the median relative amount and interquartile range over all patients was calculated and presented in graphs.

Figure 1 Schematic representation of the timeline of analysis of amount of insulin used in antidepressant (AD) users before, during and after an episode of AD use and in no-AD users before and after the index date



Besides calculating the relative amount of insulin used between the 3-months periods, we calculated in each period the percentage of patients that increased (relative amount of insulin used > 1.10), decreased (relative amount of insulin used < 0.90) and kept constant on their insulin used (relative amount of insulin used between 0.90 and 1.10). This was done in AD users as well as in no-AD users. Differences in percentages between two time points were tested with the McNemar test for paired proportions.

We stratified all analyses for the two main types of AD, namely SSRIs and TCAs, because previous studies suggested that SSRIs and TCAs have contradictory effects on glycemic control. Patients who switched from one type of AD to another within the periods that we studied, were excluded from these analyses. Differences in the amount of insulin used over time between SSRI users and TCA users were tested with a mixed between-within subjects analysis of variance (a non-parametric alternative is not available). To show the intra-individual differences over time we calculated the relative amount of insulin used in each period relative to the reference period, as described above, and presented this in graphs. Differences between SSRI and TCA users in percentage of increasers and decreasers in amount of insulin used in one time point were tested with the Chi-square test.

We calculated the change in amount of insulin used we could detect with 80% power and an alpha of 0.05 with the observed data. The change we could detect between the various time points within the patients with an AD episode ranged from 6.4 to 10.6 units of insulin used (about 10-16% change), with one outlier for the difference between D3 and D4 where we could detect a change of 14.6 units of insulin. For the patients without an AD episode the change we could detect ranged from 5.7 to 7.9 units of insulin used (about 10-13% change).

Results

Table 1 presents the baseline characteristics in AD users and no-AD users. The mean age of the AD users (54.8 +/- 19) was comparable with the age of the no-AD users (53.0 +/- 19). The no-AD users were more often male (48.3%) than the AD users (45.1%). The majority of the AD users (91%) had only one episode of AD use during the study period. The duration of the first episode was 6 to 12 months in most patients (43.6%) and more than 24 months in 25.6% of the AD users. SSRIs were most frequently used and 18 patients (13.5%) used different types of antidepressants during the first AD episode. In almost 50% of the patients the defined daily dose of AD was changed during the AD episode.

Figure 2 shows the median relative amount of insulin used in AD users before, during and after an AD episode (upper graph) and the median relative amount of insulin used in no-AD

Table 1 Baseline characteristics of patients with an antidepressant (AD) episode and without an AD episode

	AD users		No-AD users
	Whole set	*Subset* [a]	
N	133	56	180
Age, mean (SD)	54.8 (19)	50.9 (17)	53.0 (19)
Male sex	60 (45.1%)	26 (46.4%)	48.3%
Number of AD episodes			
1	121 (91.0%)	44 (78.6%)	NA
2	12 (9.0%)	12 (21.4%)	NA
Duration of first AD episode			
6 to 12 months	58 (43.6%)	29 (51.8%)	NA
12 to 18 months	26 (19.6%)	10 (17.9%)	NA
18 to 24 months	15 (11.2%)	7 (12.5%)	NA
> 24 months	34 (25.6%)	10 (17.9%)	NA
Type of AD			
SSRI	85 (63.9%)	30 (53.6%)	NA
TCA	23 (17.3%)	14 (25.0%)	NA
Other	7 (5.3%)	5 (8.9%)	NA
SSRI + TCA	8 (6.0%)	4 (7.1%)	NA
SSRI + other	7 (5.3%)	3 (5.4%)	NA
TCA + other	1 (0.8%)	0	NA
SSRI + TCA + other	2 (1.5%)	0	NA
Change of DDD of AD [b]	63 (47.4%)	27 (48.2%)	NA

[a] subset of all AD users who also had at least 6 months of insulin use after the end of the AD episode [b] within first AD episode; DDD: defined daily dose; NA: Not applicable; AD: antidepressant

users (lower graph). The upper graph showed a significant increase in the amount of insulin used in AD users before the AD episode (p=0.035). There were no significant changes over time in the amount of insulin used when analysing period B1 through D2 (p=0.094) or period B1 through A2 (p=0.916). There were no significant differences between time points before or during the AD episode. No-AD users showed a significant increase in insulin use of on average 16% over a period of two years (p<0.001; figure 2, lower graph).

From the percentage of AD users of whom the amount of insulin used increased or decreased, we observed that the variability between patients was large (Figure 3, upper graph). More than 25% of the patients had at least 10% decrease in the amount of insulin used before, during and after the AD episode, while around 40% of the patients had at least 10% increase in insulin used in these periods. No clear changes in these percentages were seen during or after the AD episode compared with before. In the no-AD users, there was

also a large variability between the patients regarding a decrease or increase in the amount of insulin used (Figure 3, lower graph). Overall, most patients had a more than 10% increase in the amount of insulin used. The percentage of increasers ranged from 39.4% in period N5 to 55.0% in period N8.

Figure 2 Median relative amount [interquartile range] of insulin used in AD users before, during and after an AD episode (upper graph) and median relative amount [interquartile range] of insulin used in no-AD users (lower graph)

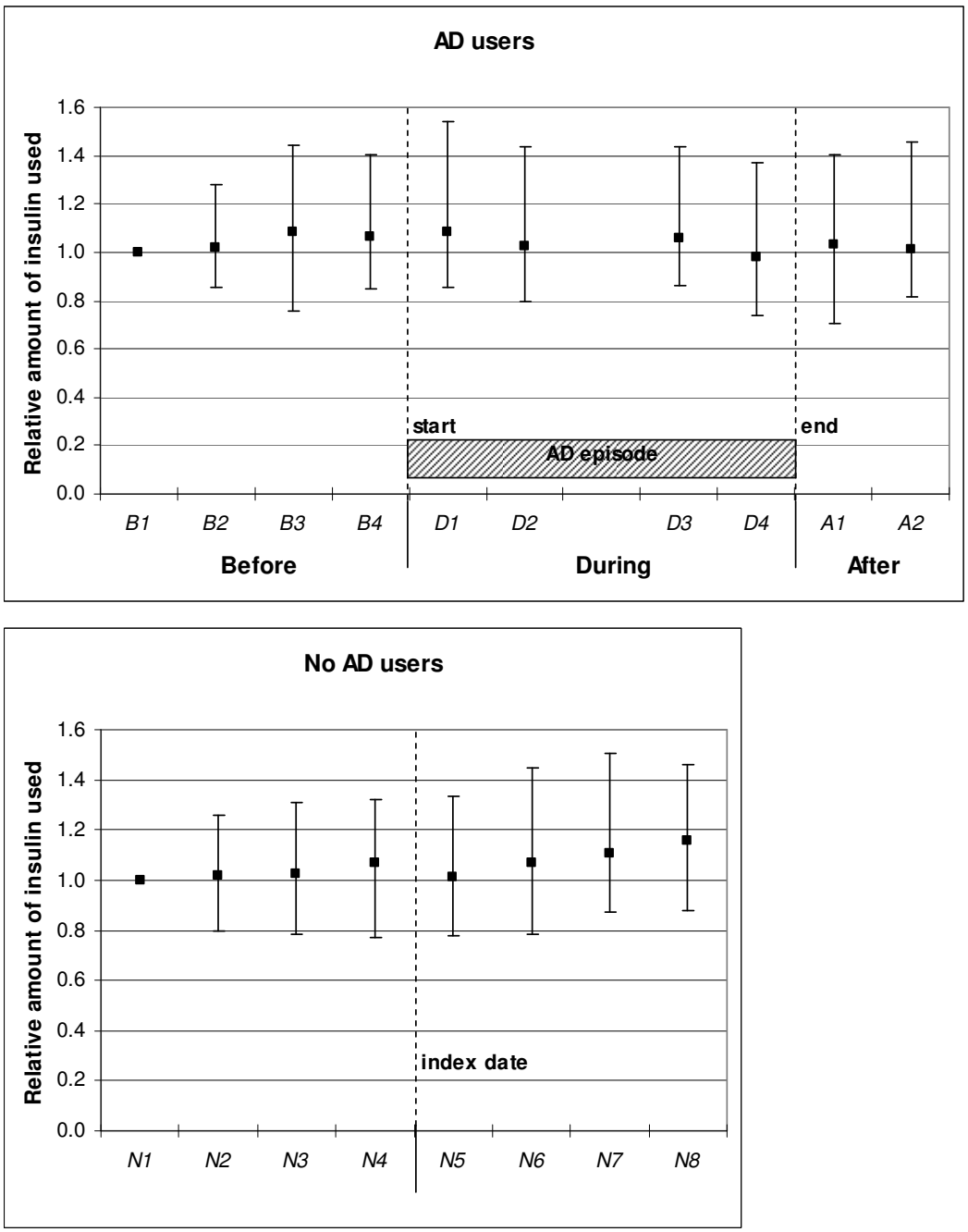
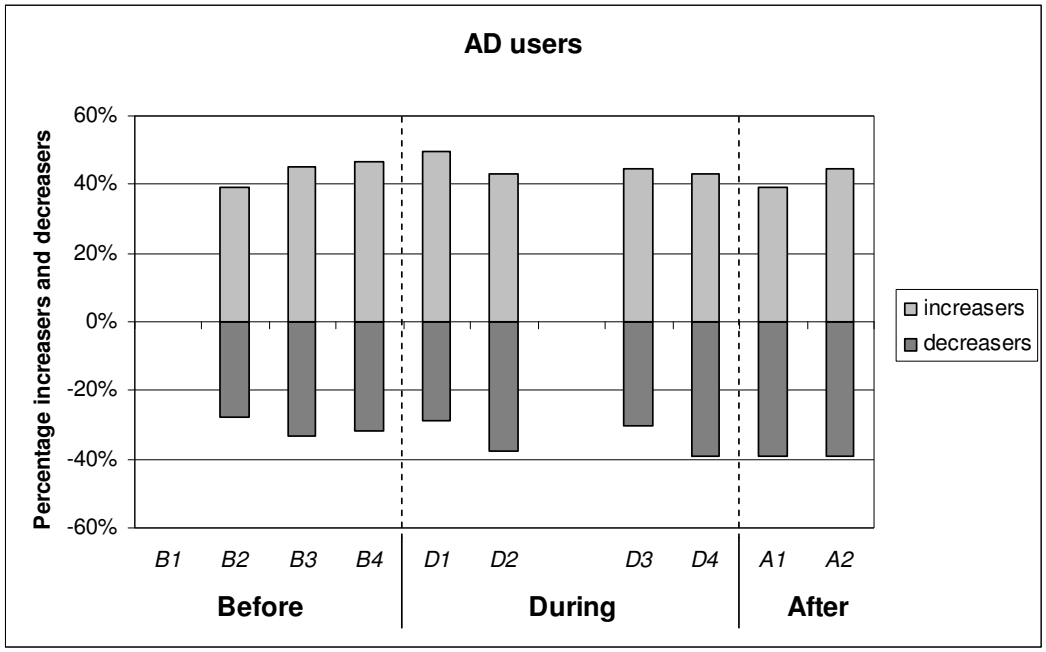


Figure 3 Percentages of patients of whom their amount of insulin used decreased with more than 10% or increased with more than 10% before, during and after an antidepressant (AD) episode, with 12-9 months before the AD episode as the reference 3-month period (*B1*)



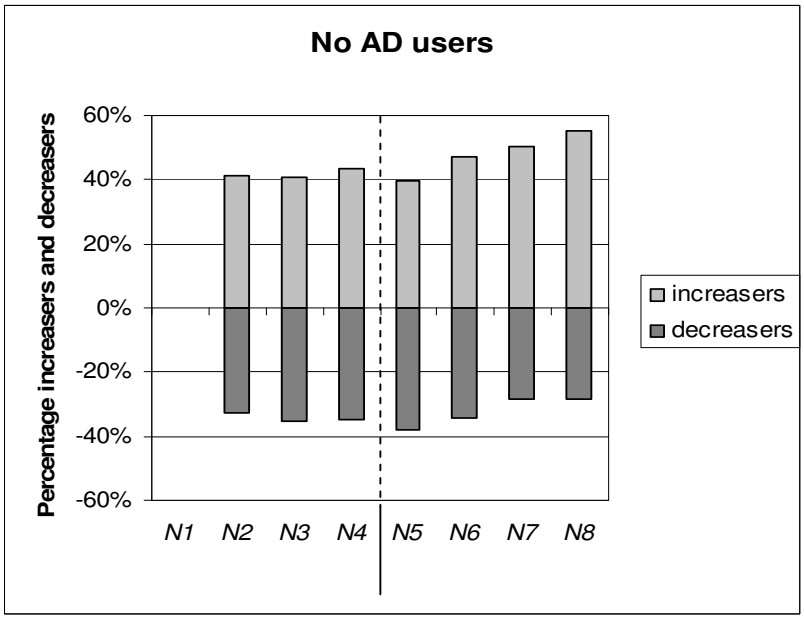
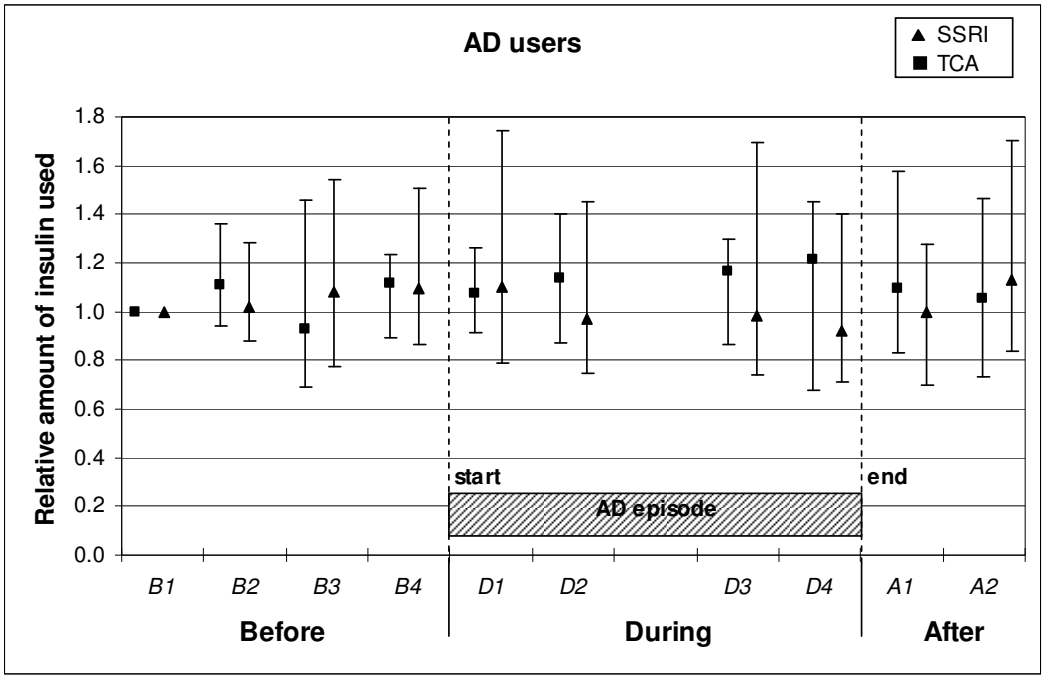


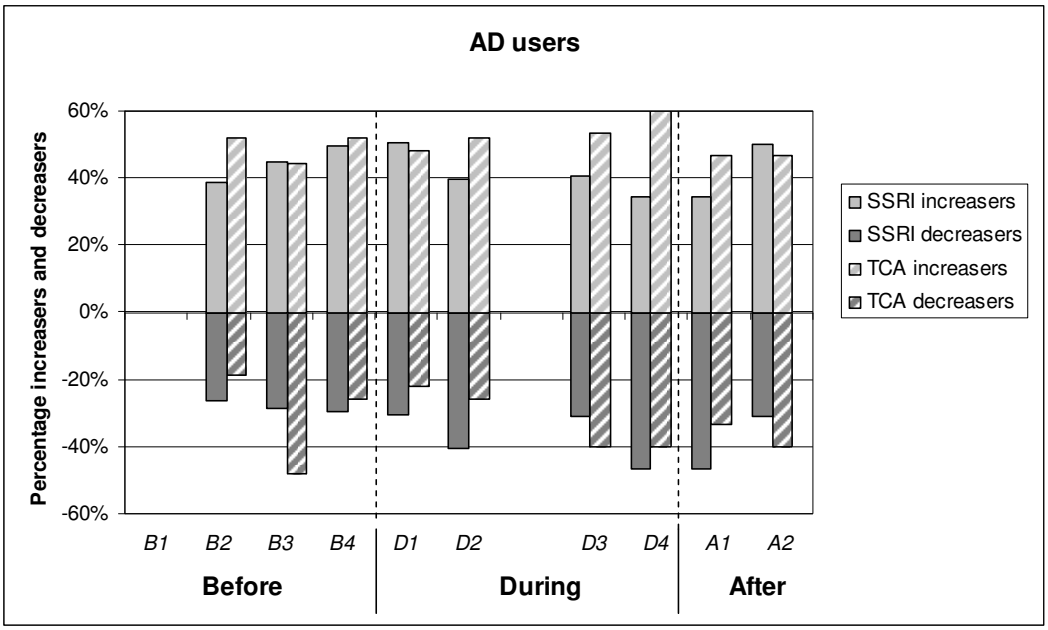
Figure 4 shows the median relative amount of insulin used in AD users, stratified for SSRIs and TCAs. In period D1 and D2, 91 patients used SSRIs, 27 used TCAs, 7 used other antidepressants, and 8 switched between type of AD and were excluded. In period D3 and D4, 32 patients used SSRIs, 15 used TCAs, 5 used other antidepressants, and 4 switched between type of AD and were excluded. SSRI users had a decrease in amount of insulin used

Figure 4 Median relative amount [interquartile range] of insulin used in AD users before, during and after an AD episode, stratified for SSRI users (n=91 for period B1 to D2; n=32 for period D3 to A2) and TCA users (n=27 for period B1 to D2; n=15 for period D3 to A2)



during the AD episode compared to before the AD episode (13% difference between period B4 and D2; p=0.029). This decrease seemed to have disappeared again after the AD episode (relative intensity is 1.13 in period A2). The TCA users used similar amounts of insulin during the AD episode compared to before the AD episode. The difference between TCA users and SSRI users in amount of insulin used over time was not statistically significant (p=0.462). Looking at the percentages of patients who increased and decreased in their amount of insulin used when stratified for type of AD, again the large interindividual variability was striking (Figure 5). The percentage of SSRI users that had a more than 10% decrease in amount of insulin used was higher in period D2 (40.7%) than in the periods before the AD episode (26.4-29.7%, p=0.036 for difference between B2 and D2). In TCA users, the percentages of patients that had a more than 10% increase or decrease were similar during the AD episode as in the periods before the AD episode. There were no significant differences between SSRI and TCA users regarding the percentages of increasers and decreasers before, during or after the AD episode.

Figure 5 Percentages of patients of whom their amount of insulin used decreased with more than 10% or increased with more than 10% before, during and after an antidepressant (AD) episode, with 12-9 months before the AD episode as the reference 3-month period (*B1*), stratified for SSRI users (n=91 for period B1 to D2; n=32 for period D3 to A2) and TCA users (n=27 for period B1 to D2; n=15 for period D3 to A2)



Discussion

Overall, antidepressants did not influence glycemic control, measured as the amount of insulin used, in diabetes patients. A difference of the effect on glycemic control was observed between SSRI users and TCA users. SSRI use seemed to decrease the amount of insulin used during antidepressant use, suggesting a beneficial effect of antidepressant use on glycemic control, while TCA use did not change the amount of insulin used during an AD episode. However, this decrement in SSRI users was not statistically significant. In diabetes patients without an AD episode the amount of insulin used significantly increased over two years. The absence of a change in AD users and an increase in no-AD users may be of clinical importance. However, we did not directly compare AD and no-AD users because it is well known in diabetes that many external factors determine differences in the amount of insulin used between subjects who did and did not use AD, i.e. a priori the between patient variability is much larger than the within patient variability. Moreover, we were not able to directly compare the AD and no-AD users as the timeline of patients with and without an AD episode was not the same because the duration of the AD episode was different in every patient. A large intra- and interindividual variability in the amount of insulin used in diabetes patients was observed.

A strong aspect of this study is that we used a longitudinal design with repeated measurements of the amount of insulin used over time and a relatively long follow-up. This design enabled us to evaluate the course of the amount of insulin used before, during and after an episode of antidepressant use. Second, we included a relatively large group of diabetes patients and because we used prescription data, we could study several types of antidepressants at once. Third, since virtually all patients in the Netherlands are registered with a single community pharmacy, pharmacy records are almost complete with regard to prescription drugs, independent of the prescriber. Furthermore, since antidepressants and insulin are only available as prescription drugs in the Netherlands, we had complete data of all our subjects. Fourth, we calculated the amount of insulin used within patients, which gives more reliable results than between subject analyses, because the relation under study is less likely to be confounded by external factors.

A limitation of this study is that our method of measuring glycemic control, i.e. the amount of insulin used of diabetes patients using pharmacy data, may not be sensitive enough to detect small changes in the amount of insulin used. Second, we assessed insulin use in periods of 3 months. This period was chosen to get a valid estimate of insulin use because insulin is mostly prescribed for periods of 3 months in The Netherlands. However, it might be that this 3-month period was too long to detect changes in insulin use, i.e. that changes were levelled out over the 3-month period. Third, as we only had prescription data we could not differentiate between type 1 and type 2 diabetes patients. We selected users of insulin without use of oral antidiabetic drugs, meaning that the study population probably included mostly patients with type 1 diabetes. Fourth, the amount of insulin used was used as a proxy for glycemic control. However, we think that our outcome measure, in a population of insulin users who will predominantly adjust insulin dose on measured blood glucose levels, is a good measure for changes in glycemic control. Other outcome measures like changes in HbA1c-values and changes in average blood glucose may be less accurate, because many patients will adjust insulin dose by strict glucose-monitoring.

To our knowledge, this is the first study that investigated the influence of antidepressants on glycemic control in an observational study with repeated measurements. Previous studies investigated the influence of different antidepressants on HbA1c levels, glucose levels and plasma insulin levels. Most of these studies had small sample sizes (< 50 patients) [11,15,16], studied non-diabetic patients [11,15,17], or had a short study duration (< 6 months of antidepressant treatment) [11,16]. The results of these studies were inconsistent. One study found a significant decrease of blood glucose after treatment with the SSRI fluoxetine, and a significant increase in blood glucose after use of the TCA imipramine [11].

Other studies did not find significant effects on blood glucose levels of paroxetine [16], bupropion [17] and several tricyclic antidepressants [15]. The study of treatment with the SSRI paroxetine found a significant decrease of HbA1c [16] and the study where several tricyclic antidepressants were evaluated found an increased insulin sensitivity after treatment [15]. The latter finding is inconsistent with other studies that reported unfavourable effects of TCAs on glucose and insulin homeostasis. A review of the effects of antidepressants on glucose homeostasis and insulin sensitivity concluded that, in general, serotonergic antidepressants had a favourable effect on blood glucose levels and insulin sensitivity [6]. Both venlafaxine and duloxetine, which are serotonin and noradrenalin reuptake inhibitors, had neutral metabolic effects and tricyclic antidepressants disrupted glucose homeostasis. Our results corroborate with the conclusions of the review but the variation in the amount of insulin used between the subjects in our study was too large to really confirm these findings. A recent case report found a close association between imipramine treatment and insulin use [18], but we were unable to replicate this finding. Most studies on the association between depression (as opposed to antidepressants) and HbA1c are cross-sectional studies and these also showed conflicting results. Some found a significant association [19,20], while others did not [21,22]. Recently, a randomized clinical trial showed that depression recovery with sertraline as well as sustained remission with and without treatment are associated with improvements in glycemic control [23].

In conclusion, in this longitudinal study antidepressant use overall did not influence glycemic control in diabetes patients. The tendency for a difference that we observed between SSRIs and TCAs is suggestive for a pharmacologic effect of antidepressants rather than a general effect of depression on glycemic control. The differences between SSRIs and TCAs in glycemic control were rather small and probably not of importance for choosing a specific type of antidepressant in patients with diabetes mellitus.

References

1. Anderson RJ, Freedland KE, Clouse RE, Lustman PJ. The Prevalence of Comorbid Depression in Adults With Diabetes: a Meta-Analysis. *Diabetes Care* 2001;24:1069-78.
2. Lustman PJ, Anderson RJ, Freedland KE, de Groot M, Carney RM, Clouse RE. Depression and Poor Glycemic Control: a Meta-Analytic Review of the Literature. *Diabetes Care* 2000;23:934-42.
3. The Effect of Intensive Treatment of Diabetes on the Development and Progression of Long-Term Complications in Insulin-Dependent Diabetes Mellitus. The Diabetes Control and Complications Trial Research Group. *N Engl J Med* 1993;329:977-86.
4. Intensive Blood-Glucose Control With Sulphonylureas or Insulin Compared With Conventional Treatment and Risk of Complications in Patients With Type 2 Diabetes (UKPDS 33). UK Prospective Diabetes Study (UKPDS) Group. *Lancet* 1998;352:837-53.

5. Goodnick PJ, Henry JH, Buki VM. Treatment of Depression in Patients With Diabetes Mellitus. *J Clin Psychiatry* 1995;56:128-36.
6. McIntyre RS, Soczynska JK, Konarski JZ, Kennedy SH. The Effect of Antidepressants on Glucose Homeostasis and Insulin Sensitivity: Synthesis and Mechanisms. *Expert Opin Drug Saf* 2006;5:157-68.
7. Lustman PJ, Clouse RE, Ciechanowski PS, Hirsch IB, Freedland KE. Depression-Related Hyperglycemia in Type 1 Diabetes: a Mediational Approach. *Psychosom Med* 2005;67:195-9.
8. Erenmemisoglu A, Ozdogan UK, Saraymen R, Tutus A. Effect of Some Antidepressants on Glycaemia and Insulin Levels of Normoglycaemic and Alloxan-Induced Hyperglycaemic Mice. *J Pharm Pharmacol* 1999;51:741-3.
9. Goodnick PJ. Use of Antidepressants in Treatment of Comorbid Diabetes Mellitus and Depression As Well As in Diabetic Neuropathy. *Ann Clin Psychiatry* 2001;13:31-41.
10. Lustman PJ, Griffith LS, Clouse RE, Freedland KE, Eisen SA, Rubin EH et al. Effects of Nortriptyline on Depression and Glycemic Control in Diabetes: Results of a Double-Blind, Placebo-Controlled Trial. *Psychosom Med* 1997;59:241-50.
11. Ghaeli P, Shahsavand E, Mesbahi M, Kamkar MZ, Sadeghi M, shti-Khavidaki S. Comparing the Effects of 8-Week Treatment With Fluoxetine and Imipramine on Fasting Blood Glucose of Patients With Major Depressive Disorder. *J Clin Psychopharmacol* 2004;24:386-8.
12. Maheux P, Ducros F, Bourque J, Garon J, Chiasson JL. Fluoxetine Improves Insulin Sensitivity in Obese Patients With Non-Insulin-Dependent Diabetes Mellitus Independently of Weight Loss. *Int J Obes Relat Metab Disord* 1997;21:97-102.
13. Park S and Choi SB. Does Fluoxetine Administration Influence Insulin Resistance in 90% Pancreatectomized Rats? *Metabolism* 2002;51:38-43.
14. Herings RM, Bakker A, Stricker BH, Nap G. Pharmaco-Morbidity Linkage: a Feasibility Study Comparing Morbidity in Two Pharmacy Based Exposure Cohorts. *J Epidemiol Community Health* 1992;46:136-40.
15. Okamura F, Tashiro A, Utumi A, Imai T, Suchi T, Tamura D et al. Insulin Resistance in Patients With Depression and Its Changes During the Clinical Course of Depression: Minimal Model Analysis. *Metabolism* 2000;49:1255-60.
16. Paile-Hyvarinen M, Wahlbeck K, Eriksson JG. Quality of Life and Metabolic Status in Mildly Depressed Women With Type 2 Diabetes Treated With Paroxetine: a Single-Blind Randomised Placebo Controlled Trial. *BMC Fam Pract* 2003;4:7.
17. Jain AK, Kaplan RA, Gadde KM, Wadden TA, Allison DB, Brewer ER et al. Bupropion SR Vs. Placebo for Weight Loss in Obese Patients With Depressive Symptoms. *Obes Res* 2002;10:1049-56.
18. Derijks HJ, De Koning FH, Meyboom RH, Heerdink ER, Spooren PF, Egberts AC. Impaired Glucose Homeostasis After Imipramine Intake in a Diabetic Patient. *J Clin Psychopharmacol* 2005;25:621-3.
19. Gross R, Olfson M, Gameroff MJ, Carasquillo O, Shea S, Feder A et al. Depression and Glycemic Control in Hispanic Primary Care Patients With Diabetes. *J Gen Intern Med* 2005;20:460-6.
20. Katon W, Von Korff M, Ciechanowski P, Russo J, Lin E, Simon G et al. Behavioral and Clinical Factors Associated With Depression Among Individuals With Diabetes. *Diabetes Care*

2004;27:914-20.

21. Ciechanowski PS, Katon WJ, Russo JE, Hirsch IB. The Relationship of Depressive Symptoms to Symptom Reporting, Self-Care and Glucose Control in Diabetes. *Gen Hosp Psychiatry* 2003;25:246-52.
22. Paschalides C, Wearden AJ, Dunkerley R, Bundy C, Davies R, Dickens CM. The Associations of Anxiety, Depression and Personal Illness Representations With Glycaemic Control and Health-Related Quality of Life in Patients With Type 2 Diabetes Mellitus. *J Psychosom Res* 2004;57:557-64.
23. Lustman PJ, Clouse RE, Nix BD, Freedland KE, Rubin EH, McGill JB et al. Sertraline for Prevention of Depression Recurrence in Diabetes Mellitus: a Randomized, Double-Blind, Placebo-Controlled Trial. *Arch Gen Psychiatry* 2006;63:521-9.

Chapter 2.4

Depressive symptoms in subjects with diagnosed and undiagnosed type 2 diabetes

Abstract

Background

To investigate if disturbed glucose homeostasis or known diagnosis of diabetes was associated with depressive symptoms. The reason for the increased prevalence of depression in patients with type 2 diabetes (DM2) is unknown.

Methods

Within the Utrecht Health Project, an ongoing longitudinal study among inhabitants of a residential area of a large city in The Netherlands, 4,747 subjects (age: 39.4 +/- 12.5 years) were classified into four mutually exclusive categories: normal fasting plasma glucose (FPG) (< 5.6 mmol/l), impaired FPG (>= 5.6 and < 7.0 mmol/l), undiagnosed DM2 (FPG >= 7.0 mmol/l), and diagnosed DM2. Presence of depressive symptoms was defined as a score of 25 or more on the depression subscale of the Symptom Check List (SCL-90) or self-reported use of antidepressants.

Results

Diagnosed DM2 was associated with an increased risk of depressive symptoms (odds ratio (OR) = 1.69; 95% confidence interval (CI) 1.06-2.72) after adjustment for demographic and lifestyle variables. Additional adjustment for number of chronic diseases reduced the OR to 1.36 (95% CI 0.83-2.23). Impaired fasting glucose and undiagnosed DM2 were not associated with depressive symptoms.

Conclusions

Our findings suggest that disturbed glucose homeostasis is not associated with depressive symptoms. The increased prevalence of depressive symptoms among patients with *diagnosed* DM2 suggests that depressive symptoms might be a consequence of the burden of diabetes. The number of chronic diseases seems to explain part of the association between DM2 and depressive symptoms.

Introduction

Diabetes and depression are both common conditions in today's society. There are about 200 million people with diabetes worldwide [1] and an estimated 121 million people currently suffer from depression [2]. Diabetes and depression are also often comorbid conditions. A recent meta-analysis showed that the prevalence of depression is doubled in patients with type 2 diabetes compared with subjects without diabetes [3]. However, the reason for the increased prevalence among diabetic patients is unknown. Also, the direction of the association between type 2 diabetes and depression is not known. Recently, a meta-analysis of longitudinal studies suggested that depression is a small risk factor for the onset of type 2 diabetes [4]. However, depression is also often seen as a consequence of type 2 diabetes [5]. In this paper we will focus on the latter.

There are two possible mechanisms underlying the association between type 2 diabetes and the onset of depression. First, biochemical changes associated with diabetes could account for the increased risk of depression [5]. For example, hyperglycemia and hyperinsulinemia increase the activity of the hypothalamic-pituitary-adrenal axis, inducing arousal of the nervous system, which in turn may promote depression [6,7]. Second, depression in patients with diabetes may be viewed as the result of the burden of the disease [5,8]. This is supported by the finding that when the burden of diabetes increases, the probability of mood symptoms increases as well [9]. Furthermore, an increased prevalence of depression is also seen in patients with chronic diseases other than diabetes [10].

The aim of the present study was to investigate if disturbed glucose homeostasis is associated with depressive symptoms or if a known diagnosis of diabetes is associated with depressive symptoms. We compared the prevalence of depressive symptoms in 4 groups: subjects with normal fasting glucose level; subjects with prediabetes, i.e., impaired fasting glucose level; subjects who did not know they had type 2 diabetes, but whose fasting plasma glucose (FPG) level indicated the presence of diabetes; and subjects who know they had type 2 diabetes because their doctor diagnosed them.

Methods

Study population

The Utrecht Health Project (UHP) was used as a data source for this study [11]. The UHP is an ongoing longitudinal study, set up in 2000, among all inhabitants of a new residential area of Utrecht, a large city in The Netherlands. Each new inhabitant who registered with a general practitioner (GP) was invited by mail to participate in the study. In the Dutch healthcare system, access to pharmacy care or secondary care is only possible via the GP. Therefore,

almost all inhabitants of The Netherlands are registered with a GP. At baseline, an individual health profile is made for every participant that consists of an interview assisted questionnaire, physical examination, and a blood sample. The questionnaire includes questions about demographic factors, life style factors, current health status, quality of life, psychopathology, and disability. Physical examination includes measurement of weight, height, and blood pressure. Blood assessment includes FPG and cholesterol level measurements. In the future, follow-up data on morbidity, medication and referrals will be obtained through the automated registry of all GPs and pharmacists in the area.

The Medical Ethics Committee of the University Medical Center Utrecht in The Netherlands approved the UHP. The UHP started to recruit participants in 2001 and since then response has been steadily increasing. By January 2005 13,128 inhabitants were invited, of whom 6,755 gave informed consent (51.4%). Baseline data were complete for 6,304 (48.0%) participants, of whom 4,950 were aged 18 years or over. Non-responders were more often male (51.5% versus 45.9%) and non-responders were somewhat younger than responders (mean age (SD): 36.5 (12) versus 38.7 (13)). Reasons for not participating in the study were: not interested in the study (44%), no recognition of personal advantage in participating (26%) and too busy (14%). A small group did not want to participate in scientific research from conviction (2%), others had practical reasons for not participating (14%).

The current analysis is based on baseline data of 4,950 subjects aged 18 years or over.

FPG and diabetes

First, subjects who reported to be diagnosed with diabetes by a physician were classified as having 'diagnosed diabetes'. Second, the remaining subjects were, according to the latest American Diabetes Association criteria [12], categorized based on their FPG concentration into: normal FPG (<5.6 mmol/l), impaired FPG (≥5.6 and <7.0 mmol/l), undiagnosed diabetes (≥7.0 mmol/l). Fasting glucose values were obtained from a venous blood sample. In 178 subjects, a finger prick sample was obtained instead of a venous blood sample, producing whole blood glucose values. These whole blood values were converted into venous values by multiplying with factor 1.11 [13]. Some patients were not fasting when the blood sample was obtained. If subjects had a nonfasting blood glucose value which was increased (≥ 5.6 mmol/l), they were excluded from further analyses (n=22). Patients with diagnosed diabetes who used insulin and no oral hypoglycemic agents were defined as having type 1 diabetes and were excluded (n=14).

Depressive symptoms

Depressive symptoms were assessed with a psychopathology questionnaire, the Symptom Check List (SCL-90) [14,15]. The SCL-90, a self-rated scale, consists of eight psychiatric symptom domains, including a 16-item depression subscale. Each item is scored on a 5-point scale. In this study, the presence of depressive symptoms was defined as a score of 25 or more on the depression subscale. A validation study showed that this cut off point was the optimal cut off point with a sensitivity of 88.5 and a specificity of 60.7 when compared with the depression section of the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) (SCID) [16]. Another study showed that a cut off point of 25 gave a sensitivity of 95.5 and specificity of 74.0 for major depression one month after a myocardial infarction, also with the SCID as the gold standard [17].

In addition, subjects reporting use of antidepressants were classified as having depressive symptoms irrespective of their score on the SCL-90. Subjects who used antidepressants for a non-psychiatric indication (n=3), such as pain and incontinence, were recoded as 'not using antidepressants'.

Covariates

Before performing the analyses, we selected sex, age, education level, body mass index, smoking, alcohol consumption, physical activity and the number of chronic diseases as potential confounders. Information on sex, age, highest attained education level, smoking, alcohol consumption, physical activity (number of days for more than 30 minutes), and chronic diseases was collected with a self-report questionnaire. We categorized education level as follows: low (no education, primary school or lower vocational training), intermediate (general secondary school or intermediate vocational training) and high (higher vocational training or university). Smoking was categorized as never, current and former smoking. Alcohol consumption was assessed with number, frequency and type of alcoholic drinks. Alcohol consumption was categorized as 0 glasses per day, 1-2 glasses per day and more than 2 glasses per day. To assess level of physical activity, the number of days with at least half an hour physical activity (including cycling, gardening or sports) in leisure time was used.

Chronic diseases present during the last year and diagnosed by a physician included asthma, COPD, severe heart disease, myocardial infarction, stroke, cancer, osteoarthritis, and rheumatoid arthritis. Weight and height were measured and body mass index was calculated.

Data-analysis

Relative risks of depressive symptoms for subjects with impaired FPG concentrations, undiagnosed and diagnosed type 2 diabetes were estimated using logistic regression and were expressed as odds ratios (ORs) with 95% confidence intervals (CI). Model 0 shows the crude ORs. In model 1, we adjusted for sex, age and education level. In model 2, we added body mass index, smoking status (never, current, former), alcohol consumption and physical activity to the model. In model 3, the number of chronic diseases, other than diabetes, was added to the model. These covariates were selected before the analyses. Sex, education level, smoking status and alcohol consumption were included as categorical covariates. Age, body mass index, physical activity and number of chronic diseases were included as continuous covariates. We performed the goodness of fit test as a measure of model fit.

Only 2.3% of the data we used was missing. However, if we performed complete case analysis, we lost 1,168 (24.6%) subjects, as these subjects had missing values on one or more variables. Missing data can severely affect the power of statistical analyses because patients with incomplete data are typically excluded. Furthermore, complete case analysis can give biased results because nonresponse is usually selective. Therefore, in the present study, multiple imputation [18] was used to fill in missing values under the assumption that these values only depend on observed values (missing at random). Five complete data sets were created, using the library Multivariate Imputation by Chained Equations [19] in the statistical package S-PLUS 2000. Although it depends on the percentage of missing data as to how many data sets should be created, in many applications 3-5 imputations are sufficient to obtain excellent results [20]. Subsequently, logistic regression modelling was performed on each data set as described above and the regression coefficients of the five models were pooled.

We reasoned that adjusting for cardiovascular disease in the model might not be appropriate as cardiovascular disease could, in part, be an intermediate factor in the association between type 2 diabetes and depressive symptoms. Therefore, we performed an additional analysis in which we excluded 79 subjects who reported severe heart disease, myocardial infarction or stroke.

All analyses were performed with SPSS version 12.0 (SPSS Inc., Chigago, Il) or S-PLUS 6.2 (Insightful Corp., Seattle, WA).

Results

The source population consisted of 4950 subjects aged 18 years or over. Questionnaire data were completely missing from 167 subjects and these subjects were excluded (mean age,

Table 1 Characteristics of the study population (observed data, before multiple imputation)

	Normal FPG *(< 5.6)* N=3499	**Impaired FPG** *(≥5.6 and < 7.0)* N=671	**Undiagnosed DM2** *(FPG ≥7.0)* N=55	**Diagnosed DM2** N=102
Sex, % male	40.4	63.6	43.6	52.9
N missing	*0*	*0*	*0*	*0*
Age, mean (SD)	37.7 (11.3)	46.7 (13.9)	56.9 (12.7)	55.8 (14.4)
N missing	*0*	*0*	*0*	*0*
Education, % low	17.1	28.5	51.9	44.9
% middle	42.9	43.9	34.6	40.8
% high	40.0	27.6	13.5	14.3
N missing	*114*	*19*	*3*	*4*
BMI (kg/m^2), mean (SD)	24.9 (4.0)	27.7 (4.3)	29.2 (6.1)	28.1 (4.6)
N missing	*6*	*1*	*0*	*0*
Smoking, % current	23.5	26.8	16.4	20.6
% former	28.1	37.3	58.2	41.2
N missing	*36*	*7*	*0*	*0*
Alcohol use, % 0 glasses/day	23.5	21.4	33.3	42.9
% 1-2 glasses/day	60.5	53.9	43.1	38.8
% >2 glasses/day	15.9	24.7	23.5	18.4
N missing	*106*	*35*	*4*	*4*
Physical activity (days/week), mean (SD)	3.2 (2)	3.2 (2)	3.3 (3)	3.3 (2)
N missing	*255*	*61*	*4*	*12*
Asthma/COPD, %	8.8	12.5	3.6	17.8
N missing	*42*	*5*	*0*	*1*
Severe heart disease or MI, %	0.6	2.4	3.6	9.8
N missing	*41*	*7*	*1*	*5*
Stroke, %	0.4	0.9	0.0	6.1
N missing	*40*	*4*	*0*	*3*
Cancer, %	0.5	2.0	1.8	5.9
N missing	*38*	*10*	*0*	*1*
Osteoarthritis, %	6.0	14.0	27.3	29.9
N missing	*57*	*7*	*0*	*5*
Arthritis, %	3.9	6.1	9.1	17.7
N missing	*85*	*13*	*0*	*6*
Depressive symptom score, mean (SD)	21.0 (7.2)	20.8 (7.1)	20.9 (6.3)	23.0 (9.2)
median (range)	18.0 (16-77)	18.0 (16-62)	19.0 (16-46)	19.0 (16-64)
N missing	*52*	*10*	*0*	*1*
Use of antidepressants, %	2.5	3.6	3.6	2.0
N missing	*19*	*6*	*0*	*0*
Depressive symptoms[1], %	19.4	17.5	20.0	29.7
N missing	*63*	*13*	*0*	*1*
Fasting plasma glucose, mean (SD)	4.9 (0.4)	5.9 (0.3)	8.3 (2.1)	8.3 (3.3)
N missing	*0*	*0*	*0*	*7*

FPG: fasting plasma glucose; DM2: type 2 diabetes mellitus; MI: myocardial infarction;

[1] defined as score of 25 or more on depression subscale of SCL-90 and/or self-reported use of antidepressants

(SD): 35.4 (11) years and 46.7% male subjects). Twenty- two subjects with only nonfasting plasma glucose concentration available and 14 subjects with type 1 diabetes were excluded leaving 4747 subjects in the study population. Of these subjects, 2121 (44.7%) were male and 2626 (55.3%) female and they had a mean age (SD) of 39.4 (12.5) years. In this population, the prevalence (95% CI) of impaired fasting glucose, undiagnosed and diagnosed type 2 diabetes was 15.5% (14.4-16.6%), 1.3% (0.96-1.6%) and 2.4% (1.9-2.9%), respectively. The overall prevalence (95% CI) of depressive symptoms (including antidepressant use) was 19.3% (18.2-20.4%). These percentages are the observed prevalences, i.e., before imputation.

Table 1 presents the subject characteristics in the four subgroups before multiple imputation was performed and the number of missing values. In 385 subjects, information on FPG concentration and diagnosed diabetes was missing and in 77 subjects depressive symptom score or information on antidepressant use was missing. Sex ratio, age, education level and body mass index differed considerably among the four subgroups. All chronic diseases were most prevalent in the patients with diagnosed type 2 diabetes. Mean depressive symptom score was highest in the group with diagnosed diabetes, whereas the percentage of subjects who used antidepressants was lowest in this group. The prevalence of depressive symptoms, defined as a score of 25 or more on the depression subscale of the SCL-90, and/or use of antidepressants, was 20.0% in the patients with undiagnosed and 29.7% in the patients with diagnosed type 2 diabetes. Mean FPG levels were similar in the undiagnosed and diagnosed type 2 diabetes patients, although the variance was higher in the patients with diagnosed diabetes.

Table 2 presents the crude and adjusted ORs and 95% CIs of depressive symptoms for impaired FPG concentration and undiagnosed and diagnosed type 2 diabetes. These are the effect estimates after multiple imputation was performed. Therefore the numbers of subjects in the exposure groups are higher than the numbers presented in Table 1. Compared with the normal glucose group, the odds of depressive symptoms was not higher in the impaired glucose group (OR=0.90 (0.72-1.12)) and the patients with undiagnosed type 2 diabetes (OR=1.00 (0.51-1.96); model 0). Adjustment for demographic, lifestyle variables and the number of chronic diseases did not change this finding. In the diagnosed type 2 diabetes group, however, the odds of depressive symptoms was increased (OR=1.79 (1.17-2.75); model 0). Adjustment for sex, age, education level, body mass index, smoking status, alcohol consumption and physical activity did not change this risk estimate (model 2). Adding the number of chronic diseases to the model lowered the OR for diagnosed type 2 diabetes (OR=1.36 (0.83-2.23); model 3). The model adjusted for all demographic and lifestyle

variables (model 2) and the model additionally adjusted for the number of chronic diseases (model 3) showed no deviance of goodness of fit with p values of 0.47 and 0.29, respectively. Model 1 did show deviance of goodness of fit with a p value of 0.01, but we considered this less relevant as this was not our primary model of interest.

Table 2 Association of impaired fasting plasma glucose concentration, undiagnosed and diagnosed type 2 diabetes with depressive symptoms, expressed as crude and adjusted odds ratio (95% confidence interval) (after multiple imputation)

	Normal FPG#	**Impaired FPG**	**Undiagnosed DM2**	**Diagnosed DM2**
	(< 5.6)	*(≥ 5.6 and < 7.0)*	*(FPG ≥ 7.0)*	
N	3853	732	58	104
Model 0	1.00	0.90 (0.72-1.12)	1.00 (0.51-1.96)	1.79 (1.17-2.75)
Model 1*	1.00	1.03 (0.81-1.31)	0.84 (0.42-1.69)	1.70 (1.07-2.70)
Model 2**	1.00	1.01 (0.78-1.29)	0.82 (0.40-1.68)	1.69 (1.06-2.72)
Model 3***	1.00	0.99 (0.77-1.27)	0.86 (0.42-1.77)	1.36 (0.83-2.23)

reference category; FPG: fasting plasma glucose; DM2: type 2 diabetes mellitus;

* adjusted for sex, age and education;

** adjusted for sex, age, education, body mass index, current and former smoking, alcohol consumption, physical activity;

*** adjusted for sex, age, education, body mass index, current and former smoking, alcohol consumption, physical activity and number of chronic diseases.

When excluding 79 subjects with cardiovascular disease, the ORs for the impaired fasting glucose group and the patients with undiagnosed diabetes were comparable with those presented in Table 2 (data not shown). For the patients with diagnosed type 2 diabetes the ORs were slightly higher than when subjects with cardiovascular disease were not excluded: when adjusting for sex, age and education (model 1), the OR for patients with diagnosed type 2 diabetes was 1.92 (95% CI 1.17-3.14). Additional adjustment for lifestyle factors (model 2) resulted in an OR of 1.95 (95% CI 1.18-3.22). The OR was 1.77 (95% CI 1.06-2.96) when adjusting for the number of (other) chronic diseases (model 3).

In the results presented above we categorized the FPG levels into three groups, which reflect clinical practice. We performed a post-hoc analysis where we included FPG level as a continuous variable and diagnosis of diabetes as a dichotomous variable into the model. The crude OR for continuous FPG (per 1 mmol/l increase) was 0.91 (95% CI 0.83-1.01) and the crude OR for a diagnosis of diabetes was 2.38 (95% CI 1.43-3.95). Adjusting for age, sex and education level (model 1) changed the ORs into 0.98 (95% CI 0.89-1.07) and 1.79 (95% CI 1.07-3.01). Additional adjustment for lifestyle factors and the number of chronic diseases (model 3) resulted in an OR for FPG level of 0.99 (95% CI 0.90-1.09) and for a diagnosis of

diabetes of 1.40 (95% CI 0.81-2.41). Again, model 2 and 3 did not show deviance of goodness of fit and model 1 showed deviance of goodness of fit. These results are similar to the results described above; namely, having a diagnosis of diabetes increased the risk of depressive symptoms while no association was seen between FPG and depressive symptoms.

Discussion

In the present study, subjects with impaired FPG concentration and undiagnosed type 2 diabetes did not have an increased risk of depressive symptoms. Patients with diagnosed type 2 diabetes had a 1.7 times increased risk of depressive symptoms compared with subjects with normal glucose concentrations after adjustment for demographic and lifestyle variables. After additional adjustment for number of chronic diseases, the risk of depressive symptoms was no longer significantly increased in patients with diagnosed diabetes.

These results suggest that disturbed glucose homeostasis is not associated with depressive symptoms. As subjects aware of their diabetes had an increased risk of depressive symptoms and subjects unaware of their diabetes had not, the increased risk of depressive symptoms in diabetes might be a consequence of its burden rather than a consequence of high glucose levels. The number of comorbid chronic diseases explains part of the increased prevalence seen in diagnosed type 2 diabetes patients.

Depressive symptoms might be a consequence of diagnosed diabetes. However, as this study has a cross-sectional design, we cannot distinguish between causes and consequences. Diagnosed diabetes can also be a consequence of having depressive symptoms. This may occur if subjects with depressive symptoms are more likely to be diagnosed with type 2 diabetes because they more often consult a physician, i.e., detection bias.

It is possible that we found a high prevalence of depressive symptoms in the patients with diagnosed type 2 diabetes and not in subjects in earlier stages of diabetes because patients with diagnosed diabetes have a higher level and longer duration of disturbed glucose homeostasis. Although, if this were the case, we would expect to see a trend between increasing glucose levels and depressive symptoms, which we did not.

Our findings support those of Palinkas et al [21], who showed an increased prevalence of depression in patients with previously diagnosed type 2 diabetes and no increased prevalence in patients unaware of their diabetes, suggesting that depression develops after a diagnosis of diabetes. The almost two-fold increased prevalence of depressive symptoms we observed in diagnosed type 2 diabetes is also in accordance with the results of a meta-analysis, showing an OR of 2 when pooling 20 cross-sectional studies [3]. Further, our finding of the absence of an association between fasting glucose concentrations and depressive

symptoms is not inconsistent with the results of two recent studies showing a negative [22] and a positive association [23] between insulin resistance and depression. Also in a recent prospective study on insulin resistance and depressive symptoms in middle aged men no association was found [24]. However, in that study, no association between clinically manifest diabetes and depressive symptoms was found either. Another prospective study showed no association between patients with diagnosed diabetes and those with depressive symptoms [25].

Two recent studies on factors associated with depression in type 2 diabetes showed that patients with diabetes and comorbid chronic somatic diseases had an increased prevalence of depression, whereas those without diabetes did not [26,27]. In our study, the number of chronic diseases also explained part of the increased prevalence of depressive symptoms in patients with diagnosed type 2 diabetes. Interestingly, a prospective study showed no increased risk of developing depression in patients with diabetes compared with those with osteoarthritis [28]. This suggests nonspecificity of diabetes as a risk factor for depression and is consistent with our finding that depression in patients with diabetes may be a consequence of the burden of the disease rather than a result of physiologic changes due to high glucose levels.

Some limitations of our study need to be addressed. First, we used self-report to define diagnosis of type 2 diabetes. It is possible that some misclassification occurred in that patients with diagnosed diabetes misclassified themselves as not diagnosed. Yet, a Dutch validation study showed that self-reported diagnosis of diabetes is in good agreement with a physician diagnosis [29]. Furthermore, if misclassification was random, our estimated relative risk of depressive symptoms in patients with diagnosed diabetes is an underestimation. Type 1 diabetes was defined as use of insulin and no use of oral hypoglycemic agents; we excluded 14 subjects based on this definition. It is possible that some of these subjects were patients with type 2 diabetes after all. An additional analysis showed that including or excluding these subjects did not change the association between diagnosed diabetes and depressive symptoms.

Second, we did not use a diagnostic interview to confirm depression according to the DSM-IV criteria. However, our aim was to compare different groups regarding depressive symptom level and not to obtain precise estimates of the prevalence of depression in these groups. Analyzing the depression score continuously might have given more power to detect an association. However, we used the validated cut off point of 25 described in the literature [16,17]. Also, two practical reasons to dichotomize the depressive symptom score were that the variable was not normally distributed, which could not be solved by several transformations,

and that we also included users of antidepressant medication into the depressive symptom group. Regarding the cut off point of 25, one study also suggested a cut off point of 27 for major and minor depression with a sensitivity of 81.1 and specificity of 83.5 [17]. Use of this more conservative cut off point did not change our findings.

Third, antidepressant use is not solely prescribed for depression but also for other psychiatric problems such as anxiety and panic disorders; this prescription practice can have confounded our results. However, in our study, only 5.9% of the subjects with depressive symptoms were defined as such based on antidepressant use. Excluding subjects who used antidepressants resulted in a slightly higher OR in the group with diagnosed type 2 diabetes. In addition, analysing antidepressant use as a covariate, instead of as an indicator of depression on the dependent variable side, did not change our conclusion.

Fourth, as the response rate of the study was about 50%, selection bias could have occurred. Nonresponders were more often male and were somewhat younger than responders, which may have influenced the prevalence of diabetes and depressive symptoms but most likely not the association between diabetes and depressive symptoms [30]. Potential selection bias through missing data was minimized because we performed multiple imputation to impute missing values. Multiple imputation is the preferred method to deal with missing values because it leads to unbiased estimates and correct standard errors as opposed to complete case analysis or single imputation techniques [31]. Because our study population was young, the number of patients with undiagnosed and diagnosed type 2 diabetes was relatively low. Therefore, the study may have had limited power. Regarding generalizability, our results may not apply to an elderly population because of our relatively young study population.

A strong aspect of our study is that we used a large sample from the general population. Second, we adjusted for the most important factors known from the literature that could disturb the association between type 2 diabetes and depressive symptoms. Third, the design of the UHP made it possible to compare subjects who were unaware of their diabetes with subjects who were aware of their diabetes within a single population and therefore we could make some inferences about the reason for the increased prevalence of depressive symptoms in patients with type 2 diabetes.

In conclusion, the results of the present study suggest that disturbed glucose homeostasis is not associated with depressive symptoms. The association we found between diagnosed diabetes and depressive symptoms suggests that the increased prevalence of depressive symptoms in patients with type 2 diabetes is a consequence of the burden of the disease. Furthermore, the number of chronic diseases seems to partly explain the

association between diagnosed type 2 diabetes and depressive symptoms.

More research is needed to confirm the association between type 2 diabetes and depressive symptoms. Instead of cross-sectional studies, longitudinal studies should be performed to investigate if diabetes is an independent risk factor for the onset of depressive symptoms. In these studies, the influence of a history of depression and the influence of comorbid chronic diseases should be taken into account. Furthermore, future research should clarify if the onset of depressive symptoms is specific for type 2 diabetes or if it is also present in other chronic diseases. Finally, the mechanisms underlying the association between diabetes and depression should be investigated.

References

1. Wild S, Roglic G, Green A, Sicree R, King H. Global Prevalence of Diabetes: Estimates for the Year 2000 and Projections for 2030. *Diabetes Care* 2004;27:1047-53.
2. Murray CJ and Lopez AD. Alternative Projections of Mortality and Disability by Cause 1990-2020: Global Burden of Disease Study. *Lancet* 1997;349:1498-504.
3. Anderson RJ, Freedland KE, Clouse RE, Lustman PJ. The Prevalence of Comorbid Depression in Adults With Diabetes: a Meta-Analysis. *Diabetes Care* 2001;24:1069-78.
4. Knol MJ, Twisk JW, Beekman AT, Heine RJ, Snoek FJ, Pouwer F. Depression As a Risk Factor for the Onset of Type 2 Diabetes Mellitus. A Meta-Analysis. *Diabetologia* 2006;49:837-45.
5. Talbot F and Nouwen A. A Review of the Relationship Between Depression and Diabetes in Adults: Is There a Link? *Diabetes Care* 2000;23:1556-62.
6. Cameron OG, Kronfol Z, Greden JF, Carroll BJ. Hypothalamic-Pituitary-Adrenocortical Activity in Patients With Diabetes Mellitus. *Arch Gen Psychiatry* 1984;41:1090-5.
7. Chan O, Inouye K, Riddell MC, Vranic M, Matthews SG. Diabetes and the Hypothalamo-Pituitary-Adrenal (HPA) Axis. *Minerva Endocrinol* 2003;28:87-102.
8. Musselman DL, Betan E, Larsen H, Phillips LS. Relationship of Depression to Diabetes Types 1 and 2: Epidemiology, Biology, and Treatment. *Biol Psychiatry* 2003;54:317-29.
9. Peyrot M and Rubin RR. Persistence of Depressive Symptoms in Diabetic Adults. *Diabetes Care* 1999;22:448-52.
10. Katon WJ. Clinical and Health Services Relationships Between Major Depression, Depressive Symptoms, and General Medical Illness. *Biol Psychiatry* 2003;54:216-26.
11. Grobbee DE, Hoes AW, Verheij TJ, Schrijvers AJ, van Ameijden EJ, Numans ME. The Utrecht Health Project: Optimization of Routine Healthcare Data for Research. *Eur J Epidemiol* 2005;20:285-7.
12. Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* 2005;28:S37-S42.
13. Stahl M, Brandslund I, Jorgensen LG, Hyltoft PP, Borch-Johnsen K, de Fine ON. Can Capillary Whole Blood Glucose and Venous Plasma Glucose Measurements Be Used Interchangeably in Diagnosis of Diabetes Mellitus? *Scand J Clin Lab Invest* 2002;62:159-66.
14. Derogatis LR, Lipman RS, Covi L. SCL-90: an Outpatient Psychiatric Rating Scale--Preliminary Report. *Psychopharmacol Bull* 1973;9:13-28.

15. Arrindell WA and Ettema JHM. Dimensionele Structuur, Betrouwbaarheid En Validiteit Van De Nederlandse Bewerking Van De Symptom Checklist (SCL-90). *Nederlands Tijdschrift Psychologie* 1981;43:381-7.
16. Aben I, Verhey F, Lousberg R, Lodder J, Honig A. Validity of the Beck Depression Inventory, Hospital Anxiety and Depression Scale, SCL-90, and Hamilton Depression Rating Scale As Screening Instruments for Depression in Stroke Patients. *Psychosomatics* 2002;43:386-93.
17. Strik JJ, Honig A, Lousberg R, Denollet J. Sensitivity and Specificity of Observer and Self-Report Questionnaires in Major and Minor Depression Following Myocardial Infarction. *Psychosomatics* 2001;42:423-8.
18. Rubin DB and Schenker N. Multiple Imputation in Health-Care Databases: an Overview and Some Applications. *Stat Med* 1991;10:585-98.
19. Buuren van S and Oudshoorn K. *Flexible Multivariate Imputation by MICE*, TNO Prevention and Health, Leiden, 1999.
20. Schafer JL and Olsen MK. Multiple Imputation for Multivariate Missing-data Problems: A Data Analyst's Perspective. *Multivariate Behavioral Research* 1998;33:545-71.
21. Palinkas LA, Barrett-Connor E, Wingard DL. Type 2 Diabetes and Depressive Symptoms in Older Adults: a Population-Based Study. *Diabet Med* 1991;8:532-9.
22. Lawlor DA, Smith GD, Ebrahim S. Association of Insulin Resistance With Depression: Cross Sectional Findings From the British Women's Heart and Health Study. *BMJ* 2003;327:1383-4.
23. Timonen M, Laakso M, Jokelainen J, Rajala U, Meyer-Rochow VB, Keinanen-Kiukaanniemi S. Insulin Resistance and Depression: Cross Sectional Study. *BMJ* 2005;330:17-8.
24. Lawlor DA, Ben Shlomo Y, Ebrahim S, Davey SG, Stansfeld SA, Yarnell JW et al. Insulin Resistance and Depressive Symptoms in Middle Aged Men: Findings From the Caerphilly Prospective Cohort Study. *BMJ* 2005;330:705-6.
25. Palinkas LA, Lee PP, Barrett-Connor E. A Prospective Study of Type 2 Diabetes and Depressive Symptoms in the Elderly: the Rancho Bernardo Study. *Diabet Med* 2004;21:1185-91.
26. Engum A, Mykletun A, Midthjell K, Holen A, Dahl AA. Depression and Diabetes: A Large Population-Based Study of Sociodemographic, Lifestyle, and Clinical Factors Associated With Depression in Type 1 and Type 2 Diabetes. *Diabetes Care* 2005;28:1904-9.
27. Pouwer F, Beekman AT, Nijpels G, Dekker JM, Snoek FJ, Kostense PJ et al. Rates and Risks for Co-Morbid Depression in Patients With Type 2 Diabetes Mellitus: Results From a Community-Based Study. *Diabetologia* 2003;46:892-8.
28. Kessing LV, Nilsson FM, Siersma V, Andersen PK. No Increased Risk of Developing Depression in Diabetes Compared to Other Chronic Illness. *Diabetes Res Clin Pract* 2003;62:113-21.
29. Klungel OH, de Boer A, Paes AH, Seidell JC, Bakker A. Cardiovascular Diseases and Risk Factors in a Population-Based Study in The Netherlands: Agreement Between Questionnaire Information and Medical Records. *Neth J Med* 1999;55:177-83.
30. Van Loon AJ, Tijhuis M, Picavet HS, Surtees PG, Ormel J. Survey Non-Response in the Netherlands: Effects on Prevalence Estimates and Associations. *Ann Epidemiol* 2003;13:105-10.
31. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a Gentle Introduction to Imputation of Missing Values. *J Clin Epidemiol* 2006;59:1087-91.

Chapter 2.5

Antidepressant use before and after the initiation of diabetes mellitus treatment

Abstract

Background

Although current evidence suggests an association between diabetes and depression, its direction is unclear. We examined the temporal association between diabetes and depression by studying antidepressant and benzodiazepine use around the initiation of diabetes treatment.

Methods

From a pharmacy registry database we selected 49,593 diabetic patients and a random sample of non-diabetic subjects (n=154,441), all >= 40 years old. Incidence of antidepressant and benzodiazepine use was calculated during seven years before and seven years after the initiation of diabetes medication. A random index date was assigned to non-diabetic subjects. Time-specific incidence rate ratios of antidepressant and benzodiazepine use were calculated in one-year, three-months and one-month intervals.

Results

The incidence of antidepressant and benzodiazepine use was increased during the two months before and three months after the initiation of diabetes treatment compared with non-diabetic subjects. The strongest increase in incidence of antidepressant and benzodiazepine use was seen in the month after the initiation of diabetes treatment with an incidence rate ratio of 2.4 (95% CI: 2.0-3.0) and 3.4 (95% CI: 3.0-3.8), after adjustment for age, sex and Chronic Disease Score.

Conclusions

The short time period between initiation of diabetes treatment and initiation of antidepressant and benzodiazepine use makes it unlikely that diabetes and depression are causally related. A plausible explanation is that depression and psychosocial complaints are more likely to be detected and treated if patients present with diabetes-related complaints to a physician, or vice versa, that diabetes is more likely to be detected if patients present with depression or psychosocial complaints.

Introduction

There is growing evidence that the prevalence of depression is increased among diabetic patients compared with subjects without diabetes. A recent meta-analysis pooled ten cross-sectional studies and reported an overall odds ratio of 1.8 [1]. However, the reason for this co-occurrence of diabetes and depression is not yet clear.

Depression might be a risk factor for developing diabetes. A pooled analysis of nine longitudinal studies showed a 1.4-fold increased risk of diabetes in depressed subjects [2]. Two subsequently published longitudinal studies confirmed a significantly increased risk of onset of type 2 diabetes in depressed subjects [3,4], whereas two other studies did not find an increased risk of diabetes in subjects older than 50 years [5] or in antidepressant users [6].

Alternatively, diabetes might be a risk factor for onset of depression. Most longitudinal studies did not find a significant association [4,7-10]. A recent study among well-functioning older adults found an increased risk of recurrent depressed mood among diabetic patients [11].

Despite the longitudinal design of these studies, they had some disadvantages. Most studies measured depression and diabetes only at two points in time with unknown events happening between these measurements. Only two studies that investigated depression as a risk factor for diabetes used incident cases of depression [6,12], and only three studies that investigated diabetes as a risk factor for depression used incident cases of diabetes [8-10]. The selection of incident cases is preferable over the selection of prevalent cases as the exposed and unexposed group are more comparable with respect to other characteristics. In addition, regarding the studies investigating diabetes as a risk factor for depression, none of these studies assessed the incidence of depression directly after the diagnosis of diabetes. Two years was the shortest period over which the incidence of depression was calculated [10]. One study assessed the incidence of depression continuously over time, but did not calculate time-specific incidence rates [9]. It is important to know when the risk of depression increases after a diagnosis of diabetes. To study this, depression should be continuously monitored.

The objective of our study was to investigate the temporal association between diabetes and depression by studying antidepressant and benzodiazepine use around the initiation of diabetes treatment. We studied antidepressant use as well as benzodiazepine use to investigate whether the association with diabetes was specific for depression or whether diabetes was also related to more general psychosocial complaints.

Methods

Source population

Our cohort was selected from the PHARMO database, as described in detail elsewhere [13]. In short, the database comprises all pharmacy dispensing records of residents of about 50 Dutch municipalities, counting for approximately two million patient histories. Since virtually all patients in the Netherlands are registered with a single community pharmacy, pharmacy records are almost complete with regard to prescription drugs independent of the prescriber. In the Netherlands, antidepressants, benzodiazepines, oral hypoglycemic agents and insulin are only available as prescription drugs. Therefore, pharmacy data will cover all use of these drugs. Variables included in the PHARMO database are sex, date of birth, dispensed drugs (coded according to the Anatomical Therapeutic Chemical (ATC) classification), drug dispensing date, amount of drug dispensed, prescribed dosage regimen and type of prescriber.

Study population

From the PHARMO database, we selected all subjects who had a prescription of any antidiabetic drug (ATC code = A10A* and A10B*) between January 1st 1996 and December 31st 2006, and were born in 1965 or before, i.e. were older than 40 years on December 31st 2006 (n=97,667). We randomly selected 195,334 subjects (twice the sample size of the group of diabetic patients) who were also older than 40 years at December 31st 2006 but did not have a prescription of any antidiabetic drug between January 1st 1996 and December 31st 2006. Because we aimed to include incident diabetes cases in the study population we excluded subjects who had less than six months between their first known prescription in the PHARMO database and their first prescription of the antidiabetic drug (n=47,123). In the Netherlands, drugs are prescribed for a maximum period of three months and therefore we chose six months to be sure to exclude prevalent diabetes cases. The index date was defined as the initiation of antidiabetic drug use in those with a prescription for antidiabetic drugs. Subjects without a prescription for antidiabetic drugs were assigned a random index date in such a way that the distribution of the index date relative to their total follow-up time was similar to the distribution of the index date relative to the total follow-up time in the diabetic patients. To ensure consistency, we excluded non-diabetic subjects who had less than six months between their first prescription in the PHARMO database and their index date (n=26,364). Despite the restriction of only including subjects older than 40 years at December 31st 2006, subjects could be younger than 40 years at the index date as the index date was earlier than December 31st 2006 for almost all subjects. We therefore excluded

subjects who were younger than 40 years of age at the index date (diabetes: n=821; non-diabetes: n=14,169). Finally, a small proportion of subjects were excluded because of missing or incorrect antidepressant or benzodiazepine prescription data, for example a missing ATC code or extremely low (< 1) or extremely high (> 1000) amount of tablets dispensed with one prescription (diabetes: n=130; non-diabetes: n=240). This resulted in a study population of 154,441 non-diabetic subjects and 49,593 diabetic subjects to be analyzed in this study.

Incidence of antidepressant and benzodiazepine use

The start of an episode of antidepressant (AD) or benzodiazepine (BD) use was defined as the first prescription for any AD (ATC code = N06A*) or BD (ATC code = N05BA*, N05CD* or N05CF*) within the study period. The end of an episode of AD or BD use was defined as no new prescription of any AD or BD within six months after the end date of the last AD or BD prescription. The end date of a prescription was calculated by adding the duration of AD or BD use (total amount dispensed divided by prescribed daily dose) to the start date of the prescription. The start of the first episode of AD or BD use was defined as an incident case if a period of at least six months was present between the first known prescription in the PHARMO database and the start of the episode.

Covariates

Age at index date, sex and the Chronic Disease Score (CDS) were considered as covariates. The CDS is a measure of the chronic disease status among drug users and can be considered as an indicator of an individual's morbidity and overall health status. Exposure to various prescription drugs has been shown to be a valid measure of chronic diseases [14]. The CDS includes the major chronic diseases such as heart disease, respiratory illness, cancer, ulcers and high cholesterol. The CDS was calculated over the period of one year before the index date as a chronic disease measure at baseline.

Data-analysis

In theory, we had data from eleven years before the index date to eleven years after the index date. In practice, we had enough person-years from seven years before until seven years after the index date and therefore we analyzed this time frame.

We divided these seven years before and after the index date in intervals of one year to investigate the pattern of incident AD and BD use over time. Within these intervals of one year we calculated the number of incident AD and BD users and the amount of person-years contributing to these incident cases in both diabetes and non-diabetic subjects. Cox

regression was used to calculate incidence rate ratios and 95% confidence intervals within each one year interval, adjusted for age, sex and the CDS. In this study, the incidence rate ratio represented the risk of a first AD or BD episode in diabetic patients relative to the non-diabetic subjects. To study the period around the index date in more detail, we subsequently studied the 2.5 years before and after the index date and divided this period in intervals of three months, and again calculated incidence rate ratios as described above. Finally, we studied the ten months before and after the index date and divided this period in intervals of one month to have a detailed picture of the incidence rate ratios around the index date.

We compared the prevalence of AD and BD use between diabetic patients and non-diabetic subjects using logistic regression, in subjects still at risk for a first episode of AD or BD use at the index date. In addition, the length of the first episode of AD and BD use was compared between subjects with and without diabetes, in subjects with a first episode of AD or BD use after the index date.

Table 1 Baseline characteristics of the study participants

	No diabetes	**Diabetes**
N	154,441	49,593
Age, mean (sd)	58.0 (13)	64.6 (12)
Male sex	46.2%	48.9%
CDS one year before index date		
0	54.6%	28.9%
1-3	27.0%	30.3%
>3	18.5%	40.8%
Mean follow-up, year	7.2	8.2
Median follow-up, year	7.9	8.6
Number of AD episodes		
0	83.5%	79.1%
1	12.1%	15.2%
2-4	4.3%	5.5%
> 4	0.1%	0.2%
Number of BD episodes		
0	55.8%	46.5%
1	25.5%	30.3%
2-4	16.2%	20.2%
> 4	2.5%	3.0%
Number of subjects with incident AD use	18,065	7631
Number of subjects with incident BD use	42,381	16,404

CDS: Chronic Disease Score; AD: antidepressant; BD: benzodiazepine

Results

Table 1 gives the baseline characteristics of the study population. Diabetic patients were on average 6.6 years older than the non-diabetic subjects and were more often male (48.9% versus 46.2%). Total follow-up time was slightly higher in diabetic patients. Almost 20% of the subjects had at least one AD episode and the proportion of subjects with more than one AD episode was small (4.7%). In contrast, almost half of the subjects had at least one episode of BD use and 19.8% had more than one episode. Both AD and BD episodes were more frequent among diabetic patients compared with non-diabetic subjects.

When examining the seven years before and seven years after the index date, the incidence rate ratio of AD use was significantly increased in diabetic patients compared with non-diabetic subjects in the first year after the index date and was not increased in the other one year intervals (Figure 1, upper graph). The middle graph of Figure 1 presents a detailed part of the upper graph showing 2.5 years before and 2.5 years after the index date. The lower graph of Figure 1 presents again a detailed part of the middle graph showing ten months before and ten months after the index date. In the month before the index date the incidence rate of AD use was increased in the diabetic patients compared with the non-diabetic patients (adjusted incidence rate ratio = 1.42, 95% CI: 1.14-1.78). In the first month after the index date the incidence rate ratio peaked with a 2.4 fold increased incidence rate in diabetic patients. In the second and third month after the index date the incidence rate in the diabetic patients decreased again and was almost similar to the rate in the non-diabetic patients. Until two years after the index date the prevalence of AD use was still increased with at least 20% (OR>1.2) among subjects at risk for a first episode of AD use at the index date (data not shown). The median duration of the incident episodes of AD use after the index date was not significantly higher in diabetic patients (120 days) than in subjects without diabetes (114 days).

Figure 2 presents the incidence rate ratios of BD use in diabetic subjects compared with non-diabetic subjects in one year intervals (upper graph), three months intervals (middle graph) and one month intervals (lower graph). The incidence rate ratio of BD use was increased two months before the index date (1.56, 95% CI: 1.34-1.82). Similar to AD use, the incidence rate ratio for BD use peaked in the month after the index date (3.38, 95% CI: 3.00-3.83) and decreased again in the second and third month after the index date. The prevalence of BD use was increased with at least 20% (OR>1.2) until 15 months after the index date among subjects at risk for a first episode of BD use at the index date (data not shown). The median duration of the incident episodes of BD use after the index date was significantly higher in diabetic patients (20 days) than in subjects without diabetes (15 days).

Figure 1 Incidence rate ratio (solid line) and 95% confidence interval (CI; dashed lines), adjusted for age, sex and chronic disease score, of antidepressant use around the index date ranging from seven years before to seven years after the index date (upper graph), ranging from 2.5 years before to 2.5 years after the index date (middle graph) and ranging from ten months before to ten months after the index date (lower graph)

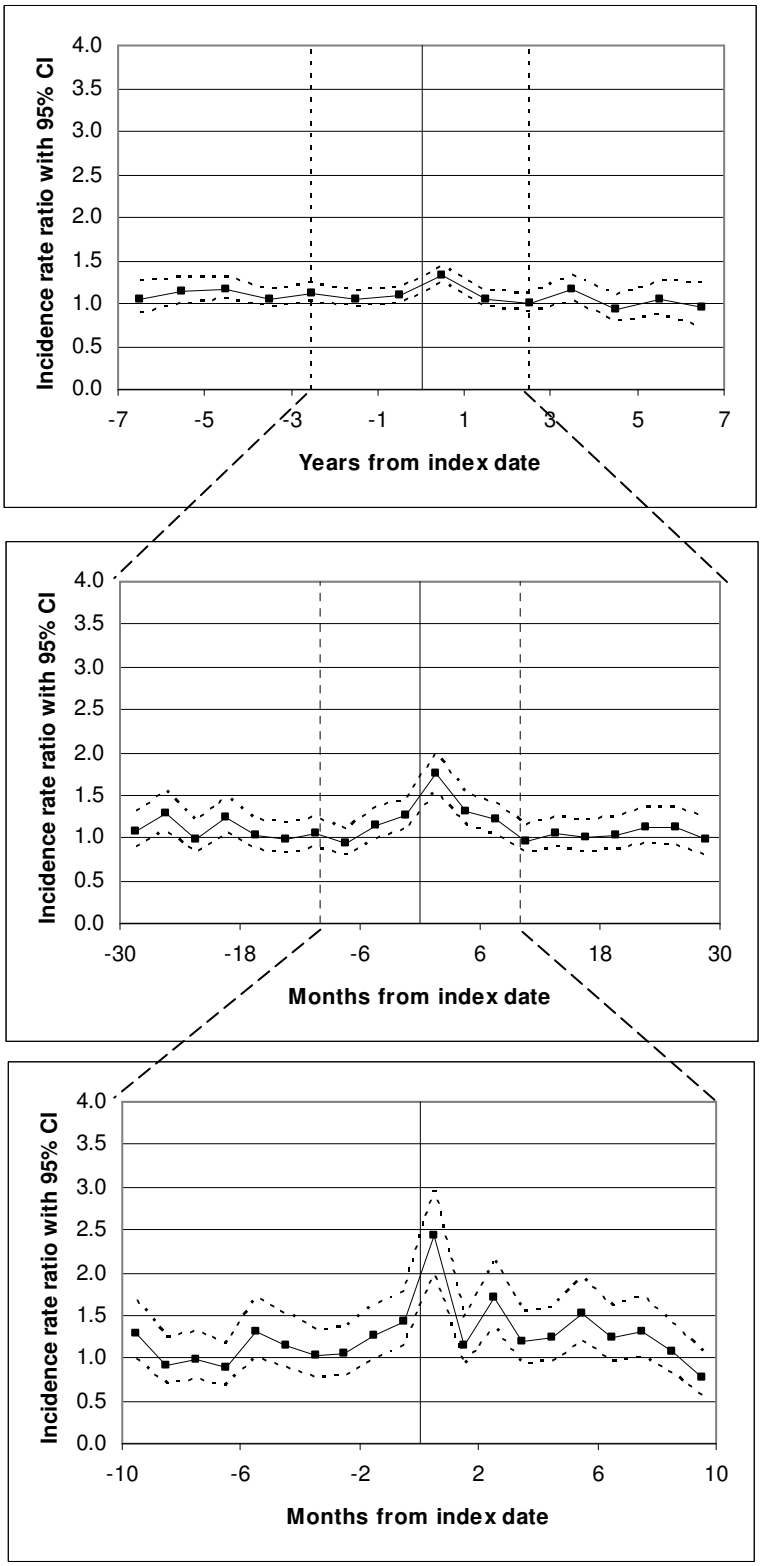


Figure 2 Incidence rate ratio (solid line) and 95% confidence interval (CI; dashed lines), adjusted for age, sex and chronic disease score, of benzodiazepine use around the index date ranging from seven years before to seven years after the index date (upper graph), ranging from 2.5 years before to 2.5 years after the index date (middle graph) and ranging from ten months before to ten months after the index date (lower graph)

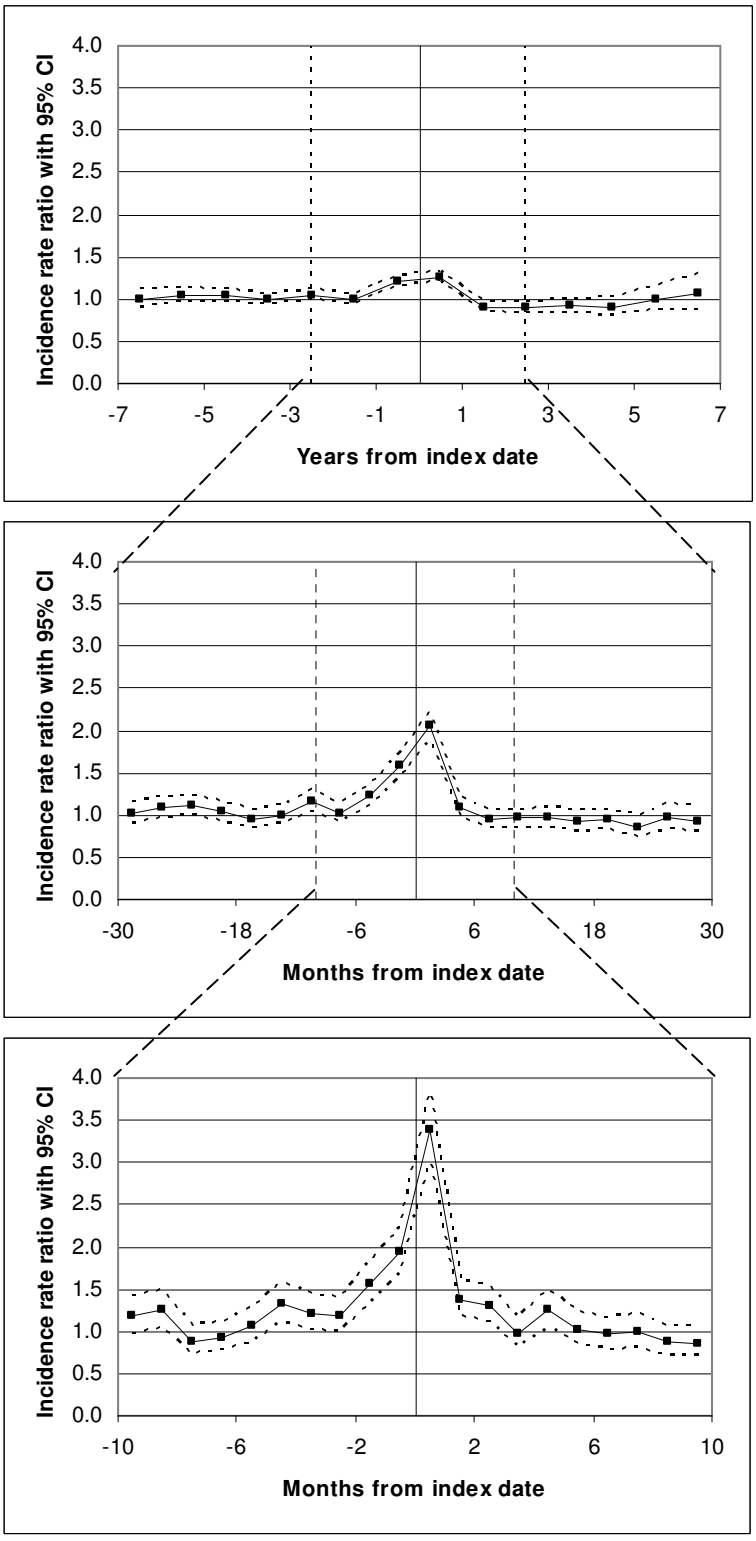


Table 2 Incidence rate in non-diabetic subjects and diabetic patients (/1000 person-years), crude and adjusted incidence rate ratios and 95% confidence intervals of antidepressant and benzodiazepine use in the three months before and after the index date

	Months before index date			**Months after index date**		
	3	*2*	*1*	*1*	*2*	*3*
Antidepressant use						
Incidence rate in no diabetes	0.053	0.050	0.064	0.056	0.060	0.059
Incidence rate in diabetes	0.066	0.079	0.100	0.157	0.082	0.109
Crude	1.24 (0.97-1.59)	1.57 (1.24-1.99)	1.56 (1.27-1.92)	2.77 (2.29-3.34)	1.37 (1.09-1.72)	1.85 (1.50-2.29)
Adjusted for age and sex	1.22 (0.94-1.58)	1.54 (1.21-1.97)	1.54 (1.21-1.97)	2.66 (2.19-3.24)	1.30 (1.02-1.64)	1.89 (1.52-2.35)
Adjusted for age, sex and CDS	1.04 (0.80-1.35)	1.27 (0.99-1.62)	1.42 (1.14-1.78)	2.43 (1.99-2.97)	1.15 (0.90-1.47)	1.71 (1.36-2.13)
Benzodiazepine use						
Incidence rate in no diabetes	0.162	0.163	0.174	0.172	0.179	0.171
Incidence rate in diabetes	0.223	0.296	0.381	0.691	0.295	0.250
Crude	1.38 (1.18-1.62)	1.82 (1.57-2.10)	2.20 (1.92-2.51)	3.90 (3.47-4.38)	1.65 (1.42-1.91)	1.46 (1.25-1.72)
Adjusted for age and sex	1.37 (1.16-1.61)	1.82 (1.57-2.12)	2.12 (1.84-2.43)	3.53 (3.13-3.98)	1.49 (1.28-1.74)	1.40 (1.18-1.65)
Adjusted for age, sex and CDS	1.18 (0.99-1.39)	1.56 (1.34-1.82)	1.94 (1.68-2.23)	3.38 (3.00-3.83)	1.37 (1.17-1.60)	1.31 (1.11-1.55)

CDS: Chronic Disease Score

The median duration of an episode of BD use was markedly lower than the duration of an episode of AD use. This is also reflected in the shorter period of increased prevalence for BD use compared with AD use after the index date.

Table 2 presents more detailed information on the absolute and relative risks for AD and BD use in the three months before and the three months after the index date, because in these months there were significant differences between diabetes and non-diabetic patients. Adjustment for age and sex did not substantially change the incidence rate ratios for AD and BD use, whereas additional adjustment for the CDS lowered the incidence rate ratios (Table 2).

Discussion

We investigated the incident use of AD and BD seven years before and seven years after the initiation of diabetes treatment. The incidence of AD use was only increased shortly before and shortly after starting diabetes treatment. The incidence of BD use was similarly increased shortly before and shortly after the index date in diabetic patients compared with non-diabetic patients. A high peak in incidence in the use of these drugs by diabetic patients was seen in the month after the index date with a 2.4- and 3.4-fold increased incidence rate for AD and BD use respectively. Although the incidence of AD and BD use was increased only shortly after the index date, prevalence of AD use was higher up to two years after the initiation of diabetes treatment and prevalence of BD use was increased until 15 months after initiation of diabetes treatment.

Before interpreting the results we will address some limitations of this study. We used antidepressant use as a proxy for depression. A limitation of this approach is that not all depressed subjects are treated with antidepressants and that not all antidepressant users have an indication of depression [15], which will have resulted in misclassification. However, more severe depression is likely to be recognized and treated [16], and therefore antidepressant use may be a valid proxy for more severe and debilitating depression. Benzodiazepine use was used as a proxy for more general psychosocial complaints. Insomnia and anxiety are the main indications for benzodiazepine use, but in practice benzodiazepines are prescribed more broadly for psychosocial related indications [17]. We therefore consider benzodiazepine use as a good proxy for general psychosocial complaints. Although misclassification may have occurred, an important strength of this study is that information on AD and BD use was available at every point in time. These continuous measurements in time gave us the opportunity to calculate time-specific incidence rates of AD and BD use to study the association between depression and diabetes in detail. As we measured diabetes

as initiation of diabetes medication, we missed undiagnosed and untreated diabetic patients. Therefore, reversed causality could be an issue in this study. We adjusted for age, sex and the CDS in the analyses but had no information about other possible confounders such as body mass index and physical activity. If these confounders are positively associated with both depression and diabetes, our estimates would be overestimations of the true effects.

This study showed an increased incidence of AD and BD use shortly *after* the initiation of diabetes medication. This suggests that there is no long lasting effect of diabetes on incidence of depression or psychosocial complaints, although the prevalence of AD and BD use remained high for more than one year. Other studies that assessed incidence of depression after a diagnosis of diabetes calculated incidence of depression after a period of at least two years [8-10] and may have missed the increased incidence shortly after the diagnosis of diabetes that we observed. The use of AD and BD after initiation of diabetes can be a consequence of the burden of diabetes, as was noticed in previous papers [18-20]. However, it is also possible that the increased incidence of AD and BD use reflects selective detection of the general practitioner: in patients who consulted their doctor for diabetes-related complaints, depressive or psychosocial symptoms might have been detected 'by coincidence'. The short time period between the initiation of diabetes treatment and incidence of AD and BD use makes it more likely that selective detection is an explanation for our results.

This study also showed an increased incidence of AD and BD use shortly *before* the initiation of diabetes medication. This suggests that depression and psychosocial complaints might be risk factors for diabetes. Conversely, the antidepressants themselves might be a risk factor for diabetes. There are indications in the literature that antidepressant use can disturb glucose levels and glycemic control [21]. However, the short time period between the increased incidence of AD and BD use and the initiation of diabetes treatment, makes it unlikely that they are causal risk factors for diabetes. A more plausible explanation for our findings is that diabetes was more likely to be detected in patients that were treated with antidepressants and benzodiazepines. In contrast to other studies investigating depression as a risk factor for diabetes, we measured incident rather than prevalent depression. This might be important as subjects with incident depression will be more comparable regarding covariates to subjects without depression than to subjects with prevalent depression. Selecting prognostically more comparable groups is better than adjusting for differences in the analyses because the latter method increases the chance of residual confounding. Also in contrast with other studies, we measured depression continuously over time. This

continuous measure might better represent depression than assessing depression at one point in time. These two aspects might be a reason that other studies found an increased risk of diabetes among depressed patients after a longer time period than we did.

We studied antidepressant use as well as benzodiazepine use to investigate whether the association with diabetes was specific for depression or whether diabetes was also related to more general psychosocial complaints. We observed a similar pattern of increased incidence rates for AD use and BD use, which suggests that the association with diabetes is not specific for depression and that diabetes is also associated with more general psychosocial complaints.

Strengths of this study are that we had complete prescription history of almost 50,000 diabetic patients and 150,000 non-diabetic patients. The design of our study allowed us to study the time-specific association between depression and incidence of diabetes and between diabetes and incidence of depression within one study population.

In summary, this study showed an increased incidence of AD and BD use shortly before and shortly after the initiation of diabetes medication. The short time period between incidence of AD and BD use and initiation of diabetes treatment makes it unlikely that depression or psychosocial complaints are causal risk factors for diabetes. Also, the short time period between initiation of diabetes treatment and incidence of AD and BD use and makes it unlikely that diabetes is a causal risk factor for depression or psychosocial complaints. A more plausible explanation is that depression, as well as more general psychosocial complaints, are more likely to be detected and treated if patients present with diabetes-related complaints to a physician, or vice versa, that diabetes is more likely to be detected if patients present with depression-related or psychosocial complaints.

References

1. Ali S, Stone MA, Peters JL, Davies MJ, Khunti K. The Prevalence of Co-Morbid Depression in Adults With Type 2 Diabetes: a Systematic Review and Meta-Analysis. *Diabet Med* 2006;23:1165-73.
2. Knol MJ, Twisk JW, Beekman AT, Heine RJ, Snoek FJ, Pouwer F. Depression As a Risk Factor for the Onset of Type 2 Diabetes Mellitus. A Meta-Analysis. *Diabetologia* 2006;49:837-45.
3. Carnethon MR, Biggs ML, Barzilay JI, Smith NL, Vaccarino V, Bertoni AG et al. Longitudinal Association Between Depressive Symptoms and Incident Type 2 Diabetes Mellitus in Older Adults: the Cardiovascular Health Study. *Arch Intern Med* 2007;167:802-7.
4. Engum A. The Role of Depression and Anxiety in Onset of Diabetes in a Large Population-Based Study. *J Psychosom Res* 2007;62:31-8.
5. Brown LC, Majumdar SR, Newman SC, Johnson JA. History of Depression Increases Risk of Type 2 Diabetes in Younger Adults. *Diabetes Care* 2005;28:1063-7.
6. Knol MJ, Geerlings MI, Egberts AC, Gorter KJ, Grobbee DE, Heerdink ER. No Increased Incidence of

Diabetes in Antidepressant Users. *Int Clin Psychopharmacol* 2007;22:382-6.

7. Palinkas LA, Lee PP, Barrett-Connor E. A Prospective Study of Type 2 Diabetes and Depressive Symptoms in the Elderly: the Rancho Bernardo Study. *Diabet Med* 2004;21:1185-91.
8. Bisschop MI, Kriegsman DM, Deeg DJ, Beekman AT, van Tilburg W. The Longitudinal Relation Between Chronic Diseases and Depression in Older Persons in the Community: the Longitudinal Aging Study Amsterdam. *J Clin Epidemiol* 2004;57:187-94.
9. Brown LC, Majumdar SR, Newman SC, Johnson JA. Type 2 Diabetes Does Not Increase Risk of Depression. *CMAJ* 2006;175:42-6.
10. Polsky D, Doshi JA, Marcus S, Oslin D, Rothbard A, Thomas N et al. Long-Term Risk for Depressive Symptoms After a Medical Diagnosis. *Arch Intern Med* 2005;165:1260-6.
11. Maraldi C, Volpato S, Penninx BW, Yaffe K, Simonsick EM, Strotmeyer ES et al. Diabetes Mellitus, Glycemic Control, and Incident Depressive Symptoms Among 70- to 79-Year-Old Persons: the Health, Aging, and Body Composition Study. *Arch Intern Med* 2007;167:1137-44.
12. Akker van den M, Schuurman A, Metsemakers J, Buntinx F. Is Depression Related to Subsequent Diabetes Mellitus? *Acta Psychiatr Scand* 2004;110:178-83.
13. Herings RM, Bakker A, Stricker BH, Nap G. Pharmaco-Morbidity Linkage: a Feasibility Study Comparing Morbidity in Two Pharmacy Based Exposure Cohorts. *J Epidemiol Community Health* 1992;46:136-40.
14. Von Korff M, Wagner EH, Saunders K. A Chronic Disease Score From Automated Pharmacy Data. *J Clin Epidemiol* 1992;45:197-203.
15. Gardarsdottir H, Heerdink ER, van DL, Egberts AC. Indications for Antidepressant Drug Prescribing in General Practice in the Netherlands. *J Affect Disord* 2007;98:109-15.
16. Thompson C, Ostler K, Peveler RC, Baker N, Kinmonth AL. Dimensional Perspective on the Recognition of Depressive Symptoms in Primary Care: The Hampshire Depression Project 3. *Br J Psychiatry* 2001;179:317-23.
17. Olfson M and Pincus HA. Use of Benzodiazepines in the Community. *Arch Intern Med* 1994;154:1235-40.
18. Palinkas LA, Barrett-Connor E, Wingard DL. Type 2 Diabetes and Depressive Symptoms in Older Adults: a Population-Based Study. *Diabet Med* 1991;8:532-9.
19. Talbot F and Nouwen A. A Review of the Relationship Between Depression and Diabetes in Adults: Is There a Link? *Diabetes Care* 2000;23:1556-62.
20. Knol MJ, Heerdink ER, Egberts AC, Geerlings MI, Gorter KJ, Numans ME et al. Depressive Symptoms in Subjects With Diagnosed and Undiagnosed Type 2 Diabetes. *Psychosom Med* 2007;69:300-5.
21. McIntyre RS, Soczynska JK, Konarski JZ, Kennedy SH. The Effect of Antidepressants on Glucose Homeostasis and Insulin Sensitivity: Synthesis and Mechanisms. *Expert Opin Drug Saf* 2006;5:157-68.

Chapter 3

Methodological issues in etiologic research

Chapter 3.1

Unpredictable bias when using the missing indicator method or complete case analysis for missing confounder values: an empirical example

Abstract

Background

In etiologic research, data on one or more confounders are often missing. The missing indicator method (MIM) and complete case analysis (CC) are frequently used to handle missing confounder data, despite evidence that these methods give biased results. Using empirical data, we demonstrated the degree of bias in the effect estimate of exposure when using MIM and CC for missing confounder data and compared this with multiple imputation (MI).

Methods

From a prospective cohort study (PREDICT-NL), we selected an exposure (marital status), an outcome (depressive disorder), and three confounder variables (age, sex and income). We created missing values in 'income' according to different patterns of missingness: missing values were created completely at random and depending upon the exposure and outcome values. Percentages of missing values were varied from 2.5% to 30%.

Results

When missing values were created completely at random, MIM gave an overestimation of the odds ratio while CC and MI gave unbiased odds ratios. MIM and CC gave under- or overestimations of the odds ratio when missing values depended on the exposure and outcome values. The magnitude and direction of the bias depended on how the missing values were related to exposure and outcome. Bias increased with an increasing percentage of missing values.

Conclusions

MIM should not be used in handling missing confounder data, because it gives an unpredictable bias of the effect estimate of the association even with small percentages of missing values. CC can be used when missing values are missing completely at random, but gives loss of statistical power. MI gives unbiased effect estimates when missing values are completely at random or when they depend on observed data, which is usually the case in medical research. MI should therefore be used in most situations with missing confounder data.

Introduction

The aim of etiologic research is to answer the question whether a certain factor (i.e. determinant, exposure) is causally related to the occurrence of an outcome (i.e. disease). To establish causality, the effect of external factors that could disturb the association between exposure and outcome, confounders, should be ruled out. Ideally, this is done by randomising subjects to exposure. Randomized trials provide a powerful approach to determine causal effects. However, because randomisation is often impossible or unethical, observational studies are performed in which it is essential to adjust for confounders.

In observational etiologic research missing data on one or more confounders can affect the possibility to adequately adjust for confounding variables. There are many methods to handle missing data [1,2]. Commonly, researchers exclude subjects with missing data from the analysis, so-called complete case analysis, as multivariable modelling in standard software packages commonly excludes persons with a missing value on any of the variables in the model. Obviously, this affects the number of subjects and thereby the statistical power, but more importantly, it may lead to seriously biased estimates [1,3-5]. This bias occurs because missing values are typically related to other observed subject characteristics including the outcome. Even in a follow up study where the outcome is not yet known at baseline, missing values can be (indirectly) related to the outcome if predisposing factors are associated with the missing covariate and with the outcome. For example, a question on medication use might be skipped by someone with low education level while education level is also associated with mortality.

Another approach to handle missing data is the 'missing indicator method', which was specifically proposed for missing confounder data in etiologic research [6,7]. This method does not exclude subjects from the analysis but adds an extra variable to the statistical model to indicate that the value of a certain variable is missing. Although it has been argued that the missing indicator method gives biased results [1,3,5,8], it is still used quite often, notably in etiologic research. A reason for this might be that it is an intuitively appealing method because it seems to adjust for missing values and it is easy to use. A more sophisticated approach to handle missing data is to impute (i.e. fill in) missing values. Imputing the overall or subgroup mean commonly yields biased effect estimates as well [2]. Imputing a missing value by a value predicted by a regression model using all other observed variables in the data set, including the outcome, seems a better approach [1,2,4,8-10]. This imputation can be done once (single imputation) or multiple times (multiple imputation). Single imputation gives unbiased effect estimates but overestimates the precision, i.e. underestimates the standard error, of the estimate as it assumes that all data is present [1,8,10]. In multiple

imputation the missing value is imputed multiple times (usually 5 to 10) and the uncertainty of the imputed values is taken into account. Multiple imputation has been described in numerous papers and it has been shown to give valid estimates and valid standard errors [1,4,8-15].

Although multiple imputation is increasingly being used (for example [13,16,17]), complete case analysis and the missing indicator method are still very common in the epidemiologic literature [18]. Many researchers, and editors alike, appear not to be aware of the degree of bias that can result from both methods. The objective of this study was to show the direction and degree of bias when using the missing indicator method or complete case analysis to handle missing confounder data in an etiological context. In contrast to earlier studies we used an empirical data set and simulated different patterns of missing data in a confounder. We studied the bias in the effect estimate of the association between exposure and outcome and varied the percentage of missing values from 2.5% to 30%. We compared the missing indicator method and complete case analysis with multiple imputation.

Methods

Data set

We used data from the PREDICT study which is described in detail elsewhere [19]. In short, the PREDICT study is a European prospective cohort study aimed to develop a multifactor risk algorithm for onset of major depression over 12 months. A total of 1338 subjects were included in the Dutch part of the study (PREDICT-NL). The outcome of interest was the occurrence of major depressive disorder according to DSM-IV criteria using the Depression Section of the Composite International Diagnostic Interview (CIDI) [20,21]. Many potential risk factors for depression were measured. For the present analysis we selected one particular exposure variable, one outcome variable, and three confounder variables, and created missing values in one of the confounders.

As exposure we used ‘marital status’, where married, including living together and enduring relationships, was coded as 0 and single, including divorced and widowed, was coded as 1. The outcome major depressive disorder was coded 0 when absent, and 1 when present. Age, sex and income were selected as potential confounders. Age was used as a continuous variable; females were coded as 0 and males as 1. Income was defined as the net yearly household income and was categorized into four levels: less than 12,000 euro was coded as 0, 12,000 to 22,800 as 1, 22,800 to 50,000 as 2, and more than 50,000 as 3.

From the original PREDICT-NL data set (n=1338) we selected all subjects without missing values on the five variables (n=1075) to start the present study with a complete data set.

This complete data set, further referred to as 'study data set', served as our reference situation. Descriptive statistics and frequencies of the five selected variables are presented in Table 1.The overall frequency of depressive disorders was 14%. The crude odds ratio between marital status and depressive disorder was 2.0 (95% CI: 1.4-2.9), indicating that being single doubled the risk of major depressive disorder in comparison with being married. Adjustment for age and sex did not change the odds ratio much (1.8; 95% CI: 1.3-2.6). Additional adjustment for income decreased the odds ratio to 1.4 (95% CI: 0.9-2.2), showing that income confounded the association. The odds ratio adjusted for age, sex and income was defined as the 'true' odds ratio to which all later estimates were compared. We emphasize that the data are here used for illustration purposes only and not to estimate the true causal association between marital status and depressive disorder.

Table 1 Distribution of age, sex and income (confounder variables) and major depressive disorder (outcome variable) in the study data set

	Marital status	
	Married	Single
	N=801	N=274
Age, mean (SD)	50.4 (15)	48.1 (20)
Male sex	43.8%	29.9%
Income		
< 12,000	5.4%	33.2%
12,000-22,800	23.0%	41.2%
22,800-50,000	53.8%	22.6%
> 50,000	17.9%	2.9%
Depressive disorder	11.6%	20.8%

Missing values

From the study data set 1000 new data sets with equal size as the study data set (n=1075) were created by simulating the outcome variable and keeping the other four variables unchanged. In each of the 1000 data sets, simulation of the outcome was done by first calculating the probability of depressive disorder (P) for each patient using the 'true' regression coefficients and the patient values. Subsequently, for each patient a random number from a uniform distribution in the interval [0,1] was sampled. An outcome status of 1 (depressive disorder present) was assigned when the random number was smaller than P and an outcome status of 0 otherwise. Accordingly, on average 14% of the subjects (per data

set) had a depressive disorder. By simulating 1000 new data sets, instead of copying the study data set 1000 times, we mimicked taking samples from a population rather than using just one sample.

In each simulated data set missing values were created in the confounder variable 'income' using four different scenarios (Table 2). In scenario 1 we created missing values according to the Missing Completely At Random (MCAR) mechanism [8]. In randomly selected subjects the confounder 'income' was set to missing irrespective of the values of the other variables. In scenario 2, 3 and 4 (Table 2) we created missing values according to the Missing At Random (MAR) mechanism [8]. Missing data are denoted as MAR when missing data occur in relation to other but observed patient characteristics. We created missing values in the confounder 'income' in relation to the exposure value as well as in relation to the outcome value. It may seem odd to create missing values in relation to the outcome but, as mentioned above, missing values in a confounder are often indirectly related to the outcome. We distinguished four subgroups of exposure and outcome in which we created missing values in the confounder 'income': 1) married and no depressive disorder; 2) married and depressive disorder; 3) single and no depressive disorder; 4) single and depressive disorder. To define the distribution of missing values of 'income' over the four categories we used odds ratios of missingness (Table 2). We defined three different sets of odds ratios. In scenario 2 (Table 2) the set of odds ratios was 3-2-3-1, representing the distribution of missing values of the variable 'income' in the original PREDICT-NL data set. In scenario 3 (Table 2) we used a set of odds ratios of 5-1-1-5, where missing values were particularly created in the married non-depressed subjects (exposure = 0, outcome = 0) and the single depressed subjects (exposure = 1, outcome = 1). In scenario 4, (Table 2) a set of odds ratios of 1-5-5-1 was used to create missing values mainly in the married depressed subjects (exposure = 0, outcome = 1) and in the single non-depressed subjects (exposure = 1, outcome = 0). In scenario 3 and 4, we created a large contrast in the percentage of missing income values between the categories that determine that being *single* is a risk factor for depression and the categories that determine that being *married* is a risk factor for depression. By creating this contrast the possible bias of the methods to handle missing data becomes most apparent.

Five different percentages of missing values in the confounder 'income' were simulated, namely 2.5%, 5%, 10%, 20% and 30%. These percentages are not uncommon in etiologic research.

We did not create missing values according to the Missing Not At Random (MNAR) mechanism as there is no general method to handle these missing values [4,8,11,22]. MNAR

Table 2 Characteristics of the four different scenarios of creating missing values in the confounder 'income'

	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 4**
Mechanism	Missing Completely At Random	Missing At Random	Missing At Random	Missing At Random
Missing values	Created in randomly selected subjects	Associated with marital status (exposure) and major depressive disorder (outcome)		
Distribution of missing values	Not applicable	Depressive disorder Marital status / No (0) / Yes (1) Married (0) / 3 / 2 Single (1) / 3 / 1	Depressive disorder Marital status / No (0) / Yes (1) Married (0) / 5 / 1 Single (1) / 1 / 5	Depressive disorder Marital status / No (0) / Yes (1) Married (0) / 1 / 5 Single (1) / 5 / 1
Explanation distribution of missing values	Not applicable	Realistic distribution of missing values in 'income' as encountered in the original PREDICT-NL data set	Missing values were mainly created in married non-depressed subjects and single depressed subjects	Missing values were mainly created in married depressed subjects and single non-depressed subjects

implies that the probability that an observation is missing depends on unobserved characteristics, like the value of the observation itself.

Methods to handle missing data

We compared three methods to handle missing data: the missing indicator method (MIM), complete case analysis (CC), and multiple imputation (MI). With MIM the missing values in 'income' were recoded as 0 and an extra indicator variable was created coded as 1 if the value on 'income' was missing and 0 otherwise. Both the recoded 'income' variable and indicator variable were included in the logistic regression model. Note that this is the same as adding an extra category for missing values to the categorical 'income' variable.

In CC all subjects with missing values were excluded from the logistic regression analysis. In MI all observed data was used to estimate the missing values with regression analysis using Multivariate Imputation by Chained Equations (MICE) [23]. This estimation was performed five times to get variation in the imputed values. Then, conventional logistic regression analysis was performed in each of the five data sets, and the results were pooled in a way that reflects the extra variability due to uncertainty of the imputed values [24].

Bias and coverage

For all situations (i.e. four different scenarios of creating missing values, five different missing value percentages, and three different methods to handle these missing values) we calculated the mean odds ratio over the 1000 simulated data sets for the association between marital status and depressive disorder adjusted for age, sex and income. These mean odds ratios were compared with the true odds ratio estimated from the study data set which was 1.4 (95% CI: 0.9-2.2). We assessed the coverage of the 95% confidence interval of the odds ratio in each situation. To this aim we calculated the 95% confidence interval around the estimated odds ratio in each simulated data set and assessed whether the true odds ratio was included. The coverage was then calculated as the percentage of 95% confidence intervals over the 1000 simulations that indeed included the true odds ratio. Coverage of 0.95 represents correct coverage.

All analyses were performed with R2.4.1 [25].

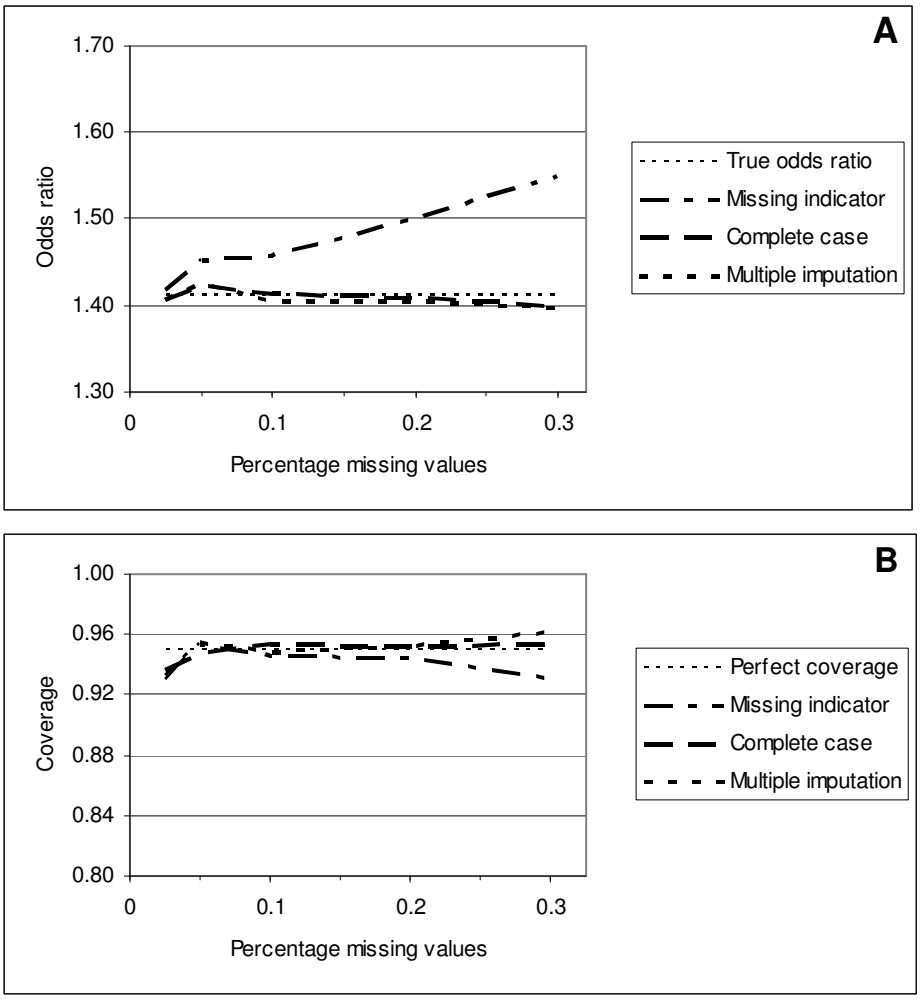
Results

Scenario 1 – Missing Completely At Random

Using the missing indicator method (MIM) resulted in an overestimation of the odds ratio of marital status (Figure 1A). This bias increased with an increasing percentage of missing

values. Both complete case analysis (CC) and multiple imputation (MI) gave an unbiased effect estimate of the association between marital status and depressive disorder. The coverage of the 95% confidence interval was around 0.95 for all methods of handling missing data (Figure 1B).

Figure 1 Scenario 1 - Odds ratio (A) of marital status and coverage of 95% confidence interval (B) for different methods of handling missing data when missing values in income were created according to the Missing Complete At Random (MCAR) mechanism, compared with the true odds ratio (1.4) and correct coverage (0.95)

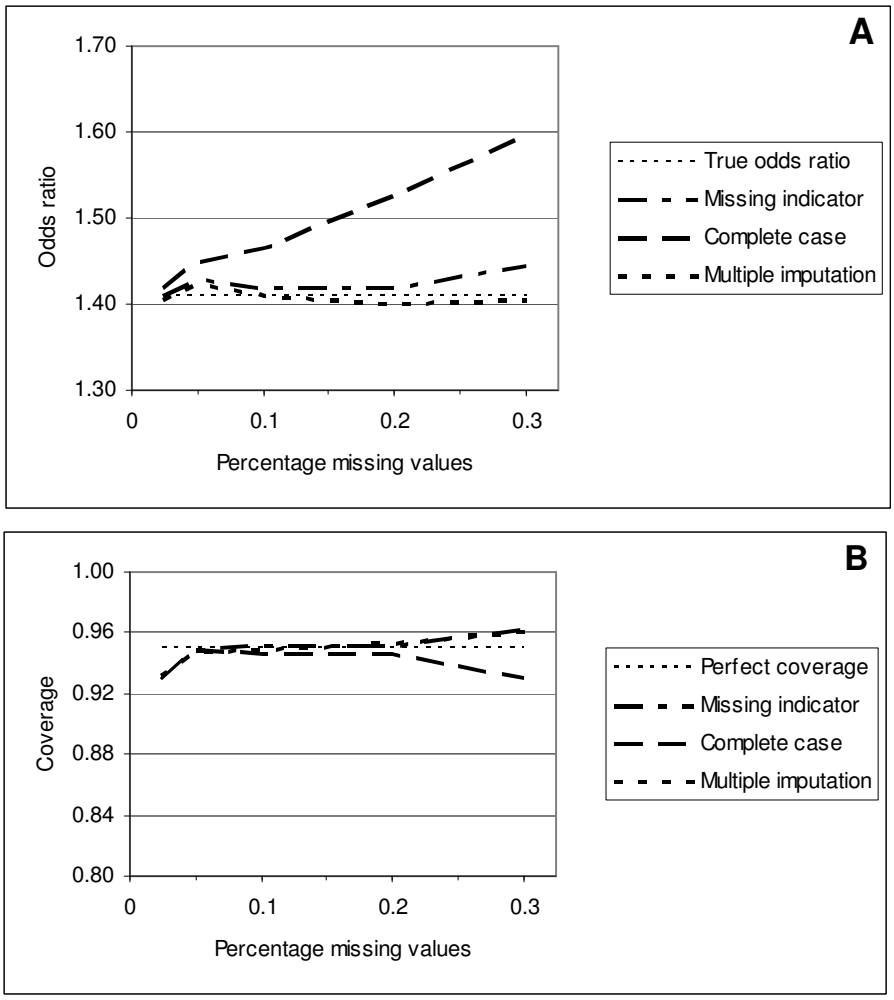


Scenario 2 – Missing At Random, odds ratios of 3-2-3-1

Using MIM resulted in a small overestimation of the odds ratio with 30% of missing values (Figure 2A). CC gave an overestimation of the odds ratio, with more bias in higher percentages of missing values up to an odds ratio of 1.6 (95% CI: 0.98-2.6) in case of 30%

missing values. The estimates of the odds ratio after MI were close to the true odds ratio for all percentages of missing values. MIM, CC and MI all showed coverage of about 0.95 (Figure 2B).

Figure 2 Scenario 2 - Odds ratio (A) of marital status and coverage of 95% confidence interval (B) for different methods of handling missing data when missing values in income were realistically created according to the Missing At Random (MAR) mechanism with odds ratios for missing values of 3-2-3-1, compared with the true odds ratio (1.4) and correct coverage (0.95)

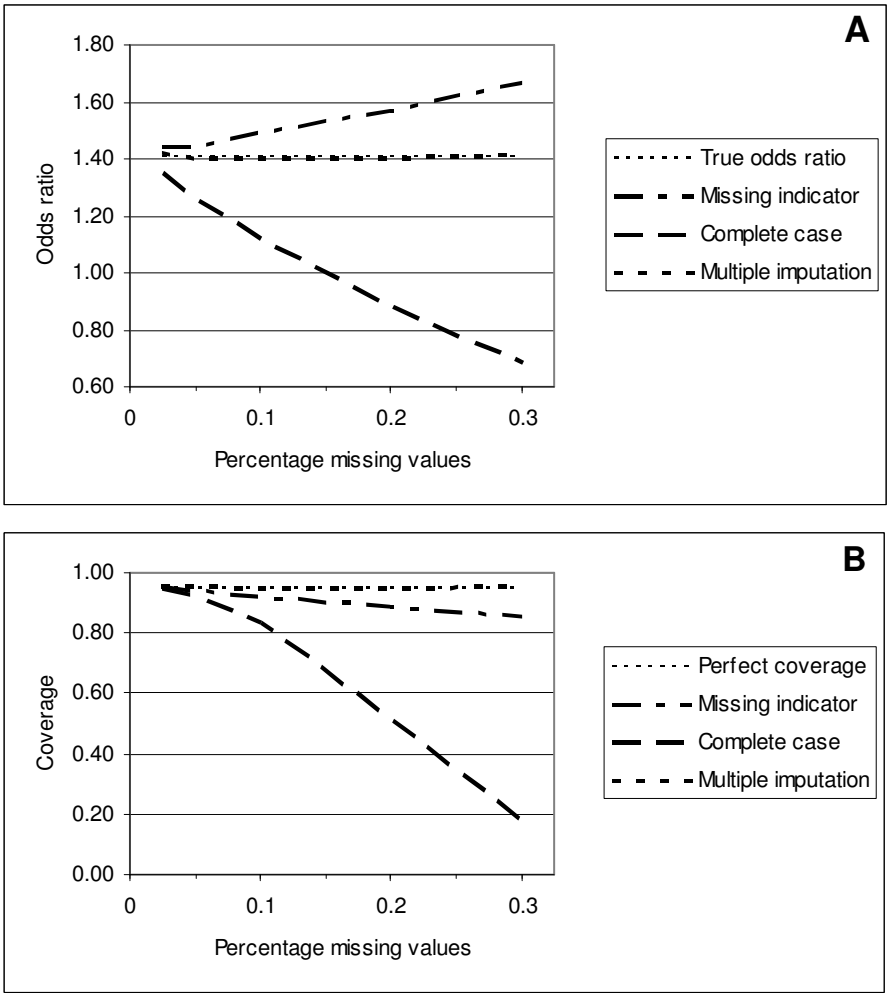


Scenario 3 – Missing At Random, odds ratios of 5-1-1-5

There was an overestimation of the odds ratio using MIM and a large underestimation of the odds ratio using CC (Figure 3A). With both methods the bias increased with an increasing percentage of missing data up to an odds ratio of 1.7 (95% CI: 1.1-2.6) for MIM and 0.68 (95% CI: 0.41-1.1) for CC with 30% missing values. MI resulted in an unbiased odds ratio for

all percentages of missing values. Coverage was lower than 0.95 in MIM and in CC, with coverage of even 0.85 and 0.17 respectively for 30% missing values (Figure 3B), indicating that the true odds ratio was not included in the 95% confidence interval in 15% and 83% of the simulations, while this should be only 5%. MI gave good coverage for all percentages of missing values.

Figure 3 Scenario 3 - Odds ratio (A) of marital status and coverage of 95% confidence interval (B) for different methods of handling missing data when missing values in income were created according to the Missing At Random (MAR) mechanism with odds ratios for missing values of 5-1-1-5, compared with the true odds ratio (1.4) and correct coverage (0.95)

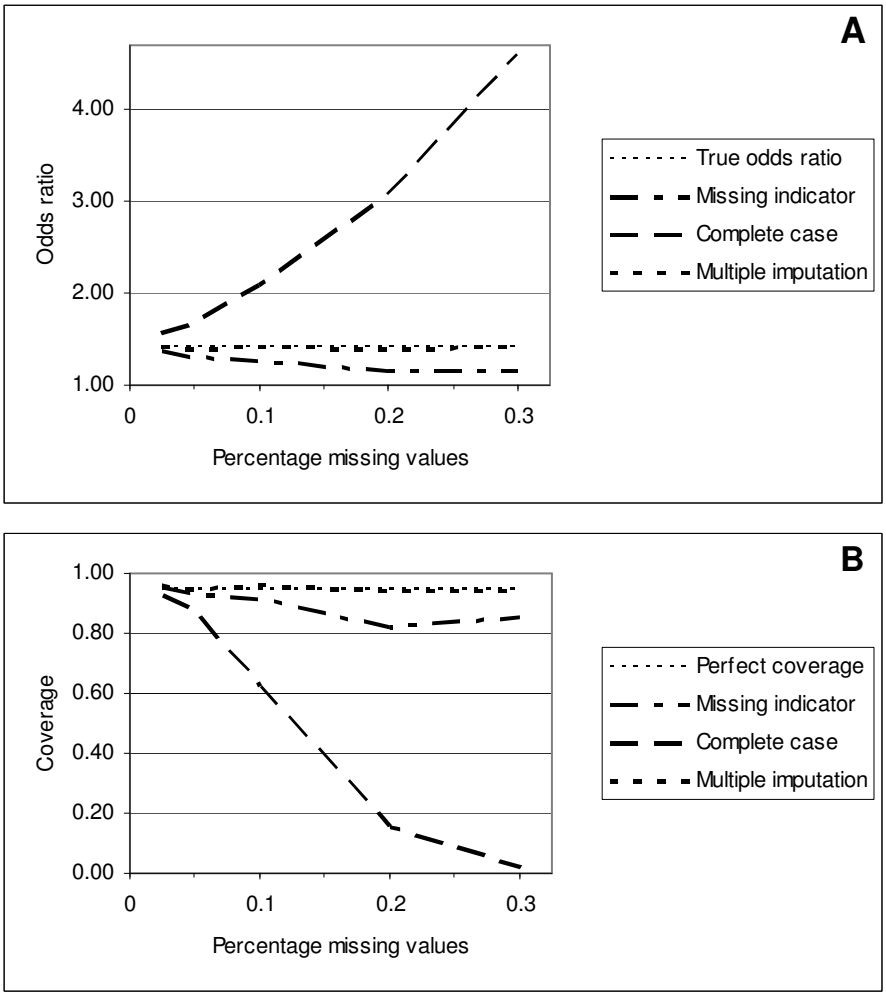


Scenario 4 – Missing At Random, odds ratios of 1-5-5-1

The odds ratio was underestimated when using MIM and largely overestimated when using CC, and showed increasing bias with an increasing percentage of missing values (Figure 4A).

MI gave unbiased odds ratios. Coverage was lower than 0.95 using MIM and was very low using CC, while MI gave good coverage (Figure 4B).

Figure 4 Scenario 4 - Odds ratio (A) of marital status and coverage of 95% confidence interval (B) for different methods of handling missing data when missing values in income were created according to the Missing At Random (MAR) mechanism with odds ratios for missing values of 1-5-5-1, compared with the true odds ratio (1.4) and correct coverage (0.95)



Discussion

This study demonstrates the degree of bias in the effect estimate of exposure when using the missing indicator method (MIM) and complete case analysis (CC) for missing confounder data in comparison with multiple imputation (MI). MIM and CC gave a biased odds ratio in almost all situations of missing confounder data. The direction and degree of that bias depended on how the missing values were related to exposure and outcome. The bias was

already present with a small percentage of missing values and increased when the percentage of missing values increased. Coverage of the 95% confidence interval when MIM and CC were used was often too low, meaning that the confidence interval was too liberal which led to a type I error. This implies that using MIM and CC to handle missing confounder data would lead to a statistical significant odds ratio of the exposure at interest while in reality there is no association. Multiple imputation gave unbiased odds ratios and good coverage in all scenarios, up to 30% of missing confounder values.

Why do MIM and CC result in an over- or underestimation of the odds ratio of marital status? If missing values are missing completely at random, MIM leads to less variation in the confounder variable because all missing values are set to 0, which results in incomplete adjustment for the confounder and therefore in an overestimation of the exposure effect. Although uncommon in practice, when missing values are MCAR the direction of bias when using MIM is indeed predictable. Obviously, the degree of bias when using MIM for missing confounder data depends on the strength of that confounder.

If missing values depend on other variables in the data set (i.e. MAR), the bias when using MIM and CC analysis is not predictable. The degree and direction of the bias depends on the distribution of missing values over the exposure-outcome categories. In scenario 2 the missing values were quite similarly distributed over the four categories, which resulted in minimal bias when using MIM, and overestimation when using CC. When the missing values were mainly created in the categories that determine that being *single* (marital status = 1) is a risk factor for depressive disorder (Scenario 3), MIM gave an overestimation and CC an underestimation of the odds ratio of marital status. Having relatively more missing values in these categories resulted in incomplete adjustment for income and thus overestimation of the odds ratio for marital status when using MIM. CC resulted in this scenario in an underestimation because the number of analyzable subjects that determine the association between being *single* and depressive disorder decreased. Creating the missing values mainly in the categories that determine that being *married* is a risk factor for depressive disorder, MIM resulted in an underestimation and CC in an overestimation (Scenario 4). Creating missing values particularly in these categories, MIM again resulted in incomplete adjustment for income but now yielded an underestimation of the effect as being married was inversely associated (i.e. odds ratio < 1) with depressive disorder. CC resulted in this scenario in an overestimation because there were fewer subjects contributing to the association between being married and depressive disorder.

Hence, in scenarios with a large contrast in number of missing values between the categories that determine that being *single* is a risk factor for depression and the categories

that determine that being *married* is a risk factor for depression, such as we created in scenario 3 and 4, it seemed possible to predict the degree and direction of bias. However, if the contrast is not so clear, such as in scenario 2, the direction of bias of using CC and MIM is less predictable. In addition, commonly the missing value mechanism is not explicitly known.

What should researchers do with missing confounder data? If missing values depend on observed data (i.e. MAR), which is usually the case in medical research [1,4,8,10,11], but the missing value mechanism is not exactly known MIM and CC give an unpredictable degree and direction of the bias and should not be used. MIM might be used safely if missing values are missing completely at random, the percentage of missing values is low (<5%) and the confounder is weak. However, the situation that all three criteria are met, is unlikely to occur. In addition, in this situation CC performs equally well as MIM, and is easier to perform. MI performs well when missing values are MCAR or MAR, even up to 30% of missing values, and is therefore the method of choice. Researchers might be reluctant to apply MI because it is believed to be time-consuming, difficult to perform and a self-fulfilling prophecy. However, an increasing number of studies appear on the application of MI and software has much improved [26], making the method better accessible. An alternative, which is still better than MIM and CC analysis [8,9], might be to perform single imputation. An advantage of single imputation is that it is easier to understand and that user-friendly statistical packages can be used. A disadvantage is that it still produces too precise estimates (too small standard errors) which increases the chance of a type I error [1].

In this study we made certain choices in creating missing values. Missing data patterns might be more complex in practice and therefore the degree and direction of bias might be even more unpredictable. First, we let the missing values depend on the exposure and outcome only, while missing values can also be related to more variables. Second, we created missing values in one confounder variable, while data can be missing in more than one confounder or in the exposure or outcome variable. In addition, a low percentage of missing data in several variables will easily result in a large percentage of subjects with at least one missing value.

In conclusion, we showed that one should not use the missing indicator method to handle missing confounder data, because it gives a biased estimation of the effect estimate of the association even with small percentages of missing values. More importantly, the direction of the bias is unpredictable. Complete case analysis can be used when missing values are missing completely at random. This, however, is hardly ever the case. Moreover, complete case analysis always leads to loss of statistical power. Multiple imputation gives unbiased effect estimates when missing values are missing completely at random and when

they depend on observed data. The latter is often the case in medical research. Hence, multiple imputation can be used in most situations with missing confounder data.

References

1. Greenland S and Finkle WD. A Critical Look at Methods for Handling Missing Covariates in Epidemiologic Regression Analyses. *Am J Epidemiol* 1995;142:1255-64.
2. Little RJ. Regression With Missing X's: a Review. *J Am Stat Assoc* 1992;87:1227-37.
3. Jones MP. Indicator and Stratification Methods for Missing Explanatory Variables in Multiple Linear Regression. *J Am Stat Assoc* 1996;91:222-30.
4. Schafer JL and Graham JW. Missing Data: Our View of the State of the Art. *Psychol Methods* 2002;7:147-77.
5. Vach W and Blettner M. Biased Estimation of the Odds Ratio in Case-Control Studies Due to the Use of Ad Hoc Methods of Correcting for Missing Values for Confounding Variables. *Am J Epidemiol* 1991;134:895-907.
6. Anderson AB, Basilevsky A and Hum DPJ. *Missing Data: A Review of the Literature.* In *Handbook of Survey Research*. (Ed. Rossi PH, Wright JD and Anderson A) pp. 415-492, Academic Press, New York, 1983.
7. Miettinen OS. *Regression Analysis.* In *Theoretical Epidemiology: Principles of Occurrence Research*. pp. 216-243, John Wiley & Sons, New York, 1985.
8. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a Gentle Introduction to Imputation of Missing Values. *J Clin Epidemiol* 2006;59:1087-91.
9. Moons KG, Donders RA, Stijnen T, Harrell FE, Jr. Using the Outcome for Imputation of Missing Predictor Values Was Preferred. *J Clin Epidemiol* 2006;59:1092-101.
10. Rubin DB. Multiple Imputation After 18+ Years. *J Am Stat Assoc* 1996;91:473-89.
11. Rubin DB and Schenker N. Multiple Imputation in Health-Care Databases: an Overview and Some Applications. *Stat Med* 1991;10:585-98.
12. van Buuren S, Boshuizen HC, Knook DL. Multiple Imputation of Missing Blood Pressure Covariates in Survival Analysis. *Stat Med* 1999;18:681-94.
13. Arnold AM and Kronmal RA. Multiple Imputation of Baseline Data in the Cardiovascular Health Study. *Am J Epidemiol* 2003;157:74-84.
14. van der Heijden GJ, Donders AR, Stijnen T, Moons KG. Imputation of Missing Values Is Superior to Complete Case Analysis and the Missing-Indicator Method in Multivariable Diagnostic Research: a Clinical Example. *J Clin Epidemiol* 2006;59:1102-9.
15. Crawford SL, Tennstedt SL, McKinlay JB. A Comparison of Analytic Methods for Non-Random Missingness of Outcome Data. *J Clin Epidemiol* 1995;48:209-19.
16. Goldstein AM, Martinez M, Tucker MA, Demenais F. Gene-Covariate Interaction Between Dysplastic Nevi and the CDKN2A Gene in American Melanoma-Prone Families. *Cancer Epidemiol Biomarkers Prev* 2000;9:889-94.
17. Knol MJ, Heerdink ER, Egberts AC, Geerlings MI, Gorter KJ, Numans ME et al. Depressive Symptoms in Subjects With Diagnosed and Undiagnosed Type 2 Diabetes. *Psychosom Med*

2007;69:300-5.

18. Burton A and Altman DG. Missing Covariate Data Within Cancer Prognostic Studies: a Review of Current Reporting and Proposed Guidelines. *Br J Cancer* 2004;91:4-8.
19. King M, Weich S, Torres F, Svab I, Maaroos H, Neeleman J et al. Prediction of Depression in European General Practice Attendees: the PREDICT Study. *BMC Public Health* 2006;6:6.
20. World Health Organisation: Composite International Diagnostic Interview (CIDI). Version 2.1, WHO, Geneva, 1997.
21. Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J et al. The Composite International Diagnostic Interview. An Epidemiologic Instrument Suitable for Use in Conjunction With Different Diagnostic Systems and in Different Cultures. *Arch Gen Psychiatry* 1988;45:1069-77.
22. Little RJ and Rubin DB. *Statistical Analysis with Missing Data.* John Wiley & Sons, Hoboken, New Jersey, 1987.
23. van Buuren S and Oudshoorn K. *Flexible Multivariate Imputation by MICE.* TNO Prevention and Health, Leiden, 1999.
24. Rubin DB. *Multiple Imputation for Nonresponse in Surveys.* John Wiley & Sons, Hoboken, New Jersey, 1987.
25. R Development Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL http://www.R-project.org, 2006.
26. Horton NJ and Kleinman KP. Much Ado About Nothing: A Comparison of Missing Data Methods and Software to Fit Incomplete Data Regression Models. *Am Stat* 2007;61:79-90.

Chapter 3.2

Estimating interaction on an additive scale between continuous determinants in a logistic regression model

Abstract

Background

To determine the presence of interaction in epidemiologic research, typically a product term is added to the regression model. In linear regression, the regression coefficient of the product term reflects interaction as departure from additivity. However, in logistic regression it refers to interaction as departure from multiplicativity. Rothman has argued that interaction estimated as departure from additivity better reflects biologic interaction. So far, literature on estimating interaction on an additive scale using logistic regression only focused on dichotomous determinants. The objective of the present study was to provide the methods to estimate interaction between continuous determinants and to illustrate these methods with a clinical example.

Methods

From the existing literature we derived the formulas to quantify interaction as departure from additivity between one continuous and one dichotomous determinant and between two continuous determinants using logistic regression. Bootstrapping was used to calculate the corresponding confidence intervals. To illustrate the theory with an empirical example, data from the Utrecht Health Project were used, with age and body mass index as risk factors for elevated diastolic blood pressure.

Conclusions

The methods and formulas presented in this paper are intended to assist epidemiologists to calculate interaction on an additive scale between two variables on a certain outcome. The proposed methods are included in a spreadsheet which is freely available at: www.juliuscenter.nl/additive-interaction.xls.

Introduction

In epidemiology, interaction refers to the situation where the effect of one risk factor (A) on a certain disease outcome is different across strata of another risk factor (B), or vice versa. This means that if interaction between A and B is present, A and B are not independent in causing a certain disease. If the combined effect of A and B is larger (or smaller) than the *sum* of the individual effects of A and B there is interaction on an additive scale, or departure from additivity. Interaction on a multiplicative scale, or departure from multiplicativity, occurs when the combined effect of A and B is larger (or smaller) than the product of the individual effects.

Rothman discerns two types of interaction: statistical and biologic interaction [1]. Statistical interaction means departure from the underlying form of a statistical model. Because there are various statistical models, statistical interaction does not have a consistent meaning. Most researchers assess interaction by entering a product term into the linear or logistic regression model. However, the interpretation of the regression coefficient of the product term depends on the statistical model. In linear regression analysis the regression coefficient of the product term means departure from additivity, whereas in logistic regression (and in Cox regression) the regression coefficient of the product term estimates departure from multiplicativity (Appendix 1). Biologic interaction means that two causes are both needed to cause disease; the two causes are component causes in the same causal model. Rothman has argued that when biologic interaction is examined, we should focus on interaction as departure from additivity rather than departure from multiplicativity [1]. In etiologic epidemiologic research, we are interested in biologic interaction rather than in statistical interaction. However, by adding a product term to a logistic model, interaction is (unknowingly) estimated as departure from multiplicativity.

Rothman [2] and Hosmer & Lemeshow [3] have shown how interaction as departure from additivity can be quantified in a logistic regression model. They proposed to make one categorical variable with four levels that combines two dichotomous determinants. Assmann *et al.* [4] demonstrated that bootstrapping may give better coverage of the 95% confidence interval of the estimate of interaction than the delta method described by Hosmer & Lemeshow [3]. These studies only focused on interaction between two dichotomous determinants. In epidemiologic research, we are often also interested in the effect of continuous determinants on an outcome and dichotomizing a continuous variable may lead to loss of information [5].

This article illustrates with a clinical example the methods and formulas to estimate interaction, as departure from additivity between continuous determinants, and its

uncertainty on both a dichotomous and a continuous outcome. We will first describe the concept of interaction on an additive scale and show the difference with interaction on a multiplicative scale using a simple 2x2 table. Then we will explain how logistic regression analysis can be used to estimate interaction on an additive scale. There are two reasons why regression analysis rather than simple tabulation might be needed to estimate interaction. First, there may be a need to adjust for confounders. Second, if one or both of the determinants are continuous it is not possible to construct a 2x2 table. In this paper we will focus on one of the three measures for interaction as departure from additivity, namely the Relative Excess Risk due to Interaction (RERI). The other two measures, Attributable Proportion (AP) and Synergy index (S), will be shortly discussed in the application section of the paper.

Example data set

To illustrate the methods and formulas, we will use data from the Utrecht Health Project (UHP), which is described in detail elsewhere [6]. In brief, the UHP is an ongoing longitudinal study, which started in the year 2001, among all inhabitants of a new residential area in the city of Utrecht, The Netherlands. At baseline, an Individual Health Profile is made, which is based on a questionnaire, physical examination and blood measurements. By January 2005, 13,128 subjects were invited of whom 6755 gave informed consent (51.4%) and entry data were complete on 6304 (48.0%) adults and children. The adult UHP population consists of 2221 (44.9%) males and 2729 (55.1%) females with a mean age (SD) of 39.3 (12.5) years.

The data set we use in this paper comprises two continuous determinants, age and body mass index (BMI), and one continuous outcome, diastolic blood pressure. Fifty-three subjects had a missing value on BMI or diastolic blood pressure and were excluded from the analyses. Table 1 presents descriptive statistics of the three variables. The three variables were dichotomized because our aim was to show how to calculate additive interaction using both dichotomous and continuous determinants and outcomes. We reasoned that using the same determinants and outcome throughout all the examples would enhance the understanding of the article. As a result, we did not use truly dichotomous determinants, such as sex. Age was dichotomized according to an arbitrarily chosen cut point of 40 years, where age <40 was coded as 0 and age >=40 was coded as 1. BMI was dichotomized according to the overweight cut point of 25 kg/m^2, where BMI <25 was coded as 0 and BMI >=25 as 1. Diastolic blood pressure was dichotomized according to a cut point of 90 mm Hg, where a normal blood pressure was coded as 0 and hypertension as 1. The percentages in the resulting categories were 65.2% and 34.8% for age, 50.9% and 49.1% for BMI and 87.5%

and 12.5% for hypertension.

Table 1 Descriptive statistics of age, body mass index and diastolic blood pressure.

	N	Mean	SD	Median	Min	Max
Age (years)	4897	39.3	12.5	35.6	18.0	91.1
Body mass index (kg/m^2)	4897	25.5	4.2	24.9	9.2	46.7
Diastolic blood pressure (mm Hg)	4897	77.8	10.5	77.0	48.5	126.5

Estimating interaction on an additive scale using a 2x2 table

Consider age (A) and BMI (B) as dichotomous risk factors for diastolic hypertension (D). A 2x2 table can be constructed with the absolute risk of disease in the four following categories: young subjects with normal BMI (A-B-), older subjects with normal BMI (A+B-), young subjects with overweight (A-B+) and older subjects with overweight (A+B+). The risk in category A-B- is called the background risk because in this category the disease frequency is caused by other factors than A and B. Table 2 shows the absolute risks of hypertension in these categories and the risk differences and relative risks in strata of age and BMI.

Table 2 Absolute risks, risk differences and relative risks of diastolic hypertension according to strata of age and BMI.

		Normal BMI (B-)		Overweight (B+)		Risk	Relative
		N	Risk	N	Risk	difference	risk
Young (A-)	**Normal tension (D-)**	1731		1232			
	Hypertension (D+)	79	4.4%	153	11.0%	6.6%	2.50
Old (A+)	**Normal tension (D-)**	581		743			
	Hypertension (D+)	100	14.7%	278	27.2%	12.5%	1.85
Risk difference			10.3%		16.2%		
Relative risk			3.34		2.47		

Interaction on an additive scale is present if the combined effect of A and B is unequal to the sum of the effects of A and B:

$$(R_{A+B+} - R_{A-B-}) \neq (R_{A+B-} - R_{A-B-}) + (R_{A-B+} - R_{A-B-}) \quad (1)$$

where R indicates the absolute risk of disease in that specific stratum. Note that the background risk is subtracted to get the effects of A alone, B alone and A and B combined. In our example: $(27.2 - 4.4) \neq (14.7 - 4.4) + (11.0 - 4.4) \Rightarrow 22.8 > 16.9$, meaning there is 'positive'

interaction on an additive scale because the combined effect is larger than the sum of the individual effects. By dividing all risks in formula (1) by the background risk, R_{A-B-}, an equivalent expression for risk ratios (RRs) is obtained:

$$(RR_{A+B+} - 1) \neq (RR_{A+B-} - 1) + (RR_{A-B+} - 1) \quad (2)$$

In our example: $(27.2/4.4 - 1) \neq (14.7/4.4 - 1) + (11.0/4.4 - 1) \Rightarrow 5.18 > 3.84$.

Interaction on a multiplicative scale is present if the combined effect of A and B is unequal to the product of the effects of A and B:

$$(R_{A+B+} / R_{A-B-}) \neq (R_{A+B-} / R_{A-B-}) * (R_{A-B+} / R_{A-B-}) \quad (3)$$

In our example: $(27.2/4.4) \neq (14.7/4.4) * (11.0/4.4) \Rightarrow 6.18 < 8.35$, meaning that there is 'negative' interaction on a multiplicative scale because the combined effect is smaller than the product of the individual effects. The fact that interaction is present can also be seen in Table 2, as the effect of BMI on hypertension is different across strata of age. Also, the effect of age on hypertension is different across strata of BMI. However, the risk difference is highest in the older age stratum and in the overweight stratum, whereas the RR is highest in the younger age stratum and in the normal BMI stratum. This agrees with the calculations above, as these showed there is positive interaction on an additive scale and negative interaction on a multiplicative scale. This example illustrates that it depends on the measure of effect (risk difference or RR) whether interaction is present or not, or in which direction the interaction operates.

The amount of interaction as departure from additivity can be derived from formula (1) for absolute risks:

$$(R_{A+B+} - R_{A-B-}) - (R_{A+B-} - R_{A-B-}) - (R_{A-B+} - R_{A-B-}) \quad (4)$$

and from formula (2) for relative risks:

$$(RR_{A+B+} - 1) - (RR_{A+B-} - 1) - (RR_{A-B+} - 1) \quad (5)$$

Rothman called this amount of interaction the relative excess risk due to interaction (RERI).[2] Rewriting formula (5) gives:

$$RERI = RR_{A+B+} - RR_{A+B-} - RR_{A-B+} + 1 \quad (6)$$

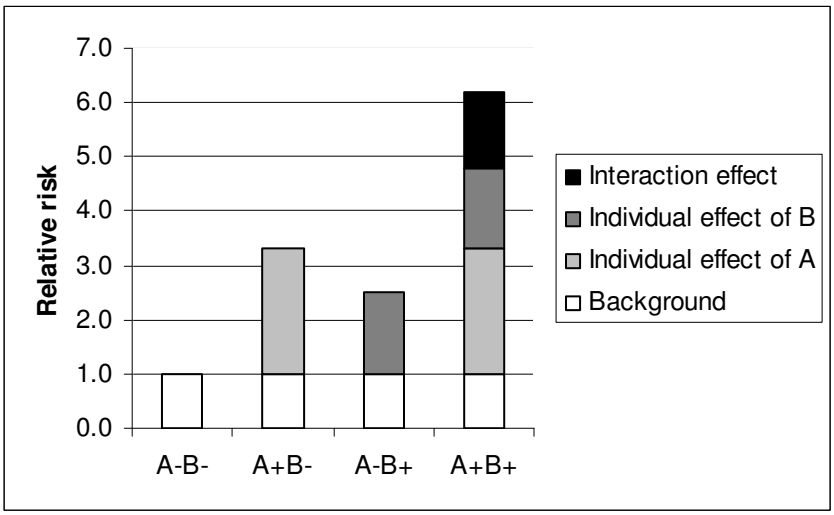
In our example the RERI is $RERI = 6.18 - 3.34 - 2.50 + 1 = 1.34$, meaning that the relative risk for hypertension in older overweight subjects is 1.34 more than if there were no interaction between age and BMI. Figure 1 shows this graphically.

Note that the absolute background risk was 4.4% and thus the absolute risk due to interaction is 5.9% (4.4% x 1.34), which is exactly the amount of interaction for absolute risks calculated with formula (4).

Assuming that the odds ratio (OR) approximates the relative risk, this formula can also be used for ORs. Note that in the absence of interaction as departure from additivity, i.e. when

there is exact additivity, RERI equals 0.

Figure 1 Relative risks in categories A-B-, A+B-, A-B+ and A+B+, divided in background, individual effect of A, individual effect of B and the interaction effect as departure from additivity.



Estimating interaction on an additive scale using logistic regression

Determinant A, determinant B and the product of A and B are included in the logistic regression model. This may seem confusing because we indicated above that a product term tests departure from multiplicativity rather than additivity in a logistic regression model. However, as shown below, the regression coefficient of the product term can also be used to calculate interaction as departure from additivity. Including determinant A, B and the product of A and B in the logistic regression formula results in:

$$\ln\left(\frac{p}{1-p}\right) = \ln(odds) = \hat{\beta}_0 + \hat{\beta}_1 A + \hat{\beta}_2 B + \hat{\beta}_3 AB \qquad (7)$$

To calculate the RERI the following three ORs are needed: 1) A+B- relative to A-B- which is $e^{\hat{\beta}_1}$, 2) A-B+ relative to A-B- which is $e^{\hat{\beta}_2}$ and 3) A+B+ relative to A-B- which is $e^{\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_3}$ (see Appendix 1). The formulas to assess presence of interaction on an additive scale and to estimate the RERI are:

$$(e^{\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_3} - 1) \neq (e^{\hat{\beta}_1} - 1) + (e^{\hat{\beta}_2} - 1) \qquad (8)$$

and

$$\hat{RERI} = e^{\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_3} - e^{\hat{\beta}_1} - e^{\hat{\beta}_2} + 1 \qquad (9)$$

The regression coefficients from the logistic regression model can be substituted into formulae (8) and (9). Important to note is that these formulas can be used for two

dichotomous as well as two continuous determinants or a combination of both.

Estimating confidence interval

A simulation study [4] showed that the first bootstrap percentile method gave better coverage of the 95% CI than the delta method [3]. Moreover, note that assessment of a continuous determinant per, e.g., 5 units instead of per 1 unit leads to a non-linear transformation of the RERI and its CI. Then the delta method cannot be used whereas bootstrapping can. For these two reasons, we adopted the first bootstrap percentile method to calculate the CI around the estimate of interaction. From the original data set 10,000 bootstrap samples (with replacement) were taken, each of which was the same size as the original sample. The RERI was then estimated in each of these new samples and the 95% CI for RERI was estimated as the 2.5th and 97.5th percentiles of the resulting bootstrap sampling distribution. The statistical program S-PLUS 6.2 was used (S-PLUS 6.2, Insightful, Seattle, USA) to carry out the bootstrapping procedure. Appendix 2 presents the script we used and an example output.

Empirical examples of estimating interaction on an additive scale

1. Two dichotomous determinants and dichotomous outcome

Consider age and BMI as risk factors for diastolic hypertension, all dichotomized as described before. Age, BMI and the product of age and BMI are entered as the independent variables and diastolic hypertension as the dependent variable in a logistic regression model. The output of the logistic regression model shows that an older person has a 3.77 times higher risk of diastolic hypertension than a young person (Table 3). Overweight subjects have a 2.72 times higher risk of diastolic hypertension compared with subjects with normal BMI. The OR and CI of the product term provide some evidence of 'negative' interaction on a multiplicative scale (OR (95% CI) = 0.80 (0.55-1.17)). Most readers will conclude that there is no interaction between age and BMI, because the OR of the product term is not significant. However, since the underlying model is a logistic regression model, this product term refers only to interaction on a multiplicative scale. The OR of 0.80 actually means that the combined effect of older age and overweight is 0.80 times the product of the individual effects of older age and overweight. To assess the amount of interaction on an additive scale we use formula (8): $(e^{\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_3}-1) \neq (e^{\hat{\beta}_1}-1)+(e^{\hat{\beta}_2}-1) \Rightarrow 7.2 > 4.5$ and formula (9): $\hat{RERI} = e^{\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_3} - e^{\hat{\beta}_1} - e^{\hat{\beta}_2} + 1 = 2.7$. There is 'positive' interaction on an *additive* scale because 7.2 is larger than 4.5. The amount of additive interaction, the RERI, is equal to 2.7. This means that the relative risk of having hypertension in older subjects with overweight is 2.7

more than if there were no interaction between age and BMI. Bootstrapping resulted in the following RERI with 95% CI: $\hat{RERI}\,(95\%CI) = 2.7\,(1.3; 4.4)$. Note that a RERI of 0 indicates exact additivity and thus no interaction on an additive scale. In this example, there is more evidence for interaction on an additive scale than for interaction on a multiplicative scale.

Table 3 Output of logistic regression model with age and BMI as dichotomous (dich) determinants and product of age and BMI entered into the model. Outcome is diastolic hypertension.

Parameter	**Estimate**	**Standard Error**	**Odds Ratio**	**95% CI of OR**	
				Lower	Upper
Age dich	1.33	0.16	3.77	2.77	5.14
BMI dich	1.00	0.14	2.72	2.05	3.61
Age dich x BMI dich	-0.23	0.19	0.80	0.55	1.17
Constant	-3.09	0.12			

Table 4 Output of logistic regression model with age (per 5 years) as continuous (cont) determinant, BMI as dichotomous (dich) determinant and product of age and BMI entered into the model. Outcome is diastolic hypertension.

Parameter	**Estimate**	**Standard Error**	**Odds Ratio**	**95% CI of OR**	
				Lower	Upper
Age (per 5 years) cont	0.26	0.03	1.29	1.22	1.36
BMI dich	1.40	0.31	4.05	2.20	7.45
Age cont x BMI dich	-0.06	0.03	0.94	0.88	1.00
Constant	-4.60	0.25			

2. One continuous and one dichotomous determinant and dichotomous outcome

Exactly the same methods as described above can be used for one dichotomous and one continuous determinant. Consider again age and BMI as risk factors for hypertension but now age is a continuous variable. Age, BMI and the product of age and BMI are entered as the independent variables and hypertension as the dependent variable in a logistic regression model. We arbitrarily chose to evaluate the effect of age per 5 years increase by dividing age by 5, and included this variable in the model. Table 4 shows that per 5 years increase of age, the risk of having diastolic hypertension increases with a factor of 1.29. Overweight subjects have a 4.05 times higher risk of diastolic hypertension compared with subjects with normal BMI. The OR and CI of the product term provide evidence for 'negative' interaction on a multiplicative scale (OR (95% CI) = 0.94 (0.88-1.00)). However, 'positive'

interaction is present on an additive scale: $(e^{\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_3}-1) \neq (e^{\hat{\beta}_1}-1)+(e^{\hat{\beta}_2}-1) = 4.0 > 3.4$ and $\hat{RERI} = e^{\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_3} - e^{\hat{\beta}_1} - e^{\hat{\beta}_2} + 1 = 0.6$ with a 95% CI of (0.3; 1.0). A RERI of 0.6 means that the relative risk of having hypertension in overweight subjects is 0.6 more with each 5 years of increase in age than if there were no interaction between age and BMI.

3. Two continuous determinants and dichotomous outcome

Consider again age and BMI as risk factors for diastolic hypertension but now as two continuous variables. Age, BMI and the product of age and BMI are entered as the independent variables and diastolic hypertension as the dependent variable in a logistic regression model. We again chose to evaluate the effect of age per 5 years. The effect of BMI was estimated per 2 units. Table 5 shows that per 5 years increase of age and per 2 units increase of BMI, the risk of having diastolic hypertension increased with a factor of 1.41 and 1.39 respectively. The OR and CI of the product term show that there is little evidence for interaction on a multiplicative scale (OR (95% CI) = 0.99 (0.98-1.01)). Calculating interaction on an additive scale gives: $(e^{\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_3}-1) \neq (e^{\hat{\beta}_1}-1)+(e^{\hat{\beta}_2}-1) \Rightarrow 0.93 > 0.80$ and $\hat{RERI} = e^{\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_3} - e^{\hat{\beta}_1} - e^{\hat{\beta}_2} + 1 = 0.14$ with a 95% CI of (0.04; 0.34). A RERI of 0.14 means that with every 5 years increase in age and 2 units increase in BMI, the relative risk of having hypertension is 0.14 more than if there were no interaction.

Table 5 Output of logistic regression model with age (per 5 years) as continuous (cont) determinant, BMI (per 2 kg/m^2) as continuous (cont) determinant and product of age and BMI entered into the model. Outcome is diastolic hypertension.

Parameter	Estimate	Standard Error	Odds Ratio	95% CI of OR	
				Lower	Upper
Age (per 5 years) cont	0.34	0.10	1.41	1.15	1.72
BMI (per 2 kg/m^2) cont	0.33	0.07	1.39	1.22	1.59
Age cont x BMI cont	-0.01	0.01	0.99	0.98	1.01
Constant	-8.09	0.92			

Robustness of RERI and CI

In the examples above, the effect of age was assessed per 5 years increase of age and the effect of BMI per 2 units increase of BMI. To assess the robustness of RERI and CI, we calculated the RERI and its 95% CI using different years of increase in age (1, 2, 5 and 10) and different units of increase in BMI (1, 2 and 5). The results are presented in Table 6. Note that

the increase in RERI with increasing units is not linear with the units increase. For example, the RERI with age per 1 unit increase and BMI per 2 units increase (0.025, Table 6) is not exactly a factor 2 larger than the RERI with age per 1 unit increase and BMI per 1 unit increase (0.011, Table 6).

Table 6 RERI and 95% confidence interval for different units increase in age (1, 2, 5 and 10) and BMI (1, 2 and 5).

RERI		**Age**			
(95% CI)		*1*	*2*	*5*	*10*
BMI	*1*	0.011 (0.004-0.026)	0.023 (0.007-0.051)	0.064 (0.020-0.158)	0.152 (0.040-0.390)
	2	0.025 (0.007-0.055)	0.051 (0.015-0.114)	0.139 (0.040-0.337)	0.327 (0.079-0.812)
	5	0.078 (0.023-0.193)	0.161 (0.047-0.412)	0.440 (0.130-1.210)	1.031 (0.287-3.106)

Table 7 Output of linear regression model with age and BMI as dichotomous determinants and product of age and BMI entered into the model. Outcome is diastolic blood pressure.

Parameter	**Estimate**	**Standard Error**	**95% CI of β**	
			Lower	Upper
Age dich	5.4	0.4	4.5	6.2
BMI dich	4.0	0.3	3.4	4.7
Age dich x BMI dich	1.1	0.6	-0.1	2.3
Constant	73.7	0.2	73.2	74.1

Estimating interaction on an additive scale using linear regression

As explained previously, in case of linear regression the regression coefficient of the product term reflects interaction as departure from additivity (Appendix 1). Consider again age and BMI as risk factors for elevated diastolic blood pressure. Age and BMI are dichotomous variables and diastolic blood pressure is a continuous variable. Age, BMI and the product of age and BMI are entered as the independent variables and diastolic blood pressure as the dependent variable in a linear regression model. The output is presented in Table 7. The combined effect of age and BMI is larger than the sum of the individual effects of age and BMI: $(84.2-73.7) \neq (79.1-73.7)+(77.7-73.7) \Rightarrow 10.5 > 9.4$ or $(5.4+4.0+1.1-0) > (5.4-0)+(4.0-0) \Rightarrow 10.5 > 9.4$.

The amount of additive interaction is: $(5.4+4.0+1.1)-5.4-4.0+0=1.1$, which (by definition) equals the regression coefficient of the product term. Note that this estimate of interaction

is not the same as "RERI" as this calculation concerns the change in absolute values of the continuous outcome instead of a change in the relative risks. The CI around the interaction estimate is easily calculated with the standard error of the regression coefficient of the product term: (-0.1; 2.3). So there is considerable evidence for 'positive' interaction on an additive scale between age and BMI as dichotomous risk factors for diastolic blood pressure on a continuous scale.

Practical use

Besides the RERI, Rothman has proposed two other measures of interaction on an additive scale: the proportion of disease among those with both exposures that is attributable to their interaction (AP) and the ratio between the combined effect and the sum of the individual effects, the synergy index (S) [2]. The formulas of these measures are:

$$AP = \frac{RERI}{RR_{A+B+}} \qquad (10)$$

and

$$S = \frac{RR_{A+B+} - 1}{(RR_{A+B-} - 1) + (RR_{A-B+} - 1)} \qquad (11)$$

Note that in the absence of interaction as departure from additivity, AP is 0 and S is 1. These measures of interaction can also be calculated in case of one or two continuous determinants using the same approach as described above for the RERI. The CI can also be obtained by bootstrapping.

Assuming that a hazard ratio approximates a relative risk, the methods to estimate interaction on an additive scale described in this paper can also be applied to Cox regression. The script for bootstrapping, however, should be adapted.

The proposed methods to estimate interaction on an additive scale with continuous determinants are included in a spreadsheet which is freely available at: www.juliuscenter.nl/additive-interaction.xls. In the spreadsheet the output of the logistic regression model (or Cox regression model) has to be filled in and all estimates of interaction (RERI, AP and S) are calculated. Furthermore, the script for bootstrapping in S-PLUS to calculate the 95% CI is presented in Appendix 2 and included in the spreadsheet.

In this article, we took Rothman's theory about the causal pie model as a starting point. This theory implies that biologic interaction is present if two causes are both needed to cause disease and therefore should be assessed as departure from additivity rather than multiplicativity. Not all researchers may agree with this view and the relevance of interaction on a multiplicative scale may be different in non-etiologic research. However, when interaction on an additive scale is the measure of interest, the methods outlined in this

paper may be used fruitfully.

Conclusions

The aim of our article was to show that interaction as departure from additivity between continuous determinants can be estimated using logistic regression analysis and to give an empirical example. The methods and formulas presented in this paper are intended to assist epidemiologists to calculate interaction on an additive scale between continuous determinants. To facilitate its use, the proposed methods are included in a spreadsheet which is freely available at: www.juliuscenter.nl/additive-interaction.xls.

References

1. Rothman KJ. *Epidemiology: An Introduction.* Oxford University Press, New York, 2002.
2. Rothman KJ. *Modern Epidemiology.* Little, Brown and Company, Boston/Toronto, 1986.
3. Hosmer DW, Lemeshow S. Confidence Interval Estimation of Interaction. *Epidemiology* 1992;3:452-6.
4. Assmann SF, Hosmer DW, Lemeshow S, Mundt KA. Confidence Intervals for Measures of Interaction. *Epidemiology* 1996;7:286-90.
5. Royston P, Altman DG, Sauerbrei W. Dichotomizing Continuous Predictors in Multiple Regression: A Bad Idea. *Stat Med* 2006;25:127-41.
6. Grobbee DE, Hoes AW, Verheij TJ, Schrijvers AJ, van Ameijden EJ, Numans ME. The Utrecht Health Project: Optimization of Routine Healthcare Data for Research. *Eur J Epidemiol* 2005;20:285-7.

Appendix 1

In linear regression the regression coefficient of the product term refers to departure from additivity whereas in logistic regression the regression coefficient of the product term refers to departure from multiplicativity. This is shown below. For simplicity we assume determinants A and B to be dichotomous (with levels 0 and 1).

Linear regression

When entering two determinants, A and B, and a product term in a linear regression model, the regression formula of the outcome Y is:

$$Y = \beta_0 + \beta_1 A + \beta_2 B + \beta_3 AB$$

The individual effect of A, assuming no effect of B, is:

$$Y = \beta_0 + \beta_1 - \beta_0 = \beta_1$$

The individual effect of B, assuming no effect of A, is:

$$Y = \beta_0 + \beta_2 - \beta_0 = \beta_2$$

The combined effect of A and B, compared to no effect of A and B, is:

$$Y = \beta_0 + \beta_1 + \beta_2 + \beta_3 - \beta_0 = \beta_1 + \beta_2 + \beta_3$$

It can be seen that the combined effect of A and B can be assessed by adding the regression coefficients of A, B and the product term. There are 3 possibilities for the regression coefficient of the product term:

1. If $\beta_3 = 0$, the combined effect of A and B = $\beta_1 + \beta_2$ → exactly additivity → no interaction as departure from additivity;
2. If $\beta_3 < 0$, the combined effect of A and B < $\beta_1 + \beta_2$ → less than additivity → 'negative' interaction as departure from additivity;
3. if $\beta_3 > 0$, the combined effect of A and B > $\beta_1 + \beta_2$ → more than additivity → 'positive' interaction as departure from additivity.

Logistic regression

When inserting two determinants, A and B, and a product term in a logistic regression model, the regression formula of the logit of p is:

$$\ln\left(\frac{p}{1-p}\right) = \ln(odds) = \beta_0 + \beta_1 A + \beta_2 B + \beta_3 AB$$

The individual effect of A, assuming no effect of B, is:

$$\ln(OR_{A+}) = \ln\left(\frac{odds(A+,B-)}{odds(A-,B-)}\right) = \ln(odds(A+,B-)) - \ln(odds(A-,B-)) = \beta_0 + \beta_1 - \beta_0 = \beta_1 \rightarrow OR_{A+} = e^{\beta_1}$$

The individual effect of B, assuming no effect of A, is:

$$\ln(OR_{B+}) = \ln\left(\frac{odds(A-,B+)}{odds(A-,B-)}\right) = \ln(odds(A-,B+)) - \ln(odds(A-,B-)) = \beta_0 + \beta_2 - \beta_0 = \beta_2 \rightarrow OR_{B+} = e^{\beta_2}$$

The combined effect of A and B, compared to no effect of A and B, is:

$$\ln(OR_{A+B+}) = \ln\left(\frac{odds(A+,B+)}{odds(A-,B-)}\right) = \ln(odds(A+,B+)) - \ln(odds(A-,B-)) = \beta_0 + \beta_1 + \beta_2 + \beta_3 - \beta_0 = \beta_1 + \beta_2 + \beta_3$$

$$\rightarrow OR_{A+B+} = e^{\beta_1+\beta_2+\beta_3} = e^{\beta_1} \times e^{\beta_2} \times e^{\beta_3} = OR_A \times OR_B \times OR_{AB}$$

It can be seen that the combined effect of A and B can be assessed by multiplying OR_A, OR_B and OR_{AB}. There are 3 possibilities for the regression coefficient of the product term:

1. If $\beta_3 = 0$, $OR_{AB} = 1$ and the combined effect of A and B = OR_A x OR_B → exactly multiplicativity → no interaction as departure from multiplicativity;
2. If $\beta_3 < 0$, $OR_{AB} < 1$ and the combined effect of A and B < OR_A x OR_B → less than multiplicativity → 'negative' interaction as departure from multiplicativity;
3. if $\beta_3 > 0$, $OR_{AB} > 1$ and the combined effect of A and B > OR_A x OR_B → more than multiplicativity → 'positive' interaction as departure from multiplicativity.

Appendix 2

Script for bootstrapping

This is a script for S-PLUS which can be used to bootstrap the RERI and its 95% confidence interval.

```
library(Design)

reri <- function(datsam)
{
fitlr <- glm(<outcome variable> ~ <variable A> * <variable B>, family=binomial, data=datsam)
reri <- exp(fitlr$coef[2]+fitlr$coef[3]+fitlr$coef[4]) - exp(fitlr$coef[2]) - exp(fitlr$coef[3]) + 1
}

summary.bootstrap(bootstrap(<data set>, reri(<data set>), B=10000, group = <data set>$<outcome variable>),
probs=c(0.025,0.5, 0.975))
```

In the general linear model (glm) command, the outcome variable and the two determinants, variable A and variable B, should be substituted. In the bootstrap command, the name of the data set should be filled in and a grouping variable should be filled in to make sure that bootstrap sampling is performed within the strata of the outcome. Furthermore, the number of samples is specified (B=10,000) in the script and the median

and 2.5th and 97.5th percentile are asked for.

Example output

This is an example of output that S-PLUS gives, when running the script described above.

```
Forming replications 1 to 100
....
Forming replications 9901 to 10000

Call:
bootstrap(data = lrgpset, statistic = reri(lrgpset), B = 10000, group = lrgpset$bpd.dich)

Number of Replications: 10000

Summary Statistics:
Observed                Bias        Mean      SE
reri      2.706         0.03215     2.738     0.7906

Empirical Percentiles:
2.5%      50%           97.5%
reri      1.28          2.709       4.377

BCa Confidence Limits:
2.5%      50%           97.5%
reri      1.293         2.701       4.388
```

Explanation of the output

Call:

Here it says that it uses the data set 'lrgpset' and it bootstraps the statistic 'reri' with a number of samples of 10,000 and the outcome 'bpd.dich' as grouping variable.

Summary statistics:

The observed or calculated RERI is 2.706. Of 10,000 samples the mean RERI is 2.738 and the standard error is 0.7906. The bias is the difference between the observed and the mean RERI.

Empirical percentiles:

Here the median value (50th percentile) and the 2.5th and 97.5th percentiles of the RERI are given.

BCa confidence limits:

These are the bias adjusted median value (50th percentile) and the bias adjusted 2.5th and 97.5th percentiles of the RERI. These 2.5th and 97.5th percentiles are the 95% confidence

interval limits. The bias adjusted median and percentiles are corrected for bias due to overfitting of the model by the bootstrap procedure.

Chapter 3.3

When one depends on the other: reporting of interaction in case-control and cohort studies published in leading journals

Abstract

Background

Interaction refers to the situation where the effect of one exposure on an outcome differs across strata of another exposure. We did a survey of epidemiological studies published in leading journals to examine current practice regarding the examination of, and reporting on, interaction.

Methods

We selected 150 case-control and 75 cohort studies published between May 2001 and May 2007 in leading general medicine, epidemiology and clinical specialist journals. Two reviewers independently extracted data on study characteristics.

Results

Of the 225 studies, 138 (61%) addressed interaction. Among the latter, 25 (18%) presented no data or only a p value or statement of statistical significance, 40 (29%) presented stratum specific effect estimates, but no meaningful comparison of these estimates and 58 (42%) presented stratum specific estimates and appropriate tests for interaction. Fifteen articles (11%) presented individual effects of both exposures as well as their joint effect or a product term, providing sufficient information to interpret interaction on an additive and multiplicative scale. Poor reporting was most prevalent in articles published in clinical specialist articles and least common in articles published in general medicine journals, with epidemiology journals in an intermediate position.

Conclusions

A majority of articles reporting cohort and case-control studies address possible interactions between exposures. However, in about half of them the information provided was unsatisfactory, and only one in ten studies reported data that allowed readers to interpret interaction effects on an additive and multiplicative scale. Efforts are required to promote more appropriate reporting on interaction.

Introduction

Interaction is an important concept in epidemiology, which refers to the situation where the effect of one exposure on an outcome differs across strata of another exposure. Other terms that have been used to describe interaction include 'effect modification', 'effect measure modification', or 'synergism', 'antagonism'.

From a statistical point of view, interaction refers to the necessity of a product term in a statistical model [1]. Depending on the form of the statistical model, interaction is examined on an additive scale, for example in linear regression, or on a multiplicative scale, for example in logistic regression. In epidemiology, there has been a long-standing debate on whether the scale should be determined by the statistical model that fits best or whether interaction should be assessed on an additive scale irrespective of the underlying statistical model [2-8]. It has been argued that the additive scale is more appropriate to assess 'biologic interaction', which reflects causal mechanisms and is implied by terms like 'synergism' and antagonism' [6,7].

An important argument for using the additive scale is that it fits with the sufficient-component concept of causality [7,8]. The presence or absence of interaction on an additive scale does not, however, indicate a particular disease mechanism: the sufficient-component theory of causation is a deliberate abstraction, which is independent of an underlying disease mechanism [7,9,10]. This may be one reason why not all researchers subscribe to the notion that interaction should always be presented on an additive scale. We believe that authors who report on interaction should provide sufficient information so that readers can interpret interaction on an additive as well as on a multiplicative scale. This approach is in line with recommendations about reporting [2,11].

Such reporting will always allow readers to interpret interaction additively, i.e., in a sufficient component-cause framework, whatever the opinion of the authors. This can be done by presenting the individual effects of both exposures as well as their joint effect, each relative to the group not exposed to either risk factor [2]. Or it can be done by presenting the full statistical model: the individual effects of both exposures and their product term, which allows readers to recalculate the joint effect [7,12,13]. Although much has been written about interaction in methodological papers, little is known about epidemiologic practice: how frequently interaction is examined in epidemiological studies, and what approach authors take when investigating and reporting interaction between exposures.

We conducted a survey of case-control and cohort studies published in leading journals to examine current practice regarding the examination of, and reporting on, interaction in cohort and case-control studies.

Methods

Selection of articles

We examined case-control and cohort studies published in five leading general medicine journals (*Annals of Internal Medicine, BMJ, JAMA, Lancet, New England Journal of Medicine*), five leading epidemiology journals (*American Journal of Epidemiology, Epidemiology, International Journal of Epidemiology, Journal of Clinical Epidemiology, Journal of Epidemiology and Community Health*), and ten leading clinical specialist journals (*American Journal of Respiratory and Critical Care Medicine, Archives of General Psychiatry, Arthritis and Rheumatism, Blood, Circulation, Clinical Infectious Diseases, Diabetes Care, Journal of American Geriatrics Society, Journal of the National Cancer Institute, Pediatrics*). We selected studies in a literature search performed at the end of March 2007 combining the journal names with the MESH term 'case-control studies' and the MESH term 'cohort studies'. Subsequently, we identified eligible studies starting with the issues published in March 2007 and went backwards in time until we identified 150 eligible case-control studies and 75 eligible cohort studies: ten case-control studies from each general medicine and each epidemiology journal, and five case-control studies from each clinical specialist journal; five cohort studies from each general medicine and each epidemiology journal, and two or three cohort studies from each clinical specialist journal. Articles that were published electronically ahead of print were included.

The inclusion of 150 case-control and 75 cohort studies was based on pragmatic considerations. We selected more case-control studies than cohort studies, because we included the case-control studies also in a separate study on the interpretation of the odds ratio in case-control research. Case-crossover studies and studies that did not report any measure of association were excluded. Finally, we only considered original articles and short reports.

Data extraction

We developed a standardized data extraction form. The extraction form was piloted on six articles not included in the study and subsequently revised. We extracted general items including the number of study participants, main exposure and condition studied, and four interaction specific items. First, we assessed whether interaction was addressed. Second, we extracted what terms the authors used to describe interaction. Third, we assessed how the interaction was presented, for example in stratified analyses or by reporting a p value. We distinguished two types of stratification: 1) presenting effect estimates of exposure and outcome in strata of the suspected effect modifying exposure; 2) presenting individual

effects of both exposures as well as their joint effect, each relative to no exposure. Table 1 and Table 2 illustrate the two types of stratification with a hypothetical example. Fourth, we assessed what statistical tests for interaction were reported, for example a Wald test or likelihood ratio test. The Wald test tests whether the regression coefficient of the product term is statistically different from zero. The likelihood ratio test tests whether the model with the product term provides a better fit than the model without the product term.

Two reviewers (MJK and PS) independently assessed all articles. Discrepancies were discussed by the two reviewers and if necessary a third person (JPV or ME) was consulted to reach consensus.

Table 1 Presentation of relative risks (95% confidence intervals) of exposure in strata of the suspected effect modifier. In this hypothetical example obesity is the risk factor of interest and smoking the suspected effect modifier.

		Suspected effect modifier	
		No smoking	Smoking
Exposure	No obesity	1.0 (reference)	1.0 (reference)
	Obesity	1.8 (1.4-2.4)	1.7 (1.2-2.3)

Table 2 Presentation of relative risks (95% confidence intervals) of individual effects of both exposures as well as their joint effect, with no exposure as the reference category. In this hypothetical example obesity is the risk factor of interest and smoking the suspected effect modifier.

Exposure	
No obesity, no smoking	1.0 (reference)
Obesity, no smoking	1.8 (1.4-2.4)
No obesity, smoking	1.3 (1.0-1.8)
Obesity, smoking	2.2 (1.6-3.0)

Levels of reporting interaction

We distinguished between four levels of reporting interaction, where each level gave more information about the degree and direction of the interaction. *Level 1* consisted of only a quantitative description that there was no interaction, or only a p value or statement of statistical significance. *Level 2* involved the presentation of separate effect estimates and confidence intervals across strata of the suspected effect modifying exposure, but no meaningful comparison of stratum specific effect estimates. *Level 3* presented effect estimates and confidence intervals across strata of the other factor as well as a p value or

Box 1 Published examples of different levels of reporting on interaction

Level	Example
1 (Descriptive, with or without statement on statistical significance)	'We found no statistically significant interactions between age and INR and age and co-morbid conditions (cerebrovascular disease, hypertension, congestive heart failure, coronary artery disease, diabetes, and cancer).' [30]
2 (Stratum specific effect estimates with confidence intervals)	'After adjustment for differences in age, ORs were higher in women compared with men for the following: abnormal WHR: women, 4.10 (95% CI, 2.59 to 6.48); men, 2.02 (95% CI, 1.54-2.65); ratio of ApoB to ApoA-1: women, 3.40 (95% CI, 2.20 to 5.25); men, 2.0 (95% CI, 1.51 to 2.66); diabetes mellitus: women, 3.52 (95% CI, 2.41 to 5.15); men, 2.23 (95% CI, 1.71 to 2.9); and hypertension: women, 3.68 (95% CI, 2.69 to 5.05); men, 2.55 (95% CI, 2.11 to 3.08). The association was stronger in men for permanent stress (OR for women, 1.97; 95% CI, 1.12 to 3.45; OR for men, 3.22 (95% CI, 2.24 to 4.63).' [31]
3 (Stratum specific effect estimates with confidence intervals and appropriate test for interaction)	'The positive association between long-term PPI therapy and hip fracture was stronger in men (OR, 1.78; 95% CI, 1.42-2.22) than women (OR, 1.36; 95% CI 1.22-1.53). The test for interaction between PPI therapy and sex was statistically significant (P=0.04).' [32]
4 (Individual and joint effect estimates using one reference category)	'The analysis of a 2x4 table for case-control designs showed that the individual OR for access thrombosis was 2.96 (95% CI 0.76-11.39) for the low TGF-β1 production haplotype and 0.81 (95% CI 0.44-1.45) for the 4G/4G PAI-I genotype. The joint OR was 5.92 (95% CI 0.82-66.21) and the synergy index was 2.47, indicating a departure from multiplicativity of the joint OR.' [33]
4 (Presentation of the full model, i.e. the individual effect estimates and the effect estimate of the product term)	'After adjustment for the use of aspirin or other NSAIDs for at least five years as well as other risk factors for colorectal cancer, statin use was still associated with as significant reduction in the risk of colorectal cancer (odds ratio, 0.55; 95 percent confidence interval, 0.40 to 0.74) (see Table). Because of reports of a synergistic effect in vitro between NSAIDs and statins, we evaluated whether the use of aspirin or other NSAIDs modified the protective effect of statins. We could find no evidence of an effect modification on a multiplicative scale (P=0.36).' [34]

Table

Variable	Unadjusted odds ratio (95% CI)	Adjusted odds ratio (95% CI)	P value
Statin use for ≥5 yr (vs. use for <5 yr or no use)	0.50 (0.40-0.63)	0.55 (0.40-0.74)	<0.001
Use of aspirin or other NSAIDs for ≥5 yr (vs. use for <5 yr or no use)	0.63 (0.51-0.77)	0.70 (0.55-0.90)	0.005
Interaction term	-	1.30 (0.74-2.27)	0.36

statement of statistical significance based on an appropriate test for interaction. Finally, *level 4* provided sufficient information to interpret interaction on an additive as well as on a multiplicative scale. This could be done in two ways: presentation of the individual effect estimates and the joint effect estimate using one reference category or presentation of the full model, i.e. the individual effect estimates and the effect estimate of the product term. Additional measures of interaction may have been presented, including the synergy index of multiplicativity or additivity, or the relative excess risk due to interaction [6,12]. Published examples of the different levels of reporting are shown in Box 1.

Data analysis

Frequencies and summary statistics were calculated stratified by journal type (general medicine, epidemiology and clinical specialist journals) and study type (case-control versus cohort studies).

Results

Our literature search for case-control studies produced 4647, 3351 and 6508 hits in the general medicine, epidemiology and clinical specialist journals, respectively. Based on this search we identified the 50 most recent eligible case-control studies for each journal type. The literature search for cohort studies resulted in 13,856, 4986 and 15,348 hits, respectively. Again, based on the search we selected the 25 most recent eligible cohort studies for each journal type. The publication date of the selected articles ranged from May 2001 to April 2007 in the general medicine journals (median: June 2006), from October 2002 to May 2007 in the general epidemiology journals (median: October 2006), and from August 2004 to April 2007 in the clinical specialist journals (median: January 2007). Fourteen of the 225 articles (6%) were short reports; six in general medicine journals, four in general epidemiology journals and four in clinical specialist journals.

Table 3 presents the characteristics of the 225 studies, stratified by journal type. Most authors and study participants came from the United States or Europe. The largest studies were published in general medicine journals, closely followed by studies published in epidemiology journals. Studies published in clinical specialist journals were smaller. As expected, cohort studies included more individuals than case-control studies (data not shown). The most frequent exposures were treatments, followed by prevalent medical conditions and lifestyle factors. Cardiovascular disease, cancer and all-cause mortality were the most frequently studied outcomes.

Table 3 Study characteristics of 225 case-control and cohort studies by type of journal.

	General medicine articles (n=75)	**General epidemiology articles** (n=75)	**Clinical specialist articles** (n=75)
Country of authors*			
United States	42 (56%)	32 (43%)	44 (59%)
Europe except UK	27 (36%)	30 (40%)	26 (35%)
United Kingdom	17 (23%)	13 (17%)	13 (17%)
Other	22 (29%)	18 (24%)	17 (23%)
Country of study participants*			
United States	33 (44%)	23 (31%)	32 (43%)
Europe except UK	20 (27%)	30 (40%)	25 (33%)
United Kingdom	15 (20%)	13 (17%)	10 (13%)
Other	20 (27%)	13 (17%)	16 (21%)
Number of subjects**	3570 (57-383,862)	3078 (100-10,766,971)	1206 (33-1,380,000)
Exposure category*			
Treatment	29 (39%)	9 (12%)	15 (20%)
Prevalent medical conditions	9 (12%)	9 (12%)	19 (25%)
Lifestyle factors	9 (12%)	15 (20%)	10 (13%)
Other	32 (40%)	46 (57%)	39 (51%)
Outcome category*			
Cardiovascular disease	26 (35%)	12 (16%)	20 (27%)
Cancer	6 (8%)	23 (31%)	14 (19%)
All-cause mortality	12 (16%)	11 (15%)	6 (8%)
Infectious disease	12 (16%)	4 (5%)	13 (17%)
Other	20 (27%)	26 (35%)	23 (31%)

* More than one possible in one article; ** Median (range) presented

Table 4 describes details on the reporting of interaction. About two-third of the studies (138 articles, 61%) examined interaction between exposures. In twelve of these studies (9%), mainly published in epidemiology journals, this was stated as one of the objectives of the study. An additional five studies (4%), of which four were published in general medicine journals, explicitly stated that the interaction analyses were pre-specified. In all journals the most frequently used terms were 'interaction' and 'effect (measure) modification' (101 of 138 articles, 73%). The term 'subgroup analysis' was used in 18 studies (13%), 'stratification' or 'stratified analysis' in 11 studies (8%). 'Synergy', 'combined effect' or 'joint effect' was mentioned in six studies (4%). Seventeen studies (12%) did not use an explicit term.

Table 4 Reporting on interaction in 225 case-control and cohort studies by type of journal.

	General medicine articles (n=75)	**General epidemiology articles** (n=75)	**Clinical specialist articles** (n=75)
Interaction examined	46 (61%)	50 (67%)	42 (56%)
Terms used*			
'Interaction'	29 (63%)	19 (38%)	24 (57%)
'Effect (measure) modification'	10 (22%)	12 (24%)	7 (17%)
'Subgroup analysis'	9 (20%)	5 (1%)	4 (10%)
'Stratification'	3 (7%)	4 (8%)	4 (10%)
'Synergy'	2 (4%)	3 (6%)	1 (2%)
No specific term used	3 (7%)	9 (18%)	5 (12%)
Presentation of interaction*			
Stratum specific effect estimates**	33 (72%)	36 (72%)	29 (69%)
Individual and joint effect estimates***	4 (9%)	6 (12%)	4 (10%)
P value	22 (48%)	13 (26%)	16 (38%)
Statement on statistical significance	8 (17%)	13 (26%)	10 (24%)
Product term	2 (4%)	0 (0%)	1 (2%)
Synergy index	2 (4%)	0 (0%)	1 (2%)
No data presented	4 (9%)	3 (6%)	1 (2%)
Statistical tests used			
Wald test for product term	11 (24%)	11 (22%)	6 (14%)
Likelihood ratio test	6 (13%)	8 (16%)	9 (21%)
Heterogeneity test	1 (2%)	1 (2%)	0 (0%)
Unclear	14 (30%)	7 (14%)	13 (31%)

* More than one possible in one article, percentages relate to studies that examined interaction

** See table 1 for example

*** See table 2 for example

The most frequent approach to reporting interactions (98 of 138 articles, 71%) involved the presentation of stratum specific effect estimates (see Table 1 for example). Only 14 studies (10%) reported individual effects of both exposures as well as their joint effect (see Table 2 for example), and few studies reported product terms or synergy indices. No study presented the relative excess risk due to interaction (RERI) or the attributable proportion due to interaction. P values or statements on statistical significance were reported in 82 (59%) articles, but the statistical test used often remained unclear. One study incorrectly examined interaction by assessing overlap of the confidence intervals around the effect estimates in the strata. Another invalid interpretation is that interaction is present because the p value in one stratum is significant but the p value in the other stratum is not. None of

the studies explicitly reported that this method was used; three of the studies in which the statistical test was unclear may have used this method.

Table 5 Different levels of reporting on interaction in 225 case-control and cohort studies by type of journal.

Level of reporting*	**Total** (n=138)	**General medicine articles** (n=46)	**General epidemiology articles** (n=50)	**Clinical specialist articles** (n=42)
Level 1	25 (18%)	8 (17%)	8 (16%)	9 (21%)
Level 2	40 (29%)	10 (22%)	16 (32%)	14 (33%)
Level 3	58 (42%)	23 (50%)	20 (40%)	15 (36%)
Level 4**	15 (11%)	5 (11%)	6 (12%)	4 (10%)

*See box 1 and main text for definitions and examples of levels of reporting

** Three of the studies of level 4 mentioned the additive scale when discussing the examined interaction

When analysing the data across “levels of reporting interaction” defined in the Methods section, 25 articles (18%) were categorized as level 1 and 40 articles (29%) met criteria for level 2 (Table 5). Fifty-eight articles (42%) presented effect estimates in strata of the suspected effect modifier and provided a p value or statement of statistical significance (level 3). Only few articles (15, 11%) presented individual effect estimates of both exposures and their joint effect, or a product term from a statistical model (level 4). Three of these studies mentioned the additive scale when discussing interaction. One of these studies presented the synergy index for additive interaction in their results. Another study mentioned in the discussion that the interaction they observed was more than could be expected under a multiplicative model. The third study described, also in the discussion, that some of the interactions they examined were consistent with an additive model and some were consistent with a multiplicative model. Two other studies, which were both case-control studies, presented a synergy index of multiplicativity. Of note, among the studies in level 4, six did not provide a confidence interval around the synergy index or a p value for interaction.

Few differences were evident when comparing reporting between groups of journals. Interestingly, the term ‘interaction’ was used somewhat less frequently in epidemiology journals than in general medicine and clinical specialist journals (Table 4). In epidemiology journals reporting on the statistical test used to assess interaction was more complete than in other journals. Poorer reporting (levels 1 and 2) was most prevalent in articles published

in clinical specialist articles (23 of 42 articles, 54%), while better reporting (levels 3 and 4) was most common in articles published in general medicine journals (28 of 46 articles, 61%), with studies published in epidemiology journals in an intermediate position (Table 5).

Discussion

This survey of current practice regarding the examination of, and reporting on, interaction in epidemiological studies found that interaction between exposures is addressed in a majority of studies published in leading journals. However, in about half of the studies the approach used or information reported was inadequate. In particular, brief descriptive statements or the presentation of stratum specific estimates, with no meaningful comparison across the strata was common. Only one in ten studies reported the individual effects of both exposures and their joint effect or product term, thus providing sufficient information to interpret interaction on an additive as well as a multiplicative scale. Reporting on interaction was not more comprehensive in epidemiology journals compared to general medicine or clinical specialist journals.

Our study was based on articles published in leading journals and our results may therefore not be applicable to cohort or case-control studies published in less prominent journals. We did not assess for each study whether the presence or absence of interaction between exposures would have been important from a clinical or public health perspective, nor did we assess whether the authors were capable of distinguishing important from unimportant interaction. Many authors may feel that if little evidence for an interaction was found, it would be unnecessary to report this in greater detail. Furthermore, space constraints and word limits may have meant that relevant material was removed after the article was accepted for publication, and reporting ceased to be comprehensive after such editorial interventions. A study of reports of clinical trials submitted to the BMJ and later published either in the BMJ or another journal does not, however, support this hypothesis: the number of tables and figures did not change markedly between submission and publication [14].

Pocock *et al* recently reviewed the practice in the analysis and reporting of epidemiological research published in epidemiology journals or general and specialist medical journals [15]. In line with our findings, a large proportion of articles (43 of 73, 59%) included subgroup analyses; the majority of these claimed differences across groups. Two surveys on reporting of clinical trials found that one invalid interpretation was quite frequent: that interaction is present because effect estimates reach conventional levels of statistical significance (i.e. $p < 0.05$) in one subgroup but not the other (9 of 17 studies [16]; 13

of 35 studies [17]). Surprisingly, in our survey only three studies may have used this method. Apparently, researchers in observational epidemiology are more aware of the shortcomings of this method of examining interaction than researchers in the field of clinical trials.

In this survey, 17 of the 138 articles that reported interaction (12%) explicitly stated that the examination of interaction was an objective of the study or that the interaction analyses were pre-specified. Some studies mentioned the assessment of interaction early on in the introduction, others in the methods or even only in the results. Analyses that were not the original aim of the study but arose during data analysis, e.g., because of additional subject matter knowledge or new literature, may often be useful and provide important insights. However, it has been advised that such analyses should be described accordingly [11,17]. This may give information whether the original data collection (assessment of exposure and confounders relevant to the exposure) was adequate for a particular new hypothesis. It also allows the reader to assess the implications of multiplicity of subgroups in the analysis [11]. To some readers this is important information as they believe that the credibility of data on a hypothesis that was specified beforehand is stronger, a view that is commonly held for randomized trials [18]. Readers with other views about inference in etiologic research, will pay less attention to this information, as they argue that the prior probability of the hypothesis (whether known before or after looking at the data) is more important and that replication will tease out which findings stand the test of time [19,20].

Only three out of the 138 studies mentioned the additive scale for the interpretation of interaction, and one of these calculated a synergy index for additivity. No study used the relative excess risk due to interaction or the attributable proportion due to interaction. We noticed some missed opportunities for assessing interaction on an additive scale, in studies that found similar odds ratios across strata of a strong risk factor, which means that there probably was quite strong interaction on an additive scale, with potential public health impact. The long-standing debate on the appropriate scale on which to assess interaction, and the call to present interaction on the additive scale to more appropriately reflect the causal structure of interactions [3-5,7,8,11], thus still does not appear to have reached main stream, applied epidemiology. A more pervasive and more widely accepted argument is that in many situations the additive scale, which uses absolute risks, will be more appropriate for public health and clinical decision making [11,21]. Nevertheless, it is clear from our study that the reporting of interaction on an additive scale is very uncommon. One little-known drawback of calculating indices for additive interaction like RERI from multiplicative models, is that for case-control studies this calculation is only exact for tables or models that do not include other variables (e.g., confounders) over which the index might vary [22,23]. If the RERI is

different among, say, men and women, the RERI calculated from a logistic model that includes sex as an additional covariate, will yield only an approximation which can be seen as an approximate weighted average of the sex-specific RERIs. The synergy index for additivity appears to be more resistant to this problem: a simulation study showed it was stable in most situations and more often stable than the RERI [23]. In cohort studies, the problem does not arise if additive risk models are fitted [22], which was not done in any paper in our survey.

In conclusion, our study demonstrates that the reporting on interaction is poor in epidemiological studies and that efforts are needed to improve this situation. The recommendations put forward by the STROBE initiative [11,24] address several aspects relevant to interaction and subgroup analyses. STROBE asks authors to describe any methods used to examine subgroups and interactions and to report which analyses were planned in advance to allow readers who desire so, to judge the implications of multiplicity. Of note, the STROBE group supports Botto and Khoury's proposal [2] to present individual effects of exposures as well as joint effects, each relative to no exposure, as the most informative approach to presenting the data. A less intuitive alternative is the presentation of the full model, including the interaction term. Only one in ten articles (15 of 138) in our survey had either of these two presentations. More than 10 years ago an international group developed the Consolidated Standards of Reporting Trials (CONSORT) to improve reporting of randomized clinical trials [25], which was revised in 2001 [26,27]. Bibliographic studies have shown that the quality of reporting of trials has improved, possibly as a consequence of the CONSORT recommendations [28,29]. Several journals now ask authors to follow the STROBE recommendations and there is therefore hope that the shortcomings evident in the present study will no longer, or to a lesser extent, be present in reports of cohort and case-control studies published in a few years time.

References

1. Ahlbom A and Alfredsson L. Interaction: A Word With Two Meanings Creates Confusion. *Eur J Epidemiol* 2005;20:563-4.
2. Botto LD and Khoury MJ. Commentary: Facing the Challenge of Gene-Environment Interaction: the Two-by-Four Table and Beyond. *Am J Epidemiol* 2001;153:1016-20.
3. Greenland S and Poole C. Invariants and Noninvariants in the Concept of Interdependent Effects. *Scand J Work Environ Health* 1988;14:125-9.
4. Rothman KJ. Synergy and Antagonism in Cause-Effect Relationships. *Am J Epidemiol* 1974;99:385-8.
5. Rothman KJ, Greenland S, Walker AM. Concepts of Interaction. *Am J Epidemiol* 1980;112:467-70.
6. Rothman KJ. *Interactions between Causes.* In *Modern Epidemiology*. pp. 311-326, Little, Brown and Company, Boston/Toronto, 1986.

7. Greenland S and Rothman KJ. *Concepts of Interaction.* In *Modern Epidemiology*. pp. 329-342, Lippincott-Raven Publishers, Philadelphia, PA, 1998.
8. VanderWeele TJ and Robins JM. The Identification of Synergism in the Sufficient-Component-Cause Framework. *Epidemiology* 2007;18:329-39.
9. Siemiatycki J and Thomas DC. Biological Models and Statistical Interactions: an Example From Multistage Carcinogenesis. *Int J Epidemiol* 1981;10:383-7.
10. Thompson WD. Effect Modification and the Limits of Biological Inference From Epidemiologic Data. *J Clin Epidemiol* 1991;44:221-32.
11. Vandenbroucke JP, von EE, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *PLoS Med* 2007;4:e297.
12. Hosmer DW and Lemeshow S. Confidence Interval Estimation of Interaction. *Epidemiology* 1992;3:452-6.
13. Knol MJ, Tweel IV, Grobbee DE, Numans ME, Geerlings MI. Estimating Interaction on an Additive Scale Between Continuous Determinants in a Logistic Regression Model. *Int J Epidemiol* 2007;36:1111-8.
14. Schriger DL, Sinha R, Schroter S, Liu PY, Altman DG. From Submission to Publication: a Retrospective Review of the Tables and Figures in a Cohort of Randomized Controlled Trials Submitted to the British Medical Journal. *Ann Emerg Med* 2006;48:750-6, 756.
15. Pocock SJ, Collier TJ, Dandreo KJ, De Stavola BL, Goldman MB, Kalish LA et al. Issues in the Reporting of Epidemiological Studies: a Survey of Recent Practice. *BMJ* 2004;329:883.
16. Moreira ED, Jr., Stein Z, Susser E. Reporting on Methods of Subgroup Analysis in Clinical Trials: a Survey of Four Scientific Journals. *Braz J Med Biol Res* 2001;34:1441-6.
17. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup Analysis and Other (Mis)Uses of Baseline Data in Clinical Trials. *Lancet* 2000;355:1064-9.
18. Rothwell PM. Treating Individuals 2. Subgroup Analysis in Randomised Controlled Trials: Importance, Indications, and Interpretation. *Lancet* 2005;365:176-86.
19. Goodman SN. Multiple Comparisons, Explained. *Am J Epidemiol* 1998;147:807-12.
20. Rothman KJ. No Adjustments Are Needed for Multiple Comparisons. *Epidemiology* 1990;1:43-6.
21. Szklo M and Nieto FJ. *Communicating Results of Epidemiologic Studies* (Chapter 9). In *Epidemiology, Beyond the Basics.* pp. 408-430, Jones and Bartlett, Sudbury, MA, 2000.
22. Greenland S. Additive Risk Versus Additive Relative Risk Models. *Epidemiology* 1993;4:32-6.
23. Skrondal A. Interaction As Departure From Additivity in Case-Control Studies: a Cautionary Note. *Am J Epidemiol* 2003;158:251-8.
24. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. *Ann Intern Med* 2007;147:573-7.
25. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I et al. Improving the Quality of Reporting of Randomized Controlled Trials. The CONSORT Statement. *JAMA* 1996;276:637-9.

26. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D et al. The Revised CONSORT Statement for Reporting Randomized Trials: Explanation and Elaboration . *Ann Intern Med* 2001;134:663-94.
27. Moher D, Schulz KF, Altman DG, for the CONSORT Group. The CONSORT Statement: Revised Recommendations for Improving the Quality of Reports of Parallel Group Randomized Trials. *Lancet* 2001;357:1191-4.
28. Moher D, Jones A, LePage L, for the CONSORT Group. Does the CONSORT Statement Improve the Quality of Reports of Randomized Trials? A Comparative Before and After Evaluation. *JAMA* 2001;285:1992-5.
29. Devereaux PJ, Manns BJ, Ghali WA, Quan H, Guyatt GH. The Reporting of Methodological Factors in Randomized Controlled Trials and the Association With a Journal Policy to Promote Adherence to the Consolidated Standards of Reporting Trials (CONSORT) Checklist. *Control Clin Trials* 2002;23:380-8.
30. Fang MC, Chang Y, Hylek EM, Rosand J, Greenberg SM, Go AS et al. Advanced Age, Anticoagulation Intensity, and Risk for Intracranial Hemorrhage Among Patients Taking Warfarin for Atrial Fibrillation. *Ann Intern Med* 2004;141:745-52.
31. Lanas F, Avezum A, Bautista LE, Diaz R, Luna M, Islam S et al. Risk Factors for Acute Myocardial Infarction in Latin America: the INTERHEART Latin American Study. *Circulation* 2007;115:1067-74.
32. Yang YX, Lewis JD, Epstein S, Metz DC. Long-Term Proton Pump Inhibitor Therapy and Risk of Hip Fracture. *JAMA* 2006;296:2947-53.
33. Lazo-Langner A, Knoll GA, Wells PS, Carson N, Rodger MA. The Risk of Dialysis Access Thrombosis Is Related to the Transforming Growth Factor-Beta1 Production Haplotype and Is Modified by Polymorphisms in the Plasminogen Activator Inhibitor-Type 1 Gene. *Blood* 2006;108:4052-8.
34. Poynter JN, Gruber SB, Higgins PD, Almog R, Bonner JD, Rennert HS et al. Statins and the Risk of Colorectal Cancer. *N Engl J Med* 2005;352:2184-92.

Chapter 3.4

Unexpected lower prevalence of depression in patients with diabetes: potential for selection bias in a waiting room population

Abstract

Introduction

Our initial aim was to study the association between diabetes and depression in a population of general practice attendees. Unexpectedly, the prevalence of depression was much lower in diabetes patients (3.6%) than in those without (14.1%; OR: 0.23, 95% CI: 0.07-0.73). We investigated whether these unexpected findings could be explained by selection bias.

Methods

We used data from a prospective cohort study (PREDICT-NL) on the development of a multifactor risk algorithm for major depression. Patients who visited the general practitioner (GP) were asked to participate. Presence of major depressive disorder was assessed with a diagnostic interview. Diabetes was defined as a doctor's diagnosis of diabetes or use of diabetes medication within the six years before inclusion into the study. Based on the questionnaire, we categorized the reason for visiting the GP into being a control visit or not.

Results

A total of 1286 participants (37.2% men; mean age +/- SD: 51 +/- 17 years) were included. The percentage of subjects visiting the GP for a control visit was higher in diabetes patients than in patients without diabetes (OR: 2.79, 95% CI: 1.78-4.36), while the percentage of subjects visiting the GP for a control visit was lower in subjects with depression than in those without (OR: 0.64, 95% CI: 0.44-0.94). However, stratification for control visit revealed that in the stratum of no control visit, the prevalence of depressive disorder was still lower in patients with diabetes than in those without (OR: 0.45, 95% CI: 0.14-1.49), although the difference was somewhat smaller.

Conclusions

This study showed that the lower prevalence of depression among diabetes patients could at least in part be explained by differential selection based on the reason for visiting the GP. Researchers should be aware of the possibility of selection bias in studies that recruit participants in waiting rooms.

Introduction

Several publications have described the association between diabetes and depression in cross-sectional as well as longitudinal studies. A meta-analysis of ten cross-sectional studies found a 1.8-fold increased prevalence of clinically relevant depression in adults with type 2 diabetes compared with those without diabetes [1]. However, most longitudinal studies reported no increased incidence of depression in diabetes patients compared with those without [2-5] or compared with subjects with other diseases [6,7]. A recent study did find an association between diabetes and recurrent depressed mood [8]. Most of these cross-sectional and longitudinal studies used self-reported questionnaires to assess depressive symptoms and used self-report to assess diabetes. The use of depressive symptoms as a measure for depression may not have been specific enough for depression. In addition, self-report questionnaires might measure diabetes related problems such as fatigue and changed eating behaviour rather than true depression. Self-reported diabetes may over- or underestimate the true prevalence of diabetes.

We aimed to assess the prevalence and incidence of major depressive disorder according to DSM-IV criteria in subjects with and without diabetes as diagnosed by a general practitioner (GP). Due to unexpected findings for the association between diabetes and depression, we subsequently investigated whether these could be explained by selection bias.

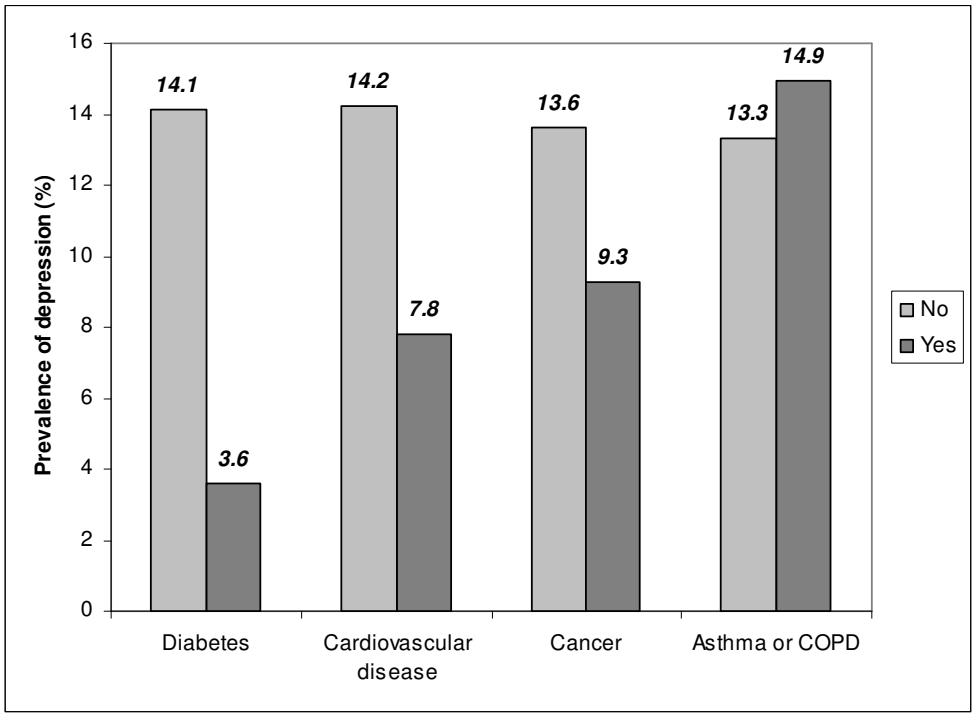
Methods and Results

We used data from the PREDICT-NL study, which is the Dutch part of the PREDICT study, a European prospective cohort study aimed to develop a multifactor risk algorithm for onset of major depression over 12 months [9]. On random days, patients of 18 years or older who visited the GP were asked to participate while waiting to see their doctor. If interested, they were asked to sign informed consent within two weeks and complete a risk factor questionnaire. When the informed consent and questionnaire returned, an appointment was made to conduct a diagnostic interview at the general practice to assess presence or absence of major depressive disorder according to DSM-IV criteria [10,11]. Follow-up measurements were performed after six and twelve months. Diabetes was defined as presence of an ICPC (International Classification of Primary Care) code for diabetes (T90) or an ATC (Anatomical Therapeutic Classification) code for diabetes medication (A10A and A10B) within the six years before inclusion into the study. A total of 1286 participants were included in the analyses of this study, of which 37.2% were male and the mean age was 51 years (SD: 17 years). Eighty-three participants (6.5%) were classified as having diabetes and

the overall prevalence of depressive disorder was 13.5%.

The prevalence of depressive disorder in subjects with and without diabetes is presented in Figure 1. Unexpectedly, the prevalence of depression was much lower in diabetes patients (3.6%) than in subjects without diabetes (14.1%; OR: 0.23, 95% CI: 0.07-0.73). Also, the incidence of depression was lower in patients with diabetes compared to those without, although numbers were small (data not shown).

Figure 1 Prevalence of depressive disorder (%) in subjects with and without diabetes, cardiovascular disease, cancer, and asthma or COPD



We investigated whether our unexpected results could be due to selection bias. We hypothesized that patients with diabetes would more frequently visit their GP for a control visit than patients without diabetes. Furthermore, we hypothesized that patients with depression would be less likely to visit their GP for a control visit than patients without a depression because patients with a depression would rather come with a complaint related to depression. Based on the questionnaire, we categorized the reason for visiting the GP at the date of recruiting into being a control visit or not. The percentage of subjects visiting the GP for a control visit was indeed higher in diabetes patients than in subjects without diabetes (51.8% vs. 27.8%; OR: 2.79, 95% CI: 1.78-4.36). The percentage of subjects visiting

the GP for a control visit was indeed lower among subjects with depressive disorder than among subjects without a depressive disorder (22.0% vs. 30.5%, OR: 0.64, 95% CI: 0.44-0.94). However, adjustment for control visit did not change the odds ratio for the association between diabetes and depression (OR adjusted for age and sex: 0.32, 95% CI: 0.10-1.03; OR adjusted for age, sex and control visit: 0.33, 95% CI: 0.10-1.09). In addition, stratification for control visit revealed that in the stratum of no control visit, the prevalence of depressive disorder was still lower in patients with diabetes (7.5%) than in subjects without diabetes (15.2%; OR: 0.45, 95% CI: 0.14-1.49), although the difference was somewhat smaller.

We investigated whether this potential selection bias was also present for the association between other chronic diseases and depression. Cardiovascular disease (ICPC codes: K73-K84, K89-K92; n=154, 12.0%), cancer (ICPC codes: A79, B72-B74, D74-D77, F74, H75, K72, L71, N74, R84, R85, S77, S80, T71, T73, U75-U77, U79, W72, X75-X77, X81, Y77, Y78; n=54, 4.2%), and asthma or COPD (ICPC codes: R91, R92, R95; n=114, 8.9%) were defined as presence of one of the respective ICPC codes within the six years before inclusion into the study. We found a lower prevalence of depressive disorder in patients with cardiovascular disease (Figure 1; OR: 0.51, 95% CI: 0.28-0.94) and cancer (OR: 0.65, 95% CI: 0.25-1.65), while patients with asthma or COPD had a similar prevalence of depression as subjects without asthma or COPD (OR: 1.14, 95% CI: 0.66-1.96). The percentage of subjects visiting the GP for a control visit was higher in cardiovascular disease patients (OR: 1.35, 95% CI: 0.94-1.92), and in patients with cancer (OR: 1.98, 95% CI: 1.14-3.44), but was not different in patients with and without asthma or COPD (OR: 0.98, 95% CI: 0.64-1.49). Similarly to diabetes, the odds ratios did not change after adjustment for control visit (data not shown).

Figure 2 describes the assumed directions of the associations between diabetes, depression and control visit. Since it is unlikely that control visit is an independent risk factor for depression, it cannot be a confounder. This concurs with our observation that adjustment for control visit did not change the associations. This observation in turn concurs with the general idea that it is not always possible to adjust for selection bias [12]. However, in patients without a control visit we still found a lower prevalence of depression in diabetes patients, which we would not expect if control visit was the only factor that caused differential selection. Figure 3 shows that the odds ratio of depression among patients with the different chronic diseases is inversely associated with the odds ratio of a control visit for patients with chronic diseases. This observation strengthens our hypothesis that differential selection of subjects visiting the GP for a control visit is a likely explanation for our findings.

Figure 2 Suspected directions of associations between diabetes, depression and control visit

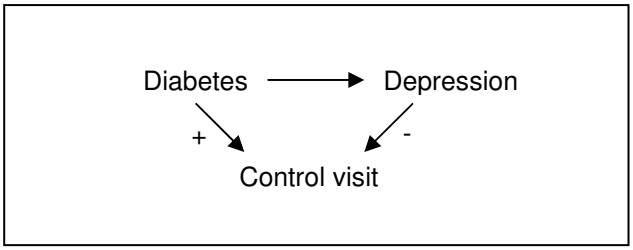
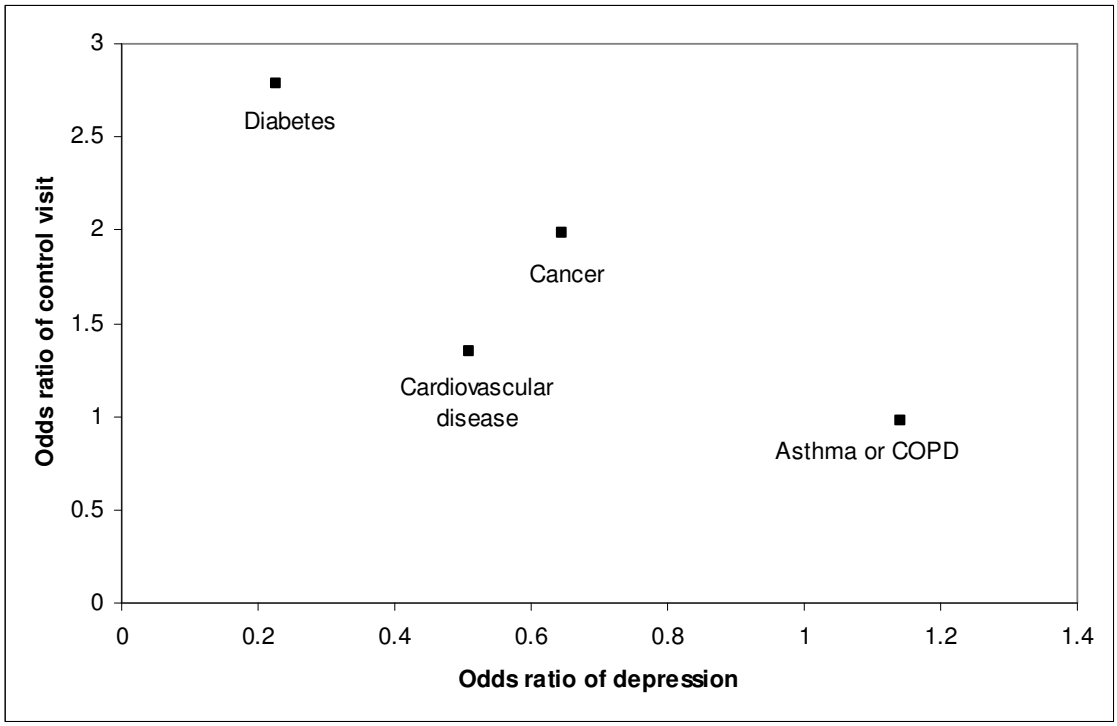


Figure 3 Association between the odds ratio of depression and the odds ratio of a control visit in subjects with diabetes, cardiovascular disease, cancer, and asthma or COPD



Discussion

In this study among patients selected from the waiting room of GPs the risk of depression was reduced among those with diabetes. This finding contradicts the currently available evidence, as the prevalence of depression is found to be higher in diabetes patients compared with those without [1,13]. Further analysis of our data strongly suggested that the lower prevalence of depression in diabetes patients was at least partly explained by selection bias. Other possible explanations for our findings are misclassification, selective non-response and differential use of antidepressant use. Misclassification, however, seems

not plausible as both chronic diseases and depression were measured objectively and without knowledge of the chronic disease when assessing depression, and without knowledge of depression status when assessing chronic diseases. Selective response may have led to the relatively high overall prevalence of depression (13.5%). However, it is not likely that this selective response was different between depressed patients with or without diabetes. When we added antidepressant users without a diagnosis of depressive disorder to the depression group, similar patterns as in Figure 1 were seen (data not shown). Alternatively, it is possible that there really is an inverse association between diabetes and depression.

We think that the potential selection bias due to control visit we described in this paper is specific for a population of GP attendees, i.e. subjects sampled from a waiting room of the GP. This selection bias would probably not have occurred if study participants were sampled from a GP population as a whole or from a general population, because in these populations selection of the patients would not depend on how often they visit the GP for a control visit. Apparently, the degree of selection bias in a population of GP attendees depends on the exposure under study as we observed differences between the chronic diseases. This difference could be due to different (adherence to) guidelines for control visits for the chronic diseases.

In conclusion, this study illustrates the potential for selection bias in a population of GP attendees due to differential selection based on the reason for visiting the GP. We showed that the degree of selection bias depends on the occurrence relation under study. Researchers should be aware of the possibility of selection bias in studies that recruit participants from a waiting room.

References

1. Ali S, Stone MA, Peters JL, Davies MJ, Khunti K. The Prevalence of Co-Morbid Depression in Adults With Type 2 Diabetes: a Systematic Review and Meta-Analysis. *Diabet Med* 2006;23:1165-73.
2. Engum A. The Role of Depression and Anxiety in Onset of Diabetes in a Large Population-Based Study. *J Psychosom Res* 2007;62:31-8.
3. Palinkas LA, Lee PP, Barrett-Connor E. A Prospective Study of Type 2 Diabetes and Depressive Symptoms in the Elderly: the Rancho Bernardo Study. *Diabet Med* 2004;21:1185-91.
4. Brown LC, Majumdar SR, Newman SC, Johnson JA. Type 2 Diabetes Does Not Increase Risk of Depression. *CMAJ* 2006;175:42-6.
5. Polsky D, Doshi JA, Marcus S, Oslin D, Rothbard A, Thomas N et al. Long-Term Risk for Depressive Symptoms After a Medical Diagnosis. *Arch Intern Med* 2005;165:1260-6.
6. Bisschop MI, Kriegsman DM, Deeg DJ, Beekman AT, van Tilburg W. The Longitudinal Relation Between Chronic Diseases and Depression in Older Persons in the Community: the Longitudinal

Aging Study Amsterdam. *J Clin Epidemiol* 2004;57:187-94.

7. Kessing LV, Nilsson FM, Siersma V, Andersen PK. No Increased Risk of Developing Depression in Diabetes Compared to Other Chronic Illness. *Diabetes Res Clin Pract* 2003;62:113-21.
8. Maraldi C, Volpato S, Penninx BW, Yaffe K, Simonsick EM, Strotmeyer ES et al. Diabetes Mellitus, Glycemic Control, and Incident Depressive Symptoms Among 70- to 79-Year-Old Persons: the Health, Aging, and Body Composition Study. *Arch Intern Med* 2007;167:1137-44.
9. King M, Weich S, Torres F, Svab I, Maaroos H, Neeleman J et al. Prediction of Depression in European General Practice Attendees: the PREDICT Study. *BMC Public Health* 2006;6:6.
10. World Health Organisation: Composite International Diagnostic Interview (CIDI). Version 2.1, WHO, Geneva, 1997.
11. Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J et al. The Composite International Diagnostic Interview. An Epidemiologic Instrument Suitable for Use in Conjunction With Different Diagnostic Systems and in Different Cultures. *Arch Gen Psychiatry* 1988;45:1069-77.
12. Rothman KJ and Greenland S. *Modern Epidemiology*. Lippincott-Raven Publishers, Philadelphia, PA, 1998.
13. Anderson RJ, Freedland KE, Clouse RE, Lustman PJ. The Prevalence of Comorbid Depression in Adults With Diabetes: a Meta-Analysis. *Diabetes Care* 2001;24:1069-78.

Chapter 3.5

What do case-control studies estimate? Survey of methods and assumptions in published case-control research

Abstract

Background

In case-control research the approach to sample cases and controls determines whether the odds ratio can be interpreted as a risk ratio, rate ratio, odds ratio or prevalence odds ratio.

Methods

The authors examined 150 case-control studies published in leading general medicine, epidemiology and clinical specialist journals from May 2001 to April 2007. Two reviewers independently extracted data on study characteristics.

Results

The majority of the studies (125 of 150, 83%) were based on incident cases and among these the source population was mostly dynamic (102 of 125, 82%). Only a minority (23 of 125, 18%) sampled from a fixed cohort. Among studies with incident cases 105 (84%) estimated the rate ratio; 57 studies (46%) required the source population to be stable whereas 48 (38%) did not need this assumption because of matching on time or concurrent sampling. In 17 studies (14% of 125) a risk ratio was estimated, with 16 requiring the rare disease assumption. The rare disease assumption was discussed in four studies, but it was not relevant to any of these. No study mentioned the need for a stable population.

Conclusions

The authors conclude that the assumption of a stable population is required considerably more frequently than the rare disease assumption. At present, researchers rarely discuss what odds ratios from case-control studies estimate.

Introduction

Case-control studies are an important study type in observational research. Given its advantages in speed and efficiency, the case-control design is often the first design choice in studies on the etiology of disease [1]. The case-control design is indispensable if the disease is rare or assessment of the exposure is expensive, and in situations where results are needed quickly to inform public health policy [2].

A crucial issue in case-control studies is the approach to identify cases and controls. A first consideration is whether cases are incident or prevalent. If cases are incident, a second consideration is whether cases and controls are from a fixed cohort or a dynamic population. In these circumstances the meaning of the odds ratio depends on the way in which controls were selected (from the population at risk at the beginning of follow-up, from the population free of disease at the end of follow-up, or from the person-time at risk), and on the underlying assumptions [3-7]. For example, much emphasis is often put on the need for the disease to be rare for the odds ratio to estimate the risk ratio if controls are sampled at the end of the follow up period from a fixed cohort. Depending on the nature of the cases, the type of source population, the sampling strategy and the underlying assumptions, the odds ratio obtained in a case-control study is interpreted as the risk ratio, rate ratio or prevalence odds ratio, or an odds ratio without such interpretation.

We did a survey of case-control studies recently published in leading general medicine, epidemiology and clinical specialist journals. We examined the methods and type of populations studied and assessed what was estimated by the odds ratio, and whether the rare disease assumption or other assumptions were important in this context.

Methods

Selection of articles

We examined case-control studies published in five general medicine journals (*Annals of Internal Medicine, BMJ, JAMA, Lancet, New England Journal of Medicine*), five general epidemiology journals (*American Journal of Epidemiology, Epidemiology, International Journal of Epidemiology, Journal of Clinical Epidemiology, Journal of Epidemiology and Community Health*), and ten clinical specialist journals (*American Journal of Respiratory and Critical Care Medicine, Archives of General Psychiatry, Arthritis and Rheumatism, Blood, Circulation, Clinical Infectious Diseases, Diabetes Care, Journal of American Geriatrics Society, Journal of the National Cancer Institute, Pediatrics*). We identified eligible studies in a literature search combining the journal names with the Medical Subject Heading (MeSH) term ‘case-control studies’. We selected 50 case-control studies from each of the three types

of journals, ten from each general medicine and epidemiology journal and five from each clinical specialist journal. We started in March 2007 with the most recently indexed items and went backwards in time until we identified 150 eligible studies. Articles that were published electronically ahead of print were included. We included original articles and short reports, but excluded letters and other editorial material. Articles that did not report any measure of association, and case-crossover studies were also excluded. The decision to include 150 studies was based on pragmatic considerations, rather than formal sample size calculations.

Definitions

Cases and controls can be selected from *fixed* cohorts (for example a birth cohort of people born in one calendar year) or from a *dynamic* population affected by births and deaths, immigration and emigration (for example the population of a city) [8]. These two concepts are also known as closed and open populations [7]. A *stable* population denotes a population in which the composition of the population, including the exposure distribution, does not change over time. A fixed population is by definition not stationary. Dynamic populations may be stable, and are likely to be stable over short time periods and for certain exposures, for example genetic factors.

Within fixed cohorts we distinguished three approaches to sample controls. First, controls can be selected from those who remain free of the disease at the *end* of the follow-up. This traditional case-control sampling design is also called *exclusive design* [6], *cumulative design* [3,7] or *cumulative incidence sampling* [3]. Second, controls can be selected at the *beginning* of the follow-up from the total study population at risk, which is also called *inclusive design* [6], *case-cohort study* [9] or *case-base study* [10]. Third, controls can be sampled concurrently with the cases, i.e. each time a new case is diagnosed a control is selected from the population at risk at that point in time. This means that controls are selected from the person-time at risk and controls are matched on time to the cases.

Within dynamic populations controls are often selected from the person-time at risk, which is also called *incidence density sampling* [3,11] or just *density sampling* [7]. This can be done by matching the controls on time (e.g. a case was diagnosed at the 5th of June 2006 and the corresponding control was randomly selected from the population that was at risk of becoming a case on the same day) or by assessing exposure in the control and case at the same point in time (e.g. controls were assigned index dates similar to the dates of diagnosis of their cases and exposure was assessed in a time window, for example 6 months before the index date). Another approach to sample controls from a dynamic population is to select

controls at some point in time, either at the end, at the beginning or during the period in which the cases are diagnosed (e.g. the cases were diagnosed between January 2003 and December 2005 and the controls were sampled from the population that was at risk of becoming a case in December 2005).

Interpretation of odds ratios

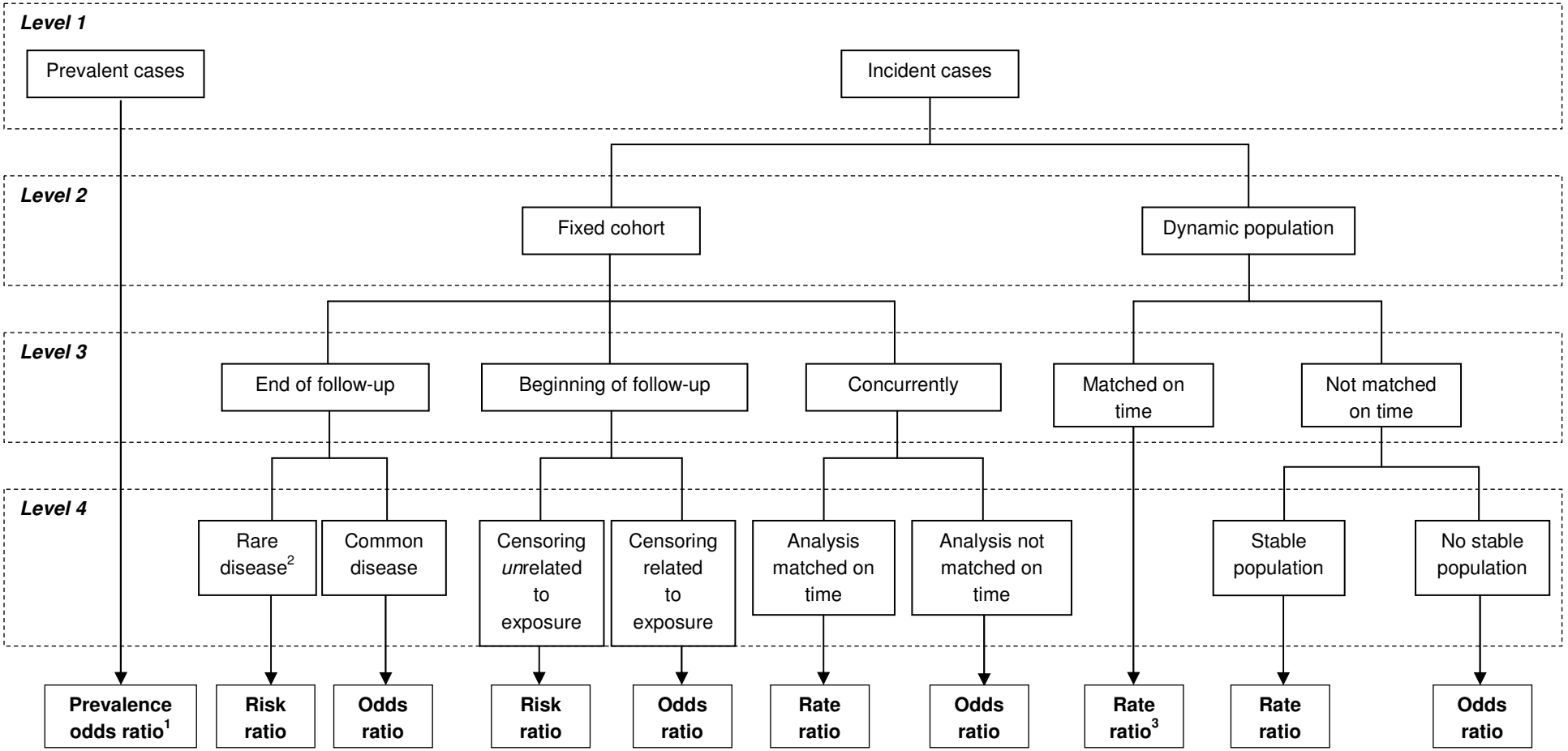
We developed a decision tree (Figure 1) to identify what is estimated by the odds ratio calculated from case-control studies, depending on the nature of the cases, the type of source population, the strategy used to select controls, and the underlying assumptions. If the cases are incident and controls are sampled at the end of the follow up period from a fixed cohort, the study will estimate the risk ratio when the assumption of a rare disease is met [4,6]. Sampling of controls at the beginning of the follow-up period in a fixed cohort will also estimate a risk ratio assuming that censoring is unrelated to exposure [4] (this assumption also applies to sampling at the end of follow-up [3], but for simplicity we focus on the rare disease assumption in that sampling scheme). The odds ratio from a case-control study that sampled the controls concurrently with the cases in a fixed cohort reflects the rate ratio if matching on time is taken into account in the analysis [4,6]. If the controls are sampled from a dynamic population and are matched on time (either sampled at the same time, or by using an index date), the odds ratio from a matched analysis estimates the rate ratio irrespective of whether the population is stable [3,4]. Of note, the impact of ignoring the matching in the analysis tends to be small unless exposures change substantially during the study period [3]. Conversely, if controls from a dynamic population are sampled at some point in time during case accrual, the source population needs to be stable in its exposure distribution, for the odds ratio to estimate the rate ratio [11].

If cases are prevalent, the odds ratio always equals the prevalence odds ratio. Its interpretation is a rate ratio if the duration of disease does not depend on exposure status, and a prevalence ratio if the disease is rare [11]. We have not pursued these distinctions or assessed them in the papers: studies based on prevalent cases were rare in our sample, and the first assumption relies on subject matter knowledge and is difficult to check.

Data extraction

A standardized data extraction form was used to assess the articles. Items covered included general items, such as journal name, year of publication, number of cases, number of controls, main exposure, condition studied; and specific items about the nature of cases (incident or prevalent), the type of source population, the sampling method, and the time

Figure 1 Decision tree to identify what is estimated by the odds ratio calculated from case-control studies, depending on the nature of the cases (prevalent or incident; level 1), the type of the source population (fixed cohort or dynamic population; level 2), sampling design to select controls (level 3), and the underlying assumptions (level 4)



[1] The prevalence odds ratio can be interpreted as rate ratio or prevalence ratio depending on assumptions (see methods)

[2] Strictly speaking the assumption that censoring is unrelated to exposure is also required when sampling at end of follow-up period (see methods)

[3] Strictly speaking this holds only if the analysis takes matching on time into account, although the impact of ignoring the matching tends to be small unless exposure trends are large

period in which cases and controls were sampled. The extraction form was piloted on six articles (two articles from each journal type) that were not included in the study and the form was modified. Two reviewers (MJK and PS) independently assessed all 150 articles. If authors referred to a previous paper for full description of the methods, information from this previous paper was used.

We defined rules on how to assess specific situations. Congenital diseases were always classified as prevalent. If incident and prevalent cases were included in one analysis, we classified the nature of cases as prevalent. If cases and controls were sampled from a fixed cohort and the controls were sampled among those who had equal or longer follow-up than the cases, we considered this as equivalent to sampling at the end of follow-up. For sampling from a dynamic population we distinguished two categories of unclear: 'unclear regarding time' meaning that it was not explicitly stated when controls were sampled in time (at beginning, end or during period of case selection), and 'unclear regarding source population' meaning that is was not clear whether the controls were sampled from the same population as cases.

For key items we calculated the initial agreement between data extracted by MJK and PS in percent and kappa values [12].

Data analysis

Frequencies and summary statistics were calculated, stratified by journal type. Differences in study characteristics between journal types were tested with the Fisher's exact test in case of proportions and the Kruskal Wallis test for non-normally distributed continuous variables. We used the decision tree shown in Figure 1 to assess what was estimated by the odds ratio.

Results

Our search gave 4647, 3351 and 6508 hits in the general medicine journals, the epidemiology journals and the clinical specialist journals, respectively. Based on this search we identified the 50 most recent eligible case-control studies for each journal type. The publication date of the selected articles ranged from May 2001 to March 2007 for studies published in general medicine journals (median November 2005), from October 2002 to March 2007 for studies published in general epidemiology journals (median April 2006), and from August 2004 to April 2007 for studies published in clinical specialist journals (median December 2006). Eleven (7%) of the 150 articles were short reports; five in general medicine journals, three in general epidemiology journals and three in clinical specialist journals.

The initial observed agreement and kappas between the two data extractors (MJK and PS) ranged from substantial to fair [12]: origin of cases, 76.7% agreement, κ=0.60; origin of controls,

Table 1 Characteristics of case-control studies included in survey by type of journal

	General medicine articles (n=50)	**General epidemiology articles** (n=50)	**Clinical specialist articles** (n=50)	**p-value**
Country of study participants				0.974
United States	17 (34%)	15 (30%)	18 (36%)	
Europe except UK	11 (22%)	14 (28%)	13 (26%)	
United Kingdom	8 (16%)	7 (14%)	5 (10%)	
Several countries (including either US, Europe or UK)	5 (10%)	3 (6%)	5 (10%)	
Other	9 (18%)	11 (22%)	9 (18%)	
Median number of cases (range)	494 (26-13,556)	611 (42-22,225)	282 (18-21,169)	0.031
Median number of controls (range)	846 (27-135,386)	1204 (85-180,220)	585 (20-423,128)	0.032
Source of cases				0.526
Population based	34 (68%)	34 (68%)	28 (56%)	
Hospital based	14 (28%)	15 (30%)	21 (42%)	
Both	1 (2%)	1 (2%)	0 (0%)	
Unclear	1 (2%)	0 (0%)	1 (2%)	
Source of controls				0.521
Population based	38 (76%)	37 (74%)	32 (64%)	
Hospital based	9 (18%)	8 (16%)	10 (20%)	
Both	2 (4%)	3 (6%)	2 (12%)	
Unclear	1 (2%)	2 (4%)	6 (4%)	
Exposures				<0.001
Drugs	20 (40%)	5 (10%)	6 (12%)	
Precursor disease states	3 (6%)	11 (22%)	11 (22%)	
Genetic factors	5 (10%)	1 (2%)	12 (24%)	
Environmental factors	0 (0%)	10 (20%)	4 (8%)	
Other*	22 (44%)	23 (46%)	17 (34%)	
Outcome category				0.001
Cardiovascular disease	18 (36%)	9 (18%)	11 (22%)	
Cancer	5 (10%)	23 (46%)	9 (16%)	
Infectious disease	12 (24%)	4 (8%)	12 (24%)	
Other**	15 (30%)	14 (28%)	19 (38%)	

* Includes two studies with two exposures (genetic factor and precursor disease state; genetic factor and other) ; ** Includes two studies with two outcome categories (both studies cardiovascular disease and cancer); P-values from Fisher's exact tests or Kruskal-Wallis tests

83.3% agreement, κ=0.68; nature of cases, 80.5% agreement, κ=0.37; type of source population, 81.9% agreement, κ=0.60; sampling design, 70.7% agreement, κ=0.54. The low agreement for nature of cases was due to disagreements on whether cases could be classified

as prevalent cases or whether this was unclear, and not because of disagreements on incident cases. Most discrepancies were resolved in discussions with the senior authors (JPV and ME).

Table 1 shows the characteristics of the case-control studies by type of journal. The number of cases and controls was highest in articles published in epidemiology journals and lowest in reports from clinical specialist journals. Drugs were the most commonly studied exposures, in particular in studies published in general medicine journals. Cardiovascular disease outcomes were mainly studied in general medicine journals, while cancer outcomes were common in epidemiology journals.

Table 2 presents the nature of cases included in studies and the source population and sampling method used to select controls. Based on this information the effect measure estimated by the odds ratio, conditional on assumptions, is also listed. Studies based on incident cases and a dynamic source population were most common, and particularly common among studies published in epidemiology journals. Among the 125 studies with incident cases, 105 (84% of 125) estimated a rate ratio: 48 studies (38% of 125) estimated a rate ratio without requiring any further assumption and 57 (46% of 125) estimated the rate ratio when assuming that the source population was a stable dynamic population. The stable population assumption might be more likely to be met for the studies with shorter duration of accrual of cases. Accrual was one year or less in nine studies (16% of 57), one to five years in 29 studies (51% of 57), five to ten years in ten studies (18% of 57), more than ten years in three studies (5% of 57), and unclear in six studies (11% of 57). Of the studies that sampled incident cases, a minority (18% of 125) sampled from a fixed cohort. In 17 of these studies (14% of 125), the estimated odds ratio reflected the risk ratio, with 16 requiring the rare disease assumption. Twelve studies (8% of 150) estimated a prevalence odds ratio, which can be interpreted as a rate ratio or prevalence ratio depending on assumptions not further considered in this study. In 16 studies (11% of 150), it was unclear what the odds ratio estimated. Ten of these studies were published in clinical specialist journals.

Table 3 compares the interpretation of odds ratio and assumptions required as determined in this study with measure of association reported and assumptions discussed by authors. Almost all studies (135 of 150, 90%) presented results as an odds ratio. Eighteen of these studies stated that the odds ratio was an approximation of the relative risk and two stated that their odds ratio was an unbiased estimate of the incidence rate ratio (see footnote to Table 3). Two studies inappropriately reported a rate ratio and one study inappropriately reported a risk ratio. Four studies discussed the rare disease assumption, but in none of these studies the rare disease assumption was required. None of the studies that needed a stable population for the odds ratio to estimate the rate ratios discussed this assumption.

Table 2 The interpretation of the odds ratio in published case-control studies according to the nature of cases (prevalent or incident), the type of source population (fixed cohort or dynamic population) and the sampling method used to select controls (end of follow-up, beginning of follow-up, concurrently, matched on time and not matched on time), stratified by type of journal

	General medicine articles n=50 (100%)	**General epidemiology articles** n=50 (100%)	**Clinical specialist articles** n=50 (100%)	**Total** n=150 (100%)	**Interpretation of odds ratio**	
					Effect measure	**Assumption to be met**
Incident cases	**44 (88%)**	**46 (92%)**	**35 (70%)**	**125 (83%)**		
Fixed cohort	*9 (18%)*	*3 (6%)*	*11 (22%)*	*23 (15%)*		
Sampling at end of follow-up	6	1	9	16	Risk ratio	Rare disease
Sampling at beginning of follow-up	0	0	1	1	Risk ratio	Censoring unrelated to exposure
Sampling concurrently	3	1	1	5*	Rate ratio	None
Unclear sampling	0	1	0	1	Unclear	-
Dynamic population	*35 (70%)*	*43 (86%)*	*24 (48%)*	*102 (68%)*		
Matched on time	17	17	9	43**	Rate ratio	None
Not matched on time	11	8	6	25	Rate ratio	Stable population
Unclear regarding time	7	17	8	32	Rate ratio	Stable population
Unclear regarding source population	0	1	1	2	Unclear	-
Prevalent cases	**5 (10%)**	**1 (2%)**	**6 (12%)**	**12 (8%)**	Prevalence odds ratio	None
Unclear nature of cases	**1 (2%)**	**3 (6%)**	**9 (18%)**	**13 (9%)**	Unclear	-

* All studies used an analysis matched on time; ** 32 studies used an analysis matched on time

Table 3 Interpretation of odds ratio and assumptions required as determined in this study versus measure of association reported and assumptions discussed by authors of original studies

Interpretation of odds ratio (assumption required)	**Total**	**Measure of association reported by authors**					**Assumption discussed by authors**		
		Odds ratio	Risk ratio	Rate ratio*	Relative risk	Likelihood ratio	Rare disease	Stable population	None
Rate ratio (population stable)	57 (100%)	56 (98%)[a]	0	1	0	0	1	0	56 (98%)
Rate ratio (none)	48 (100%)	38 (79%)[b]	1	6	3	0	2	0	46 (96%)
Risk ratio (disease rare)	16 (100%)	13 (81%)[c]	0	1	1	1	0	0	16 (100%)
Risk ratio (censoring unrelated to exposure)	1 (100%)	0 (0%)	0	1	0	0	0	0	1 (100%)
Prevalence odds ratio (none)	12 (100%)	12 (100%)[d]	0	0	0	0	1	0	11 (92%)
Unclear (unclear)	16 (100%)	16 (100%)[e]	0	0	0	0	0	0	16 (100%)
Total	150 (100%)	135 (90%)	1	9	4	1	4	0	146 (97%)

* Including incidence rate ratio and hazard ratio

[a] Six studies reported that the odds ratio could be interpreted as relative risk

[b] Six studies reported that the odds ratio could be interpreted as relative risk and two studies reported that the odds ratio could be interpreted as rate ratio

[c] Two studies reported that the odds ratio could be interpreted as relative risk

[d] Two studies reported that the odds ratio could be interpreted as relative risk

[e] Two studies reported that the odds ratio could be interpreted as relative risk

Discussion

This survey of 150 published case-control studies found that in most studies the odds ratio estimated the rate ratio, however, in a substantial proportion of these the assumption of a stable population was required. In contrast, the rare disease assumption was needed only in relatively few studies for the odds ratio to estimate the risk ratio. In most studies the authors reported odds ratios and very few investigators interpreted them as estimates of the risk or rate ratio or discussed the assumptions that may be required in this context.

The different sampling designs in case-control studies and their implication in terms of what is estimated by the odds ratio have been described in detail in the methodological literature [2-7,11,13], but we are not aware of another survey that examined the approaches actually used to select controls in published case-control research. A survey of epidemiological studies identified several issues of concern regarding the design, analysis and reporting of epidemiological research [14], but it did not address what the odds ratios estimated in case-control studies. Several assumptions need to be considered in this context. We found that the well known and extensively discussed rare disease assumption was needed in relatively few studies (16 of 150, 11%), whereas assuming that the exposure distribution was stable in the population over time was required in 57 (38%) studies. The underlying reason was that only relatively few studies sampled from fixed cohorts while about two third sampled from dynamic populations. Our results thus support the notion that the rare disease assumption is less important in case-control research than generally assumed. Greenland and Thomas [3] pointed out that the bias associated with a more frequent disease becomes substantial only when the cumulative incidence over the study period is greater than about 10% percent, which is uncommon in practice (although other figures have been reported in this context, ranging from 5% [6] to 20% [7]). In contrast, Greenland and Thomas showed that changes in the proportion of a dynamic population that is exposed can lead to biased estimates [3]. We did not check for each study included in our survey whether relevant assumptions were in fact met, i.e. that the disease was sufficiently rare, the population stable and censoring is unrelated to exposure. We considered this was not feasible because too little information was reported in the articles to reliably check these assumptions.

Our survey has other limitations. In 13 of the 150 studies (9%) the nature of cases remained unclear and it was not possible to determine what the odds ratio estimated, and whether certain assumptions were required. There may have been additional studies in this group requiring the rare disease assumption. Furthermore, the initial agreement between the two observers who extracted data was low for the nature of cases, although consensus

could generally be reached after discussion or consultation with a third reviewer. Our experience confirms the results of previous studies, which found that the reporting of important methodological aspects is often wanting in epidemiological studies [7,14-17]. For example, to decide whether cases were incident or prevalent cases, we often had to rely on a single word, such as 'consecutive', indicating incident cases. We sometimes also needed tacit knowledge about health care systems, for example when databases of health maintenance organisations were used to identify cases and controls. However, we refrained from second guessing and coded items as 'unclear' if the information provided was clearly insufficient.

We acknowledge that some case-control studies may have been missed by our search, which was exclusively based on the term 'case-control studies'. For example, we probably missed case-control studies that were not described as such by the authors and not indexed as case-control studies. These studies might well differ in relevant aspects from those included in our survey. Also, case-cohort studies may have been underrepresented in our study population although an additional, specific search for case-cohort studies in the journals and time periods selected for this study revealed that we may have missed only three such studies. We included only journals with high impact factors and the results of this study cannot therefore be applied to all journals that publish case-control research. Finally, case-control studies from general medicine and clinical specialist journals may have been overrepresented in our sample, which might have influenced combined results.

The most widely used case-control design involves sampling of controls from a dynamic population, which often needed the assumption of a stable population for the odds ratio to estimate a rate ratio. A stable population means that the exposure distribution of the controls does not change over time in this dynamic population. For example, genetic exposures tend to be more stable in populations than life style exposures. For many exposures it applies that the shorter the period over which cases accrued the more likely the population will be stable. Some environmental or life style exposures may, however, not be stable even over short periods of time and matching on time is advisable in these situations. In our survey, the interpretation of the odds ratio as a rate ratio required the stable population assumption in many studies, but this was not discussed in any of the articles.

Our study has implications for the reporting of case-control studies. The STROBE initiative (Strengthening the Reporting of Observational Studies in Epidemiology) recently published a checklist of items that should be addressed in reports of observational studies, including items that are specific to case-control studies [18,19]. Although the appropriate use and potential of the STROBE initiative is a matter of debate [20-23], we believe these

recommendations can help researchers to report more transparently on the nature of cases, the source population and methods used to select controls. Also, we and others [5] believe that investigators should report and discuss what measure of association is estimated by the odds ratio that is calculated from their case-control study. Our study also has important implications for teaching on case-control studies. In many basic textbooks, the need for the rare disease assumption is emphasized in sections covering case-control studies. However, this only concerns studies that sample controls at the end of the follow-up period in fixed cohorts, and our survey shows that this situation is rare in practice. In more advanced textbooks, the sampling of controls at the beginning of the follow-up period and concurrent sampling in fixed cohorts are sometimes also covered in detail, but in actual practice these situations are even less common. We examined 26 English language textbooks of epidemiology from the Medical School Library in Utrecht, and from our personal and institutional libraries (see list in Appendix). We found that 16 textbooks did not mention any assumption about the odds ratio in case-control studies (eight books) or only the rare disease assumption (eight books). Two textbooks vaguely referred to different modes of sampling, and only eight discussed the different sampling methods in fixed cohorts and dynamic populations in more detail.

In conclusion, since the majority of case-control studies sample from a dynamic population, and since most studies seem to rely on the assumption of a stable population, this type of sampling and the need for (discussing) the stability assumption should be emphasized in textbooks. Finally, we hope our survey will alert authors to the need for complete and transparent reporting of the strategies used to select cases and controls as well as to the need for a discussion of what measure of association is estimated by the odds ratio calculated from case-control studies.

References

1. Rosendaal FR. Bridging Case-Control Studies and Randomized Trials. *Curr Control Trials Cardiovasc Med* 2001;2:109-10.
2. Schlesselman JJ. *Case-control Studies. Design, Conduct, Analysis.* Oxford University Press, New York, 1982.
3. Greenland S and Thomas DC. On the Need for the Rare Disease Assumption in Case-Control Studies. *Am J Epidemiol* 1982;116:547-53.
4. Greenland S, Thomas DC, Morgenstern H. The Rare-Disease Assumption Revisited. A Critique of "Estimators of Relative Risk for Case-Control Studies". *Am J Epidemiol* 1986;124:869-83.
5. Pearce N. What Does the Odds Ratio Estimate in a Case-Control Study? *Int J Epidemiol* 1993;22:1189-92.
6. Rodrigues L and Kirkwood BR. Case-Control Designs in the Study of Common Diseases: Updates

on the Demise of the Rare Disease Assumption and the Choice of Sampling Scheme for Controls. *Int J Epidemiol* 1990;19:205-13.

7. Rothman KJ and Greenland S. *Modern Epidemiology.* Lippincott-Raven Publishers, Philadelphia, PA, 1998.
8. *A Dictionary of Epidemiology.* University Press, Oxford, 2001.
9. Prentice RL. A Case-Cohort Design for Epidemiologic Studies and Disease Prevention Trials. *Biometrika* 1986;73:1-11.
10. Miettinen O. Design Options in Epidemiologic Research. An Update. *Scand J Work Environ Health* 1982;8 Suppl 1:7-14.
11. Miettinen O. Estimability and Estimation in Case-Referent Studies. *Am J Epidemiol* 1976;103:226-35.
12. Landis JR and Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 1977;33:159-74.
13. Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of Controls in Case-Control Studies. I. Principles. *Am J Epidemiol* 1992;135:1019-28.
14. Pocock SJ, Collier TJ, Dandreo KJ, De Stavola BL, Goldman MB, Kalish LA et al. Issues in the Reporting of Epidemiological Studies: a Survey of Recent Practice. *BMJ* 2004;329:883.
15. Lee W, Bindman J, Ford T, Glozier N, Moran P, Stewart R et al. Bias in Psychiatric Case-Control Studies: Literature Survey. *Br J Psychiatry* 2007;190:204-9.
16. Tooth L, Ware R, Bain C, Purdie DM, Dobson A. Quality of Reporting of Observational Longitudinal Research. *Am J of Epidemiol* 2005;161:280-8.
17. von Elm E and Egger M. The Scandal of Poor Epidemiological Research. *BMJ* 2004;329:868-9.
18. Vandenbroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *Epidemiology* 2007;18:805-35.
19. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. *Ann Intern Med* 2007;147:573-7.
20. Egger M, Altman DG, Vandenbroucke JP. Commentary: Strengthening the Reporting of Observational Epidemiology the STROBE Statement. *Int J Epidemiol* 2007;36:948-50.
21. Macmahon B and Weiss NS. Is There a Dark Phase of This STROBE? *Epidemiology* 2007;18:791.
22. Kuller LH and Goldstein BD. Suggestions for STROBE Recommendations. *Epidemiology* 2007;18:792-3.
23. Ebrahim S and Clarke M. STROBE: New Standards for Reporting Observational Epidemiology, a Chance to Improve. *Int J Epidemiol* 2007;36:946-8.

Appendix

First author	Title	Year
Ahlbom, A	Introduction to modern epidemiology	1984
Austin, DF	Epidemiology for the health sciences: a primer on epidemiologic concepts and their uses	1982
Barker, DJP	Epidemiology in medical practice	1998
Clayton, D	Statistical models in epidemiology	1993
Feinstein, AR	Clinical epidemiology: the architecture of clinical research	1985
Fletcher, RH	Clinical epidemiology: the essentials	1996
Gordis, L	Epidemiology	2000
Gordis, L	Epidemiology and health risk assessment	1988
Hennekens, CH	Epidemiology in medicine	1987
Kahn, HA	An introduction to epidemiologic methods	1983
Kelsey, JL	Methods in observational epidemiology	1986
Kleinbaum, DG	Epidemiologic research	1982
Lilienfeld, DE	Foundations of epidemiology	1994
MacMahon, B	Epidemiology: principles and methods	1996
Miettinen, OS	Theoretical epidemiology: principles of occurrence research in medicine	1985
Morton, RF	A study guide to epidemiology and biostatistics	1996
Norell, SE	Workbook of epidemiology	1995
Olsen, J	Teaching epidemiology: what you should know and what you should do	1992
Rothman, KJ	Epidemiology: an introduction	2002
Rothman, KJ	Modern epidemiology	1998
Sackett, DL	Clinical epidemiology: a basic science for clinical medicine	1991
Schlesselman, JJ	Case-control studies: design, conduct, analysis	1982
Szklo, M	Beyond the basics	2003
Timmreck, TC	An introduction to epidemiology	1994
Unwin, N	An introductory study guide to public health and epidemiology	1997
Weiss, NS	Clinical epidemiology: the study of the outcome of illness	1996

Chapter 4

General discussion

In this general discussion we will critically review the literature on the association between depression and diabetes. We will summarize the results of all longitudinal studies we are aware of that investigated this association. We will discuss several methodological issues of these studies and examine whether there is evidence for causality. First, the cross-sectional association between depression and diabetes will be briefly discussed. Then, we will review whether depression is a risk factor for diabetes and subsequently whether diabetes is a risk factor for depression. Finally, we will discuss the implications of our findings and give recommendations for future studies.

Prevalence of depression in diabetes patients

A meta-analysis from 2001 identified 20 cross-sectional studies that compared the prevalence of depressive disorder or depressive symptoms in subjects with and without type 1 and type 2 diabetes [1]. That study found a two-fold increased prevalence of depression in diabetes patients. However, the meta-analysis did not evaluate the possibility of confounding, so it remains possible that this association could be explained by characteristics that differed between diabetes patients and subjects without diabetes and that were related to depression status. Furthermore, only five studies used a diagnostic interview to assess depression and none of these studies measured glucose levels to establish diabetes.

A more recent meta-analysis of cross-sectional studies investigated the prevalence of depression or depressive symptoms specifically in patients with type 2 diabetes and included ten studies [2]. Only five of the 20 studies included in the previous meta-analysis were also included in this recent meta-analysis because the recent meta-analysis was restricted to patients with type 2 diabetes and included only studies with at least 100 participants. The more recent meta-analysis found a 1.8-fold increased prevalence of depression in type 2 diabetes patients compared to those without. The majority of these studies, however, did not adjust for differences in demographic and lifestyle factors. Three studies performed a diagnostic interview to diagnose depression, but none of these used glucose measurements to assess type 2 diabetes. Two of these studies found a small non-significant increased risk of diabetes among depressed persons [3,4], while the other study found a five-fold increased risk [5].

While there is evidence for an increased prevalence of depression among diabetes patients, these results might be due to demographic, socioeconomic and lifestyle differences between subjects with and without diabetes and therefore do not imply a causal association. In addition, none of the studies assessed depression and diabetes using the current

diagnostic standards, i.e. a diagnostic interview for depressive disorder and glucose measurement for diabetes. Moreover, cross-sectional studies typically cannot determine the temporality of an association; to do so longitudinal studies are needed.

Depression as a risk factor for diabetes

We performed a meta-analysis summarizing nine longitudinal studies on depression and the incidence of type 2 diabetes (Chapter 2.1 of this thesis [6], [3,7-14]). This meta-analysis found a 37% increased risk of type 2 diabetes among depressed subjects compared with non-depressed. Since the conduct of our meta-analysis, four additional longitudinal studies investigating depression as a risk factor for diabetes have been published (Chapter 2.2 of this thesis [15], [16-18]). Updating our meta-analysis with these studies using a random effects model resulted in a pooled relative risk of 1.26 (95% CI: 1.11-1.44). Figure 1 presents the relative risks and 95% confidence intervals of these thirteen studies. Table 1 presents characteristics of these studies concerning confounders, interaction, selection bias and information bias. We will address these methodological issues in the following section and we will discuss whether there is evidence that depression is a causal risk factor for diabetes.

Figure 1 Relative risks and 95% confidence interval of all longitudinal studies investigating depression as a risk factor for diabetes

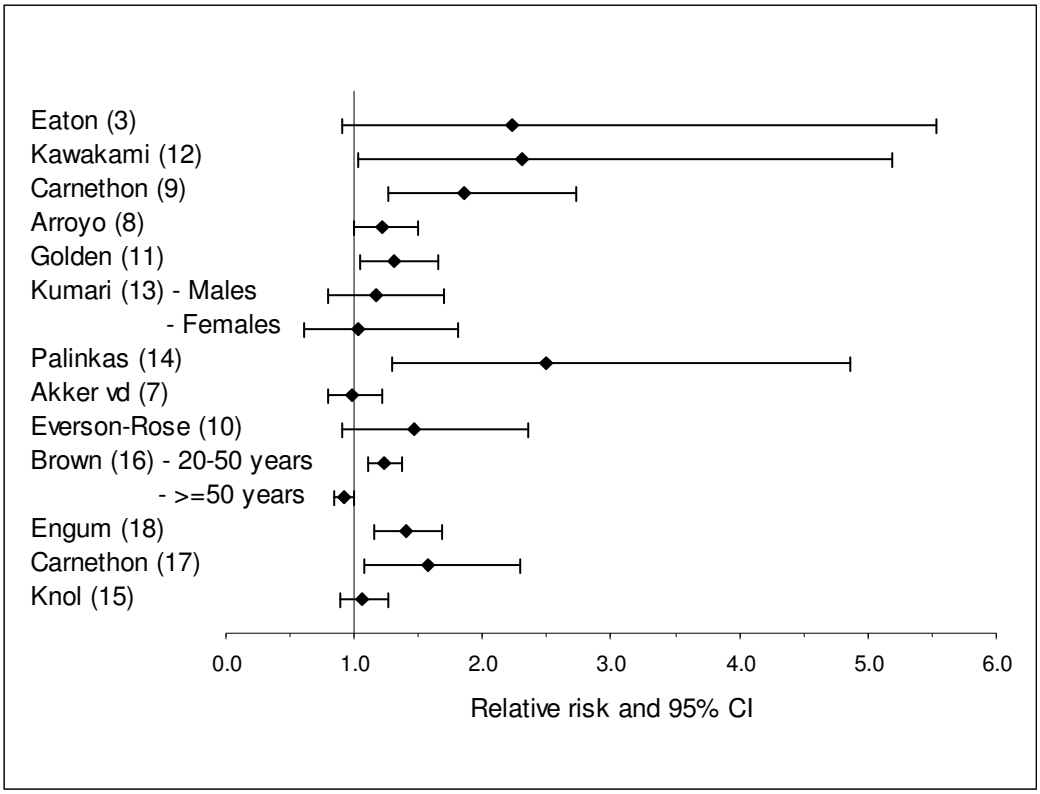


Table 1 Characteristics of longitudinal studies investigating depression as a risk factor for diabetes

First author (reference)	**Year**	**Males / females**	**Relative risk (95% CI)**	**Confounders**				**Interaction**		
				SES	*BMI*	*Life-style*	*Chronic diseases*	*Potential effect modifier*	*Strati-fication*	*Testing*
Eaton (3)	1996	649 / 1066	2.23 (0.90-5.54)	+	+	-	-			
Kawakami (12)	1999	2380 / 0	2.31 (1.03-5.19)	+	+	+	cardiovascular disease			
Carnethon (9)	2003	2858 / 3332	1.86 (1.27-2.72)	-	+	+	-	sex, education	+	+
Arroyo (8)	2004	0 / 72,178	1.22 (1.00-1.49)	-	+	+	-	BMI, smoking, alcohol, history dm	+	+
Golden (11)	2004	5183 / 6432	1.31 (1.04-1.65)	+	+	+	-			
Kumari (13)	2004	5773 / 2547	males: 1.17 (0.8-1.7) females: 1.03 (0.6-1.8)	+	+	+	-	sex	+	-
Palinkas (14)	2004	418 / 553*	2.50 (1.29-4.86)	-	+	+	-	sex	-	+
Akker vd (7)	2004	33,186 / 34,818	0.98 (0.79-1.21)	+	+	-	-	age, sex	+	-
Everson-Rose (10)	2004	0 / 2662	1.46 (0.90-2.36)	+	+	+	-	race	+	+
Brown (16)	2005	Total: 47255 / 45412	20-50 years: 1.23 (1.10-1.37) >=50 years: 0.92 (0.84-1.00)	-	-	-	physician visits	age	+	+
Engum (18)	2007	17599 / 19692*	1.40 (1.16-1.69)	+	+	+	several diseases	sex, type of diabetes	+	+
Carnethon (17)	2007	1910 / 2771	1.57 (1.07-2.29)	+	+	+	-			
Knol (15)	2007	25,477 / 35,039	1.06 (0.89-1.26)	-	-	-	chronic disease score	chronic disease score	+	-

* before excluding diabetes patients at baseline

** cut off points based on literature

*** cut off points based on quartiles

**** cut off point based on factor analysis

***** cut off point based on validation within study

Table 1 continued

First author	**Selection bias**			**Information bias**			
(reference)	*Loss to follow-up*	*Incident depression*	*Handling missing data*	*Depression assessment*	*Categories of depression*	*Diabetes assessment at baseline*	*Diabetes assessment at follow-up*
Eaton (3)	1572/3481=45.2%	-	complete case	DIS	2	self-report	self-report
Kawakami (12)	384/2764=14%	-	complete case	Zung SDS	3**	self-report	glucose
Carnethon (9)	1024/14,407=7%	-	?	GWB-DS	3**	self-report	self-report
Arroyo (8)	?	-	complete case	MHI-5	2**	self-report	self-report
Golden (11)	?	-	complete case	VES	4***	glucose	glucose
Kumari (13)	24%	-	complete case	GHQ-D	2****	self-report	glucose
Palinkas (14)	1195/2375=50%	-	complete case	BDI	2**	glucose	glucose
Akker vd (7)	censoring	+	?	doctor's diagnosis	2	doctor's diagnosis	doctor's diagnosis
Everson-Rose (10)	?	-	complete case	CES-D	2**	glucose	glucose
Brown (16)	censoring	-	?	doctor's diagnosis	2	doctor's diagnosis	doctor's diagnosis
Engum (18)	37,706/74,997=50%	-	complete case	ADI	2*****	self-report	self-report
Carnethon (17)	?	-	imputation	CES-D	2**	glucose, medication use	glucose, medication use
Knol (15)	censoring	+	complete case	medication use	2	medication use	medication use

* before excluding diabetes patients at baseline

** cut off points based on literature

*** cut off points based on quartiles

**** cut off point based on factor analysis

***** cut off point based on validation within study

Confounding

The studies differed considerably regarding adjustment for potential confounders (Table 1). All studies adjusted for age and sex; mostly by including age and sex as covariates in the logistic regression model, sometimes by restricting the sample to male or female participants, and one study by stratifying for sex. Most studies adjusted for one of the three groups of confounders: socioeconomic status or education level, body mass index (BMI) or waist circumference, and lifestyle factors such as physical activity, smoking or alcohol consumption. Only three studies adjusted for one or more comorbid chronic diseases: cardiovascular disease [12]; cardiovascular disease, thyroid disease, musculoskeletal disease, cancer and asthma [18]; and chronic disease score [15]. One study adjusted for physician visits as a proxy for comorbidity [16]. The only study that adjusted for age, sex, socio-economic status, BMI, lifestyle factors as well as several chronic diseases found an effect estimate of 1.4 [18].

We did not observe a consistent pattern between adjustments for certain confounders and the size of the effect estimates.

Interaction

Several studies examined whether the strength of the association between depression and the incidence of diabetes was modified by certain characteristics, including age, sex, education, race, BMI, smoking, alcohol consumption, type of diabetes, parental history of diabetes and comorbid chronic diseases (Table 1). Almost all studies that examined interaction presented the effect estimate in strata of the suspected effect modifier and tested for interaction (Table 1). Significant interaction on a multiplicative scale was found for race, where the odds ratio of depression and onset of diabetes was higher in African Americans than in Caucasians [10], and for age, where the effect estimate was higher in subjects aged 20 to 50 years than in subjects older than 50 years [16]. However, when we plotted the mean age of the study population of each study against the relative risk of each study there was no association between age and the size of the relative risk (data not shown). None of the studies provided enough information to examine interaction between depression and a certain factor on an additive scale (see Chapter 3.2 [19] and Chapter 3.3 of this thesis).

Selection bias

Selection bias is a distortion of the association under study that occurs if the selection of study participants is related to both exposure and outcome. In case-control studies selection bias occurs if the selection of subjects with and without the outcome depends on the

exposure status. Only one of the studies investigating depression as a risk factor for diabetes was a case-control study [16]. This study selected both cases and controls from an administrative database, i.e. source population, and assessed exposure status similarly in cases and controls, thereby minimizing the chance of selection bias. As cases and controls came from a dynamic population and they were matched on time, the estimated odds ratio in this study can be interpreted as a rate ratio (see Chapter 3.5 of this thesis).

In cohort studies selection bias occurs if the selection of participants with and without the exposure depends on the outcome status. Although selective recruitment is usually not considered to be an issue in cohort studies because the outcome is measured later on, we provide a compelling example of potential selective recruitment that was related to both exposure and outcome (Chapter 3.4 of this thesis). Studies that use a registry or administrative database will not suffer from selective participation into the study as all eligible study participants are automatically and anonymously included into the study. In contrast, studies that actively recruit participants are prone to selective response which results in selection bias if response is related to both exposure and outcome. Most studies recruited study participants from a certain population, while three studies used a registry or administrative database [7,15,16]. The studies performed within an administrative database found effect estimates closer to one than studies that recruited their study participants specifically for the study.

Only two studies included subjects with incident instead of prevalent depression (Table 1) by taking first doctor's diagnosis of depression [7] and first use of antidepressants [15]. This may be important as subjects with incident depression will be more comparable to subjects without depression than to subjects with prevalent depression regarding covariates. Selecting prognostically more comparable groups is better than adjusting for differences in the analyses because of the chance of residual confounding in the latter option. The two studies that included more comparable groups found effect estimates close to one.

Another source of selection bias in cohort studies is selective loss to follow-up, although some epidemiology textbooks consider selective loss to follow-up to be information bias [20]. The percentage of participants that were lost to follow-up ranged from 7 to 50%, while in four studies is was unclear how many participants were lost (Table 1). The three studies that used registration databases [7,15,16] had continuous information on the outcome and censored the participants at the date they were lost to follow-up.

If data on exposure, outcome or covariates is not missing completely at random, performing complete case analysis can lead to biased effect estimates (Chapter 3.1 of this thesis, [21,22]). Nine of the thirteen studies performed complete case analysis and in three

studies it was not clear if there were missing data and how they were handled (Table 1). One study reported to have imputed missing values but it was not specified how this was done.

Selective recruitment, selection of subjects with prevalent depression, selective loss to follow-up and missing data may have resulted in biased effect estimates. We observed that studies that probably suffered least from these potential selection biases found effect estimates closer to one [7,15].

Information bias

For the assessment of depression, a diagnostic interview is seen as the gold standard. Other methods to determine the presence of depression are measurement of depressive symptoms, doctor's diagnosis or use of antidepressants. Depressive symptoms can be measured by a wide variety of, mostly self-report, questionnaires. A disadvantage of using depressive symptoms as a measure for depression is that they might not be specific enough for depression. In addition, self-report questionnaires might rather measure diabetes related problems such as fatigue and changed eating behaviour than depression itself. A study compared patient characteristics with and without depressive disorder, assessed by the CIDI, and with and without high depressive symptoms, assessed by the CES-D, among 506 diabetes patients [23]. They showed that the CES-D, compared with the CIDI, seemed to be more sensitive to the stresses associated with chronic health conditions and socioeconomic factors rather than to depressive disorder itself. A disadvantage of using doctor's diagnosis of depression is that depression cases might be missed because they have not been diagnosed. Using antidepressant use to measure depression is also not ideal because not all depressed subjects are treated with antidepressants and not all antidepressant users have clinically diagnosed depression [24]. However, using doctor's diagnosis and antidepressant use might detect at least more severe and debilitating depression. Only one study of the thirteen longitudinal studies assessed depression with a diagnostic interview [3]. Most studies used self-report questionnaires, although many different questionnaires were used. Two studies used doctor's diagnosis to assess depression [7,16] and one used antidepressant use [15]. Studies that used a diagnostic interview or a questionnaire to assess depression found higher effect estimates than studies using doctor's diagnosis or antidepressant use.

Three of the studies that used a questionnaire to assess depression constructed more than two categories on the basis of the depressive symptom score [9,11,12] and presented the relative risk comparing the groups with the lowest and highest group as the main effect. Analyzing these data differently, e.g., on a continuous scale or with two categories, might have led to lower effect estimates. Most of the studies that assessed depressive symptoms

with a self-reported questionnaire based their cut off point on literature. One study used factor analysis to determine the cut off point [13], one study performed a validation study within their study to determine the cut off point [18], and one study defined quartiles of the depressive symptom score [11].

In all studies, diabetes was assessed at baseline to exclude prevalent diabetes and at follow-up to detect incident diabetes cases. Diabetes should ideally be assessed by measuring fasting blood glucose twice within one week or by performing an oral glucose tolerance test. Four of the thirteen studies assessed diabetes at baseline by measuring glucose levels, meaning that these studies also excluded undetected diabetes at baseline (Table 1). This decreases the chance of reversed causality. At follow-up, six studies assessed diabetes by means of glucose levels, whereas seven studies had to rely on self-report, doctor's report, or medication use (Table 1). These latter studies will have identified only diagnosed diabetes patients and will have missed undetected diabetes patients. As diabetes can be undetected for a long period, reversed causality cannot be excluded in these studies, meaning that diabetes could have been already present when depression status was measured and depression can therefore be a consequence of diabetes instead of vice versa.

Studies using doctor's diagnosis or medication use as a measure of depression or diabetes are prone to a specific form of information bias, namely detection bias. It is possible that diabetes is more likely to be detected in depressed subjects than in non-depressed subjects as depressed subjects will visit the doctor more often. The opposite can also be true: diabetes is less likely to be detected in depressed subjects as the diabetes symptoms or complaints are ascribed to the depression. The first possibility would give an overestimation of the association between depression and diabetes incidence while the second possibility would give an underestimation of the association. The studies using doctor's diagnoses or medication use consistently showed effect estimates close to one (Table 1).

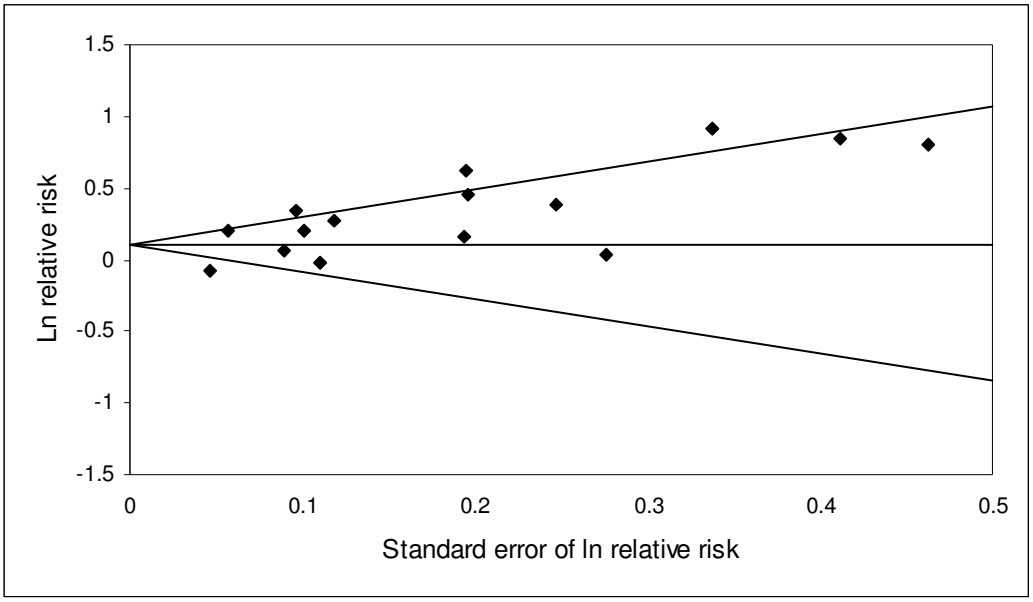
None of the thirteen studies assessed depressive disorder and diabetes at baseline and follow-up using the current diagnostic standards. In studies using a self-reported questionnaire the size of the effect estimates might depend on the number and definition of categories. On the other hand, the studies that used doctor's diagnosis or medication use to assess depression and diabetes might be biased due to selective detection. Reversed causality could not be excluded in most studies as these studies did not assess undetected diabetes at baseline and follow-up.

Publication bias

Publication bias occurs if studies with a significant result are more likely to be published than

studies that did not find a significant association. A method to examine the presence of publication bias is to plot the standard error of the natural logarithm of the effect estimate against the natural logarithm of the effect estimate itself. The studies should then be equally distributed around the line depicting the pooled effect estimate of the fixed effects model. Figure 2 shows such a plot where each dot represents one of the thirteen longitudinal studies that investigated depression as a risk factor for diabetes. This graph shows that the studies are not well distributed around the line of the pooled effect estimate. Small studies with a relatively small effect estimate are missing, which is an indication of publication bias. The rank correlation test of Begg [25] for publication bias gave a p-value of 0.30, whereas the weighted regression test of Egger [26] for publication bias gave a p-value smaller than 0.01; the latter test, which is said to have more power to detect bias [20,27], indeed indicated publication bias.

Figure 2 Examination of publication bias for depression as a risk factor for diabetes



Most of the studies were performed within large cohort studies and some within registration databases [7,15,16]. These studies were not specifically designed to study this research question and therefore the results are more prone to publication bias. Using existing data is a relatively easy and inexpensive approach to study a research question and therefore the decision to submit it might depend on the results of the study. Moreover, most editors will rather publish a statistically or clinically significant finding as these studies

have more impact than a null result. Since a few years all randomized clinical trials have to be registered to reduce the possibility of publication bias [28,29]. For observational studies, however, it is less feasible to oblige registration of study questions. As an alternative, it has been suggested to let papers be reviewed without the results section but with only an introduction, methods and strengths and limitations section [30,31]. With this approach, editors and reviewers can assess the paper on criteria that a paper ought to be assessed on, namely on the importance of the research question and the quality of the study methods. Once a paper has been accepted the author would be asked to return it with the results included.

Publication bias indeed seems to play a role in the literature on depression as a risk factor for diabetes.

Evidence for causality

If depression is a causal risk factor for diabetes, we would expect to observe a dose-response relation, i.e. increasing incidence of diabetes with increasing severity of depression or increasing episodes of depression. None of the longitudinal studies that categorized depressive symptoms in more than two categories, found a clear dose-response relation between severity of depressive symptoms and incidence of diabetes [9,11,12]. Another study found a similar hazard ratio of diabetes in subjects with two consecutive high scores of depressive symptoms as in subjects with a single high score of depressive symptoms [17]. That study also found no dose-response relation when quartiles of the depressive symptom score were studied [17]. Use of doctor's diagnosis or antidepressant use to assess depression could reflect more severe depression and therefore we would expect higher effect estimates in studies using these methods. However, the opposite is true, studies using doctor's diagnosis and antidepressant use found effect estimates closer to one.

The ultimate study design to proof that depression is a causal risk factor for type 2 diabetes would be a randomized clinical trial where subjects are assigned to depression or no depression and followed up until diabetes occurs. Of course it is not possible to randomize subjects to having a depression. However, within a group of diabetes patients it is possible to assign active depression treatment or placebo depression treatment and monitor glycemic control and diabetes complications. Five randomized controlled trials have been performed within diabetes patients, studying the effect of depression treatment on diabetes outcomes [32-36]. While treatment significantly improved depression outcomes in all studies, diabetes outcomes did not significantly improve. Only two of the five studies showed a larger, although not significant, decrease of HbA1c in the treatment group compared with the placebo group [33,34]. The number of subjects in these studies ranged from 13 to 152 and

all studies used different antidepressants. Although observational, our study also showed no evidence of an effect of antidepressant use on glycemic control within diabetes patients using insulin (Chapter 2.3 of this thesis).

Lack of evidence of a dose-response relation between depression and incidence of diabetes, and no proven improvement of diabetes outcomes in randomized depression treatment trials makes it less likely that depression is a causal risk factor for diabetes.

In summary

The longitudinal studies that investigated depression as a risk factor for diabetes are heterogeneous with regard to study characteristics and the observed strength of the association. All studies had methodological advantages and disadvantages.

For example, two of the three studies with the relative risk closest to one adjusted for comorbid chronic diseases, all those studies used a registration or administrative database and therefore did not suffer from selective response, and two of the three studies assessed incident rather than prevalent depression. However, two of these three studies did not adjust for socioeconomic status, body mass index and lifestyle factors, and all three studies suffered from an increased risk for detection bias. Of the three studies with a relative risk furthest away from one, all studies adjusted for socioeconomic status, BMI, or lifestyle factors, and diabetes was assessed by glucose measurements in two of the three studies. However, all these studies suffered from considerable loss to follow-up, performed complete case analysis and had the potential for selective response.

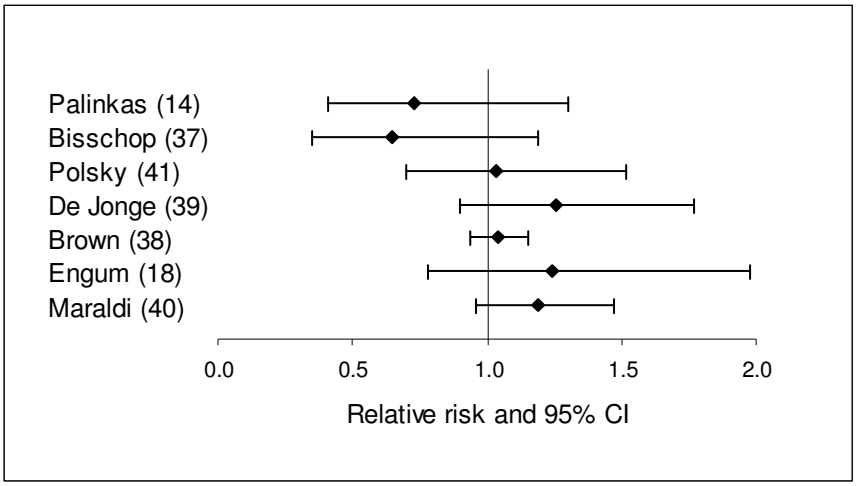
We are not able to say which potential biases are most important and therefore we cannot draw a definite conclusion on whether depression is a risk factor of diabetes and, if so, what the strength of the association is.

Diabetes as a risk factor for depression

Seven studies investigated the incidence of depression in diabetes patients and subjects without diabetes [14,18,37-41]. None of these studies found a significantly increased incidence of depression among diabetes patients. Relative risks ranged from 0.7 to 1.3 and combining these studies using a random effects model gave a pooled relative risk of 1.07 (95% CI: 0.96-1.19). Our longitudinal study showed that the incidence of antidepressant use was only increased shortly after (one to three months) the start of diabetes treatment (Chapter 2.5 of this thesis). Figure 3 presents the relative risks and 95% confidence intervals of the seven studies. Table 2 presents characteristics of these studies concerning confounders, interaction, selection bias, information bias and publication bias. We will discuss these

methodological issues in the following section and we will discuss whether there is evidence that diabetes is a causal risk factor for depression.

Figure 3 Relative risks and 95% confidence interval of all longitudinal studies investigating diabetes as a risk factor for depression



Confounding

All studies adjusted for age and sex, while one study adjusted for body mass index and three studies adjusted for some lifestyle factors including physical activity, smoking and alcohol consumption (Table 2). Four of the seven studies adjusted for one or more chronic diseases. We could not identify a pattern between adjustments for certain confounders and the size of the effect estimates.

Interaction

One study investigated whether sex modified the association between diabetes and depressive symptoms by testing the interaction on a multiplicative scale and found no statistically significant interaction [14]. Another study observed no association between diabetes and incident depression in subjects without other chronic somatic diseases, while there was an association between diabetes and depression incidence in subjects with other somatic diseases [39]. However, the interaction was not statistically significant on a multiplicative scale. In both studies, the authors did not provide enough information to examine interaction on an additive scale (see Chapter 3.2 [19] and Chapter 3.3 of this thesis).

Table 2 Characteristics of longitudinal studies investigating diabetes as a risk factor for depression

First author (reference)	**Year**	**Males / females**	**Relative risk (95% CI)**	**Confounders**			**Interaction**
				BMI	*Lifestyle*	*Chronic diseases*	*Potential effect modifier*
Palinkas (14)	2004	418 / 553*	0.73 (0.41-1.30)	+	+	-	sex
Bisschop (37)	2004	1064 / 1214*	0.65 (0.35-1.19)	-	-	+	
Polsky (41)	2005	4043 / 4344	1.03 (0.70-1.52)	-	-	?	
De Jonge (39)	2006	2011 / 2746*	1.26 (0.90-1.77)	-	+	+	somatic diseases, disability
Brown (38)	2006	46010 / 42766	1.04 (0.94-1.15)	-	-	+	
Engum (18)	2007	17599 / 19692*	1.24 (0.78-1.98)	-	-	-	
Maraldi (40)	2007	1270 / 1252	1.19 (0.96-1.47)	-	+	+	

* before excluding depressed patients at baseline

Table 2 continued

First author (reference)	**Selection bias**			**Information bias**		
	Incident diabetes	*Handling missing data*	*Loss to follow-up*	*Diabetes assessment*	*Depression assessment*	*Follow-up measurements*
Palinkas (14)	-	complete case	1195/2375=50%	glucose	BDI	8 years
Bisschop (37)	+	complete case	562/3107=18%	self-report	CES-D	3 and 6 years
Polsky (41)	+	?	399/9772=4%	self-report	CES-D	2 years
De Jonge (39)	-	complete case	2400/4803=50%	self-report	diagnosis	2 and 5 years
Brown (38)	+	?	censoring	doctor's diagnosis	doctor's diagnosis	continuous over 8 years
Engum (18)	-	complete case	37706/74997=50%	self-report	HADS	10 years
Maraldi (40)	-	complete case	42/2522=2%	glucose	CES-D	2, 3, 5 and 6 years

Selection bias

Three of the seven studies used incident cases of diabetes [37,38,41]. The selection of incident cases is preferable over the selection of prevalent cases as the exposed and unexposed group are more comparable with respect to other characteristics.

Five studies used complete case analysis to handle their missing data, while in the other two studies it was unclear whether there were missing data and how they were handled (Table 2). The percentage of loss to follow-up ranged from 2 to 50% (Table 2). The study that used a registration database [38] had continuous information on the outcome and censored the participants at the date they were lost to follow-up.

Information bias

Two studies [14,40] used glucose measurements to assess diabetes and therefore also included undiagnosed diabetes patients, which reduces the chance of reversed causality. Four studies asked the study participants themselves whether they had diabetes and one study relied on doctor's diagnosis and medication use to assess diabetes [38]. The two studies that used glucose measurements did not find lower or higher relative risks than the other studies.

Depression was assessed by a diagnostic interview in one study, in five studies by self-reported questionnaires and in one study by doctor's diagnosis and medication use (Table 2). No pattern between the methods of depression assessment and the size of the effect estimates was observed.

Most studies assessed depression many years after the diagnosis of diabetes or after the assessment of prevalent diabetes (Table 2). None of the studies assessed the incidence of depression directly after the diagnosis of diabetes. A period of two years was the shortest period over which the incidence of depression was calculated [41]. One study assessed the incidence of depression continuously over time but only gave an overall effect estimate over a time period of eight years and did not calculate time-specific incidence rates [38]. We determined the incidence of antidepressant use in intervals of months after the initiation of diabetes treatment and found an increased incidence of antidepressant use within the three months after the diagnosis of diabetes (Chapter 2.5 in this thesis).

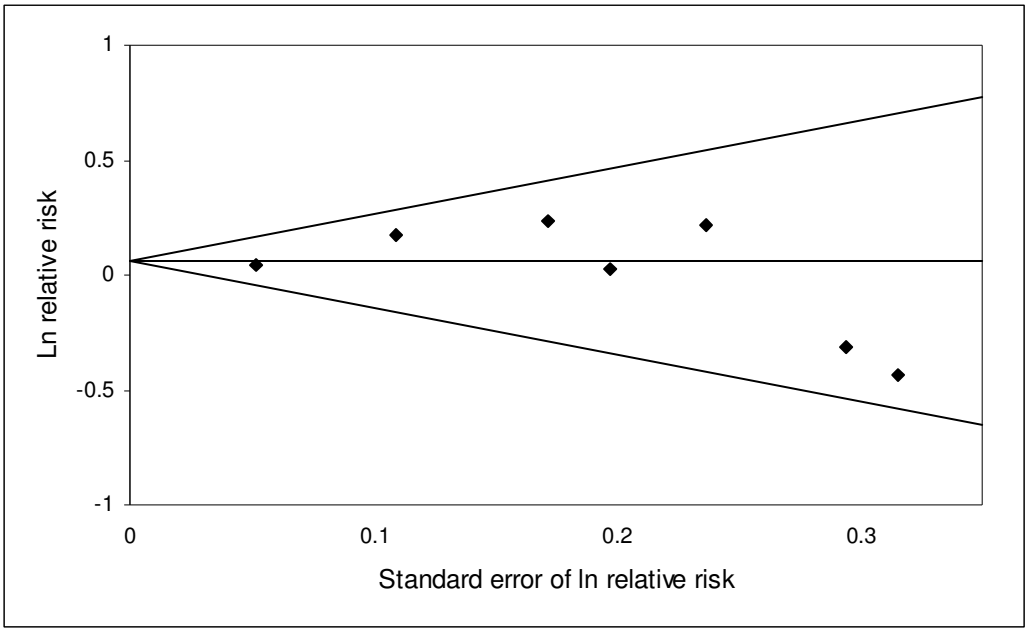
In studies using doctor's diagnosis or medication use, detection bias can occur as depression might be more likely or less likely to be detected in diabetes patients than in subjects without diabetes. However, the study using doctor's diagnosis [38] found a similar effect estimate as studies using other methods to assess diabetes and depression. Detection bias might explain the results we found in our study on antidepressant use after the initiation of diabetes treatment (Chapter 2.5 in this thesis).

None of the studies assessed diabetes as well as depression using the current diagnostic standards, i.e. a diagnostic interview for depressive disorder and glucose measurement for diabetes. Except for our study, none of the studies assessed the trend over time in depression incidence. Detection bias is a potential limitation of studies using doctor's diagnosis or medication use.

Publication bias

We plotted the standard error of the natural logarithm of the effect estimate against the natural logarithm of the effect estimate for all seven studies to examine the potential for publication bias (Figure 4). The figure shows that the studies are quite symmetrically distributed around the line indicating the pooled effect estimate according to the fixed effects model, meaning there is no evidence of publication bias. The Begg's test [25] for publication bias gave a p-value of 0.13, while Egger's test [26] for publication bias gave a p-value of 0.71, both also indicating no evidence for publication bias.

Figure 4 Examination of publication bias for diabetes as a risk factor for depression



Evidence for causality

Presence of a dose-response relationship would strengthen the idea that diabetes is a causal risk factor for depression. However, most studies investigating the relation between impaired glucose tolerance, a pre-cursor of diabetes, or levels of fasting glucose levels and

depressive symptoms did not find a significant association (Chapter 2.4 of this thesis [42], [14,43,44]). In addition, we found an association between diagnosed diabetes and depressive symptoms, but not between undiagnosed diabetes and depressive symptoms (Chapter 2.4 of this thesis [42]). On the other hand, within diabetes patients micro vascular complications and worse glycemic control have been shown to be associated with depressive symptoms [40,45,46]. There is no consistent evidence for a dose-response relationship between diabetes and the incidence of depression and therefore it is unlikely that diabetes is a causal risk factor for depression.

In summary

The longitudinal studies investigating diabetes as a risk factor for depression are quite heterogeneous with regard to study characteristics but not with regard to the observed strength of the association. There seems to be no increased risk of depression in diabetes patients after a period of at least two years after the diagnosis. There might be an increased risk of depression or depressive symptoms shortly after the diagnosis of diabetes due to psychological distress which reduces again after a certain adaptation period.

Conclusions, implications, and future studies

Depression as a risk factor for diabetes

The available evidence does not allow a firm conclusion on whether depression is a risk factor for diabetes. If depression is a risk factor for diabetes, the relative risk is rather small, but even a small relative risk can impact many individuals as both depression and diabetes are common diseases. If depression is not a causal risk factor for diabetes, it may still be a prognostic factor meaning that patients with a depression may be more likely to have diabetes due to other factors. Adequate detection and treatment of depression, however, will in that case not help to prevent the onset of diabetes.

Future studies should not only measure depressive symptoms but also depressive disorder. At baseline, preferably only incident depression cases are included. Incidence of diabetes should be assessed according to diagnostic standards, which are currently the measurement of two fasting glucose levels within one week. Ideally, depression and diabetes status are assessed at baseline and every following year for at least four years. Potential confounders that should be taken into account are age, sex, socioeconomic status or education level, body mass index, lifestyle factors and comorbid chronic diseases. A study design like this hopefully gives a definitive answer to the question whether depression is a risk factor for diabetes.

Diabetes as a risk factor for depression

After at least two years after the diagnosis, diabetes is not a risk factor for depression. Despite the lack of evidence of a causal association, diabetes and depression may still co-occur in patients due to other factors. In addition, the incidence of depression might be increased shortly after the diagnosis of diabetes due to psychological distress.

Future studies should particularly investigate at what point in time the risk of depression might be increased after a diagnosis of diabetes. To study this, it is important to monitor depression continuously. This can be achieved by using a registration database with continuous records of diagnoses and medication use. However, these databases also have limitations such as the imperfect method of the assessment of depression and the potential for detection bias. Comorbid chronic diseases should be taken into account as a confounder or effect modifier. More knowledge on when the risk of depression is high after a diagnosis of diabetes, might direct specific monitoring of depression in diabetes patients.

References

1. Anderson RJ, Freedland KE, Clouse RE, Lustman PJ. The Prevalence of Comorbid Depression in Adults With Diabetes: a Meta-Analysis. *Diabetes Care* 2001;24:1069-78.
2. Ali S, Stone MA, Peters JL, Davies MJ, Khunti K. The Prevalence of Co-Morbid Depression in Adults With Type 2 Diabetes: a Systematic Review and Meta-Analysis. *Diabet Med* 2006;23:1165-73.
3. Eaton WW, Armenian H, Gallo J, Pratt L, Ford DE. Depression and Risk for Onset of Type II Diabetes. A Prospective Population-Based Study. *Diabetes Care* 1996;19:1097-102.
4. Thomas J, Jones G, Scarinci I, Brantley P. A Descriptive and Comparative Study of the Prevalence of Depressive and Anxiety Disorders in Low-Income Adults With Type 2 Diabetes and Other Chronic Illnesses. *Diabetes Care* 2003;26:2311-7.
5. Saeed AK and Al-Dabbagh TQ. Type 2 Diabetes and Its Association With Hypertension and Depression in an Iraqi Population. *Ann Saudi Med* 2003;23:254-9.
6. Knol MJ, Twisk JW, Beekman AT, Heine RJ, Snoek FJ, Pouwer F. Depression As a Risk Factor for the Onset of Type 2 Diabetes Mellitus. A Meta-Analysis. *Diabetologia* 2006;49:837-45.
7. Akker van den M, Schuurman A, Metsemakers J, Buntinx F. Is Depression Related to Subsequent Diabetes Mellitus? *Acta Psychiatr Scand* 2004;110:178-83.
8. Arroyo C, Hu FB, Ryan LM, Kawachi I, Colditz GA, Speizer FE et al. Depressive Symptoms and Risk of Type 2 Diabetes in Women. *Diabetes Care* 2004;27:129-33.
9. Carnethon MR, Kinder LS, Fair JM, Stafford RS, Fortmann SP. Symptoms of Depression As a Risk Factor for Incident Diabetes: Findings From the National Health and Nutrition Examination Epidemiologic Follow-Up Study, 1971-1992. *Am J Epidemiol* 2003;158:416-23.
10. Everson-Rose SA, Meyer PM, Powell LH, Pandey D, Torrens JI, Kravitz HM et al. Depressive Symptoms, Insulin Resistance, and Risk of Diabetes in Women at Midlife. *Diabetes Care* 2004;27:2856-62.

11. Golden SH, Williams JE, Ford DE, Yeh HC, Paton SC, Nieto FJ et al. Depressive Symptoms and the Risk of Type 2 Diabetes: the Atherosclerosis Risk in Communities Study. *Diabetes Care* 2004;27:429-35.
12. Kawakami N, Takatsuka N, Shimizu H, Ishibashi H. Depressive Symptoms and Occurrence of Type 2 Diabetes Among Japanese Men. *Diabetes Care* 1999;22:1071-6.
13. Kumari M, Head J, Marmot M. Prospective Study of Social and Other Risk Factors for Incidence of Type 2 Diabetes in the Whitehall II Study. *Arch Intern Med* 2004;164:1873-80.
14. Palinkas LA, Lee PP, Barrett-Connor E. A Prospective Study of Type 2 Diabetes and Depressive Symptoms in the Elderly: the Rancho Bernardo Study. *Diabet Med* 2004;21:1185-91.
15. Knol MJ, Geerlings MI, Egberts AC, Gorter KJ, Grobbee DE, Heerdink ER. No Increased Incidence of Diabetes in Antidepressant Users. *Int Clin Psychopharmacol* 2007;22:382-6.
16. Brown LC, Majumdar SR, Newman SC, Johnson JA. History of Depression Increases Risk of Type 2 Diabetes in Younger Adults. *Diabetes Care* 2005;28:1063-7.
17. Carnethon MR, Biggs ML, Barzilay JI, Smith NL, Vaccarino V, Bertoni AG et al. Longitudinal Association Between Depressive Symptoms and Incident Type 2 Diabetes Mellitus in Older Adults: the Cardiovascular Health Study. *Arch Intern Med* 2007;167:802-7.
18. Engum A. The Role of Depression and Anxiety in Onset of Diabetes in a Large Population-Based Study. *J Psychosom Res* 2007;62:31-8.
19. Knol MJ, van der Tweel I, Grobbee DE, Numans ME, Geerlings MI. Estimating Interaction on an Additive Scale Between Continuous Determinants in a Logistic Regression Model. *Int J Epidemiol* 2007;36:1111-8.
20. Hennekens CH, Buring JE. *Epidemiology in Medicine.* Little, Brown and company, Boston/Toronto, 1987.
21. Greenland S and Finkle WD. A Critical Look at Methods for Handling Missing Covariates in Epidemiologic Regression Analyses. *Am J Epidemiol* 1995;142:1255-64.
22. Schafer JL and Graham JW. Missing Data: Our View of the State of the Art. *Psychol Methods* 2002;7:147-77.
23. Fisher L, Skaff MM, Mullan JT, Arean P, Mohr D, Masharani U et al. Clinical Depression Versus Distress Among Patients With Type 2 Diabetes: Not Just a Question of Semantics. *Diabetes Care* 2007;30:542-8.
24. Gardarsdottir H, Heerdink ER, van DL, Egberts AC. Indications for Antidepressant Drug Prescribing in General Practice in the Netherlands. *J Affect Disord* 2007;98:109-15.
25. Begg CB and Mazumdar M. Operating Characteristics of a Rank Correlation Test for Publication Bias. *Biometrics* 1994;50:1088-101.
26. Egger M, Davey SG, Schneider M, Minder C. Bias in Meta-Analysis Detected by a Simple, Graphical Test. *BMJ* 1997;315:629-34.
27. Sterne JAC, Egger M, Davey Smith G. *Investigating and Dealing with Publication and Other Biases.* In *Systematic Reviews in Health Care: Meta-analysis in Context*. (Ed. Egger M, Davey Smith G and Altman DG) BMJ publishing group, London, 2001.
28. Abaid LN, Grimes DA, Schulz KF. Reducing Publication Bias Through Trial Registration. *Obstet Gynecol* 2007;109:1434-7.

29. De Angelis C, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R et al. Clinical Trial Registration: a Statement From the International Committee of Medical Journal Editors. *Lancet* 2004;364:911-2.
30. Lawlor DA. Quality in Epidemiological Research: Should We Be Submitting Papers Before We Have the Results and Submitting More Hypothesis-Generating Research? *Int J Epidemiol* 2007;36:940-3.
31. Walster GW and Cleary TA. A Proposal for a New Editorial Policy in the Social Sciences. *Am Stat* 1970;24:16-9.
32. Lustman PJ, Griffith LS, Clouse RE, Freedland KE, Eisen SA, Rubin EH et al. Effects of Nortriptyline on Depression and Glycemic Control in Diabetes: Results of a Double-Blind, Placebo-Controlled Trial. *Psychosom Med* 1997;59:241-50.
33. Lustman PJ, Freedland KE, Griffith LS, Clouse RE. Fluoxetine for Depression in Diabetes: a Randomized Double-Blind Placebo-Controlled Trial. *Diabetes Care* 2000;23:618-23.
34. Paile-Hyvarinen M, Wahlbeck K, Eriksson JG. Quality of Life and Metabolic Status in Mildly Depressed Women With Type 2 Diabetes Treated With Paroxetine: a Single-Blind Randomised Placebo Controlled Trial. *BMC Fam Pract* 2003;4:7.
35. Lustman PJ, Clouse RE, Nix BD, Freedland KE, Rubin EH, McGill JB et al. Sertraline for Prevention of Depression Recurrence in Diabetes Mellitus: a Randomized, Double-Blind, Placebo-Controlled Trial. *Arch Gen Psychiatry* 2006;63:521-9.
36. Paile-Hyvarinen M, Wahlbeck K, Eriksson JG. Quality of Life and Metabolic Status in Mildly Depressed Patients With Type 2 Diabetes Treated With Paroxetine: a Double-Blind Randomised Placebo Controlled 6-Month Trial. *BMC Fam Pract* 2007;8:34.
37. Bisschop MI, Kriegsman DM, Deeg DJ, Beekman AT, van Tilburg W. The Longitudinal Relation Between Chronic Diseases and Depression in Older Persons in the Community: the Longitudinal Aging Study Amsterdam. *J Clin Epidemiol* 2004;57:187-94.
38. Brown LC, Majumdar SR, Newman SC, Johnson JA. Type 2 Diabetes Does Not Increase Risk of Depression. *CMAJ* 2006;175:42-6.
39. de Jonge P, Roy JF, Saz P, Marcos G, Lobo A. Prevalent and Incident Depression in Community-Dwelling Elderly Persons With Diabetes Mellitus: Results From the ZARADEMP Project. *Diabetologia* 2006;49:2627-33.
40. Maraldi C, Volpato S, Penninx BW, Yaffe K, Simonsick EM, Strotmeyer ES et al. Diabetes Mellitus, Glycemic Control, and Incident Depressive Symptoms Among 70- to 79-Year-Old Persons: the Health, Aging, and Body Composition Study. *Arch Intern Med* 2007;167:1137-44.
41. Polsky D, Doshi JA, Marcus S, Oslin D, Rothbard A, Thomas N et al. Long-Term Risk for Depressive Symptoms After a Medical Diagnosis. *Arch Intern Med* 2005;165:1260-6.
42. Knol MJ, Heerdink ER, Egberts AC, Geerlings MI, Gorter KJ, Numans ME et al. Depressive Symptoms in Subjects With Diagnosed and Undiagnosed Type 2 Diabetes. *Psychosom Med* 2007;69:300-5.
43. Lawlor DA, Smith GD, Ebrahim S. Association of Insulin Resistance With Depression: Cross Sectional Findings From the British Women's Heart and Health Study. *BMJ* 2003;327:1383-4.
44. Lawlor DA, Ben Shlomo Y, Ebrahim S, Davey SG, Stansfeld SA, Yarnell JW et al. Insulin Resistance and Depressive Symptoms in Middle Aged Men: Findings From the Caerphilly Prospective Cohort

Study. *BMJ* 2005;330:705-6.

45. de Groot M, Anderson R, Freedland KE, Clouse RE, Lustman PJ. Association of Depression and Diabetes Complications: a Meta-Analysis. *Psychosom Med* 2001;63:619-30.

46. Lustman PJ, Anderson RJ, Freedland KE, de Groot M, Carney RM, Clouse RE. Depression and Poor Glycemic Control: a Meta-Analytic Review of the Literature. *Diabetes Care* 2000;23:934-42.

Chapter 5

Summary

Samenvatting

Dankwoord

Curriculum vitae

Depression and diabetes are both common conditions in today's society. They also often co-occur as is shown in several cross-sectional studies. The reason for this co-occurrence is unclear. It could be due to chance; depression may be a risk factor for diabetes, or diabetes may be a risk factor for depression. The first aim of this thesis was to investigate whether depression is a risk factor for depression and whether diabetes is a risk factor for diabetes.

Methodological issues, such as confounding, interaction, selection bias and information bias, are important and commonly encountered topics in etiologic research. Potential biases play a substantial role in the interpretation of an effect estimate, especially when the strength of the 'true' effect estimate is small (relative risk between 0.5 and 2), which is probably the case for both directions of the association between depression and diabetes. The second aim of this thesis was to discuss several methodological issues in etiologic research in order to improve the conduct and reporting of this type of research.

Chapter 2 describes five studies with different designs and in different settings to investigate depression as a risk factor for diabetes and diabetes as a risk factor for depression.

Chapter 2.1 provides a quantitative summary of published longitudinal studies that investigated whether depression is a risk factor for type 2 diabetes. Medline and PsycInfo were searched up to January 2005 for articles that examined the relationship between depression and the incidence of type 2 diabetes. Nine studies were included in this meta-analysis. The pooled relative risk was 1.37 (95% CI: 1.14-1.63) using a random effects model. Heterogeneity between studies could not be explained by (1) whether studies controlled for undetected diabetes at baseline, (2) the method of diabetes assessment at follow-up, (3) the baseline risk of diabetes in the study population, and (4) follow-up duration. We concluded that depressed adults have a 37% increased risk of developing type 2 diabetes mellitus.

In **chapter 2.2** we used the pharmacy registration database PHARMO to investigate the association between depression and incidence of diabetes, where we used antidepressant and benzodiazepine use as proxies for depression and psychosocial complaints, respectively. A total of 60,516 individuals were followed from their first prescription for antidepressants or benzodiazepines until a first prescription for antidiabetic drugs or end of registration in the database. After adjustment for age, sex and chronic diseases, the hazard ratios for diabetes mellitus were 1.05 (95% CI: 0.88-1.26) for antidepressant but no benzodiazepine users, 1.21 (95% CI: 1.02-1.43) for benzodiazepine but no antidepressant users and 1.37 (95% CI: 1.12-1.68) for antidepressant and benzodiazepine users compared with no antidepressant and no benzodiazepine users. As the risk of diabetes was not increased in antidepressant but no benzodiazepine users, this study does not confirm the association

between depression and diabetes that we found in the meta-analysis.

In **chapter 2.3** we investigated the association between antidepressant use and glycemic control, measured as the amount of insulin used, within diabetes patients. Diabetes patients with and without an episode of antidepressant use were selected from the pharmacy registry database PHARMO. The amount of insulin used did not change during or after an episode of antidepressant use compared with before, while the amount of insulin used in diabetes patients without an antidepressant episode increased with 16% over a period of two years ($p<0.001$). SSRI (selective serotonin reuptake inhibitors) users showed a decrease of 13% in amount of insulin used during the antidepressant episode ($p=0.029$), while no change was seen in TCA (tricyclic antidepressants) users. The tendency for a difference in amount of insulin used between SSRIs and TCAs is suggestive for a pharmacologic effect of antidepressants rather than a general effect of depression on glycemic control.

Chapter 2.4 describes a study on the association between type 2 diabetes and the risk of depressive symptoms, using data from the Utrecht Health Project which is an ongoing longitudinal study among all inhabitants of a new residential area of a large city in The Netherlands. We assessed depressive symptoms in subjects with normal fasting glucose, impaired fasting glucose, undiagnosed type 2 diabetes (fasting glucose > 7.0 mmol/l) and diagnosed type 2 diabetes. The prevalence of depressive symptoms was increased in patients with diagnosed type 2 diabetes while it was not increased in subjects with impaired fasting glucose or undiagnosed type 2 diabetes. These results suggest that depressive symptoms are a consequence of the burden of type 2 diabetes.

In **chapter 2.5** we assessed the use of antidepressants and benzodiazepines before and after the initiation of diabetes treatment to investigate the temporal association between diabetes and depression and between diabetes and psychosocial complaints. The incidence of antidepressant use and benzodiazepine use was increased two months before and three months after the initiation of diabetes treatment compared with non-diabetes subjects. The strongest increase in incidence of antidepressant and benzodiazepine use was seen in the month after the initiation of diabetes treatment with incidence rate ratios of 2.4 (95% CI: 2.0-3.0) and 3.4 (95% CI: 3.0-3.8), after adjustment for age, sex and chronic diseases. The short time period between incidence of antidepressant and benzodiazepine use and initiation of diabetes treatment makes it unlikely that depression and diabetes are causally related. A plausible explanation is that depression and psychosocial complaints are more likely to be detected and treated if patients present with diabetes-related complaints to a physician or, vice versa, that diabetes is more likely to be detected if patients present with depression or psychosocial complaints.

In **chapter 3** we discussed several methodological issues in etiologic research.

In **chapter 3.1** we assessed the degree and direction of bias when comparing two methods of handling missing confounder data (missing indicator method and complete case analysis) with multiple imputation. From the PREDICT-NL study we selected an exposure (marital status), an outcome (depressive disorder), and three confounder variables (age, sex and income). We created missing values in the confounder 'income' according to different patterns of missingness. We showed that one should not use the missing indicator method to handle missing confounder data, because it gives a biased estimation of the odds ratio of exposure even with small percentages of missing values. More importantly, the direction of the bias is unpredictable. Complete case analysis can be used when missing values are missing completely at random. This, however, is hardly ever the case. Moreover, complete case analysis always leads to loss of statistical power. Multiple imputation gives unbiased effect estimates when missing values are missing completely at random and when they depend on observed data and can therefore be used in most situations with missing confounder data.

In **chapter 3.2** we illustrated with data from the Utrecht Health Project how interaction on an additive scale can be estimated when considering continuous determinants and a dichotomous outcome. From existing papers we derived the methods and formulas to estimate interaction and its confidence interval. We showed how these formulas can be used and how the results can be interpreted using an example with age and body mass index as risk factors for hypertension. The methods and formulas presented in this chapter are intended to assist epidemiologists to calculate interaction on an additive scale.

Chapter 3.3 shows how interaction was examined and reported in 225 published cohort and case-control studies. A majority of the studies addressed possible interactions between exposures (61%). However, in about half of them the information provided was unsatisfactory, and only one in ten studies reported sufficient data to interpret interaction effects on an additive and multiplicative scale. This could be done in two ways: presentation of the individual effect estimates and the joint effect estimate using one reference category or presentation of the full model, i.e. the individual effect estimates and the effect estimate of the product term. This study showed that efforts are required to promote more appropriate reporting on interaction.

The possibility of selection bias in a waiting room population was illustrated with empirical data from the PREDICT-NL study in **chapter 3.4**. The initial aim was to study the association between diabetes and depression in a population of general practice attendees. Unexpectedly, the prevalence of depression was much lower in diabetes patients (3.6%)

than in those without (14.1%; odds ratio: 0.23, 95% CI: 0.07-0.73). We subsequently investigated whether these unexpected findings could be explained by differential selection of subjects who visited their general practitioner (GP) for a control visit. Indeed, the percentage of subjects visiting the GP for a control visit was higher in diabetes patients than in patients without diabetes (odds ratio: 2.79, 95% CI: 1.78-4.36), while the percentage of subjects visiting the GP for a control visit was lower in subjects with depression than in those without (odds ratio: 0.64, 95% CI: 0.44-0.94). This study showed that the lower prevalence of depression among diabetes patients could at least in part be explained by differential selection based on the reason for visiting the GP.

In **chapter 3.5** we extracted data of 150 published case-control studies on the method of selecting cases and controls to examine how the reported odds ratio could be interpreted. In 17 studies (11%) the odds ratio could be interpreted as a risk ratio, with 16 of those requiring the rare disease assumption. In 57 studies (38%) a stable population was required for the odds ratio to estimate the rate ratio, while 48 studies (32%) estimated a rate ratio without needing any assumption. A prevalence odds ratio was estimated in 12 studies (8%) and in 16 studies (11%) insufficient information was reported to state how the odds ratio could be interpreted. The rare disease assumption was inappropriately discussed in four studies; no study mentioned the need for a stable population. We concluded that the assumption of a stable population is a much more frequently encountered issue in current case-control research practice than the rare disease assumption. In addition, researchers rarely discuss what odds ratios from case-control studies estimate.

Chapter 4 provides a general discussion of the current literature including the studies presented in this thesis on whether depression is a risk factor for diabetes and on whether diabetes is a risk factor for depression (thereby giving an update of the meta-analysis presented in Chapter 2.1). The available evidence was summarized and methodological issues of the studies were discussed.

The longitudinal studies that investigated depression as a risk factor for diabetes were heterogeneous with regard to study characteristics and the observed strength of the association. All studies had methodological advantages as well as disadvantages. Since it is difficult to say which potential biases have the largest influence on the results, we cannot draw a firm conclusion on whether depression is a risk factor of diabetes and, if so, what the strength of the association is. Future, well-designed, studies might give a definitive answer to the question whether depression is a risk factor for diabetes.

The longitudinal studies investigating diabetes as a risk factor for depression are quite

heterogeneous with regard to study characteristics but not with regard to the observed strength of the association. There seems to be no increased risk of depression in diabetes patients after a period of at least two years after the diagnosis. There might be an increased risk of depression shortly after the diagnosis of diabetes. Future studies should particularly investigate at what point in time the risk of depression may be increased after a diagnosis of diabetes. This might direct specific monitoring of depression in diabetes patients.

Chapter 5

Summary

Samenvatting

Dankwoord

Curriculum vitae

Dwarsdoorsnede onderzoeken hebben laten zien dat depressie en diabetes vaak samen voorkomen bij één individu. De reden voor dit samen voorkomen is onbekend. Het kan toeval zijn; het kan ook zijn dat depressie een risico factor is voor het ontwikkelen van diabetes of dat diabetes een risico factor is voor het ontwikkelen van een depressie. Het eerste doel van dit proefschrift was om deze beide relaties te onderzoeken.

Belangrijke en veelvoorkomende methodologische aspecten in etiologisch onderzoek, ofwel onderzoek naar causale risico factoren van ziekten, zijn: confounding (verstoring door externe factoren), interactie (wederzijdse samen- of tegenwerking van twee of meer factoren), selectie bias (vertekening door fouten in de selectie van de onderzoekspopulatie) en informatie bias (vertekening door fouten in metingen van de onderzoeksvariabelen). Mogelijke bias speelt een belangrijke rol bij de interpretatie van een effect schatter, met name als de sterkte van het effect klein is, wat waarschijnlijk het geval is voor de relatie tussen depressie en diabetes. Het tweede doel van dit proefschrift was om verschillende methodologische aspecten van etiologisch onderzoek te bespreken om de uitvoering en rapportage van etiologisch onderzoek te verbeteren.

Hoofdstuk 2 beschrijft vijf onderzoeken naar de relatie tussen depressie en het ontstaan van diabetes en de relatie tussen diabetes en het ontstaan van depressie, waarbij gebruik is gemaakt van verschillende studie opzetten en onderzoekspopulaties.

Hoofdstuk 2.1 geeft een kwantitatieve samenvatting van gepubliceerde longitudinale onderzoeken waarin onderzocht werd of depressie een risico factor is voor type 2 diabetes. Twee literatuur databases, Medline en PsycInfo, werden doorzocht tot januari 2005 op onderzoeken naar de relatie tussen depressie en het ontstaan van type 2 diabetes. Negen onderzoeken werden geïncludeerd in de meta-analyse. Het gepoolde relatieve risico was 1,37 (95% BI: 1,14-1,63), berekend met behulp van een ‘random effects model’. Heterogeniteit tussen de onderzoeken kon niet worden verklaard door (1) het al dan niet uitsluiten van onontdekte diabetes patiënten aan het begin van de studie, (2) de methode waarop diabetes was gemeten aan het einde van de studie, (3) het achtergrondrisico van diabetes in de studie populatie, en (4) het aantal jaren dat de deelnemers gevolgd werden. In dit onderzoek toonden we aan dat volwassenen met een depressie een 37% hoger risico hebben op het ontwikkelen van type 2 diabetes dan volwassenen zonder een depressie.

In **hoofdstuk 2.2** hebben we gebruik gemaakt van een registratie database met afleverdata van meerdere apotheken, PHARMO, om de relatie tussen depressie en incidentie van diabetes te onderzoeken. Daarbij werd het gebruik van antidepressiva als proxy voor depressie en het gebruik van benzodiazepines als proxy voor meer algemene

psychosociale klachten genomen. In totaal werden 60.516 individuen gevolgd vanaf hun eerste prescriptie voor antidepressiva of benzodiazepines tot aan hun eerste prescriptie voor diabetesmedicatie of tot aan het einde van registratie in de database. Na correctie voor leeftijd, geslacht en chronische ziekten waren de hazard ratio's voor het krijgen van diabetes 1,05 (95% BI: 0,88-1,26) voor antidepressiva gebruikers die geen benzodiazepines gebruikten, 1,21 (95% BI: 1,02-1,43) voor benzodiazepine gebruikers die geen antidepressiva gebruikten en 1,37 (95% BI: 1,12-1,68) voor mensen die antidepressiva zowel als benzodiazepines gebruikten in vergelijking met mensen die geen antidepressiva en geen benzodiazepines gebruikten. Bij mensen die alleen antidepressiva gebruikten was het risico op diabetes dus niet verhoogd. Deze studie bevestigt daarmee niet de eerder gevonden associatie tussen depressie en diabetes in de meta-analyse.

In **hoofdstuk 2.3** onderzochten we, binnen een groep diabetes patiënten, de relatie tussen antidepressiva gebruik en glycemische controle, waarbij glycemische controle werd gemeten als de hoeveelheid insuline die iemand gebruikte. In de PHARMO database selecteerden we diabetes patiënten met en zonder een episode van antidepressiva gebruik. De hoeveelheid insuline die patiënten gebruikten veranderde niet tijdens of na een episode van antidepressiva gebruik ten opzichte van de periode daarvoor, terwijl bij diabetes patiënten zonder een episode van antidepressiva gebruik de hoeveelheid gebruikte insuline toenam met 16% over een periode van twee jaar ($p<0,001$). Bij patiënten die SSRI's (selectieve serotonerge heropname remmers) gebruikten daalde de hoeveelheid gebruikte insuline met 13% tijdens een episode van antidepressiva gebruik ($p=0,029$), terwijl mensen die TCA's (tricyclische antidepressiva) gebruikten geen verandering in hoeveelheid insulinegebruik lieten zien. Dit mogelijke verschil tussen SSRI en TCA gebruikers maakt het waarschijnlijker dat antidepressiva een farmacologisch effect hebben dan dat de depressie invloed heeft op glycemische controle.

Hoofdstuk 2.4 beschrijft een studie naar de relatie tussen type 2 diabetes en het risico op depressieve symptomen, waarbij gebruik werd gemaakt van data van het Leidsche Rijn Gezondheidsproject, een longitudinaal onderzoek onder inwoners van een nieuwe woonwijk in Utrecht. Depressieve symptomen werden gemeten bij mensen met een normale nuchtere glucose waarde, met een verstoorde nuchtere glucose waarde, met ongediagnosticeerde type 2 diabetes (nuchtere glucose waarde > 7,0 mmol/l) en met zelfgerapporteerde gediagnosticeerde type 2 diabetes. Het risico op depressieve symptomen was verhoogd bij mensen die aangaven type 2 diabetes te hebben, terwijl het risico niet verhoogd was bij mensen met verstoorde glucose waarden of ongediagnosticeerde type 2 diabetes. Deze resultaten suggereren dat depressieve symptomen een gevolg zijn van de ziektelast van type

2 diabetes.

In **hoofdstuk 2.5** bepaalden we het gebruik van antidepressiva en benzodiazepines voor en na het starten van diabetes medicatie om de tijdsrelatie tussen diabetes en depressie en tussen diabetes en algemene psychosociale klachten te onderzoeken. De incidentie van antidepressiva en benzodiazepine gebruik was in de twee maanden voor en in de drie maanden na de start van diabetes medicatie verhoogd ten opzichte van mensen zonder diabetes. De sterkste toename in incidentie van antidepressiva en benzodiazepine gebruik was te zien in de eerste maand na de start van diabetes behandeling met een incidentie dichtheid ratio van 2,4 (95% BI: 2,0-3,0) en 3,4 (95% BI: 3,0-3,8), na correctie voor leeftijd, geslacht en chronische ziekten. De korte tijd tussen de start van diabetes medicatie en incidentie van antidepressiva en benzodiazepine gebruik maakt het onwaarschijnlijk dat er een causale relatie bestaat tussen depressie en diabetes. Een aannemelijke verklaring is dat depressie en algemene psychosociale klachten sneller worden gedetecteerd en behandeld als patiënten bij een arts komen voor diabetes gerelateerde klachten of, vice versa, dat diabetes eerder wordt ontdekt en behandeld als mensen voor depressieve of psychosociale klachten bij een arts komen.

Hoofdstuk 3 beschrijft verschillende methodologische aspecten van etiologisch onderzoek.

In **hoofdstuk 3.1** bepaalden we de mate en richting van bias die ontstond als we twee methoden om met missende waarden in een confounder om te gaan (missing indicator methode en complete case analyse) vergeleken met multipele imputatie. Uit de PREDICT-NL studie selecteerden we een determinant (huwelijkse staat), een uitkomst (depressieve stoornis) en drie confounder variabelen (leeftijd, geslacht en inkomen). We creëerden missende waarden in de confounder ‘inkomen’ volgens verschillende patronen. Deze studie liet zien dat de missing indicator methode vermeden moet worden om missende waarden in een confounder te hanteren, omdat het een vertekende schatting geeft van de odds ratio tussen determinant en uitkomst, zelfs als het percentage missende waarden laag is. Nog belangrijker is dat de richting van deze vertekening onvoorspelbaar is. Complete case analyse kan worden gebruikt als de missende waarden volledig willekeurig zijn, maar dit is zelden het geval. Bovendien leidt complete case analyse altijd tot verlies van gegevens en daardoor tot verlies van statistische power. Multipele imputatie geeft juiste effect schattingen als missende waarden volledig willekeurig zijn maar ook als ze afhangen van geobserveerde data. De laatste situatie is vaak het geval in medisch onderzoek. Daarom dient multipele imputatie gebruikt te worden in de meeste situaties waarbij er missende waarden in een confounder zijn.

In **hoofdstuk 3.2** illustreerden we met data van het Leidsche Rijn Gezondheidsproject hoe interactie op een additieve schaal kan worden geschat in een situatie met continue determinanten en een dichotome uitkomst. Bestaande literatuur werd gebruikt om methoden en formules af te leiden voor het schatten van de interactie en een betrouwbaarheidsinterval. Met behulp van een voorbeeld (met leeftijd en body mass index als risico factoren voor hypertensie) lieten we zien hoe deze formules gebruikt en de resultaten geïnterpreteerd kunnen worden. Dit hoofdstuk kan epidemiologen ondersteunen in het berekenen van interactie op een additieve schaal.

In **hoofdstuk 3.3** hebben we onderzocht hoe interactie gerapporteerd werd in 225 cohort en patiëntcontrole onderzoeken. De meeste van deze onderzoeken onderzochten inderdaad interactie (138 onderzoeken, 61%). Echter, in ongeveer de helft van deze artikelen was de gepresenteerde informatie onvoldoende. Slechts één op de tien artikelen presenteerde voldoende informatie om interactie zowel op een additieve als op een multiplicatieve schaal te kunnen interpreteren. Deze studie laat zien dat de rapportage van interactie verbeterd kan worden.

In **hoofdstuk 3.4** werd de mogelijkheid van selectie bias in een wachtkamerpopulatie geïllustreerd met behulp van data van de PREDICT-NL studie. Het oorspronkelijke doel van deze studie was om de relatie tussen diabetes en depressie te onderzoeken. Tegen onze verwachting in zagen we dat het aantal mensen met een depressie veel lager was bij diabetes patiënten (3,6%) dan bij mensen zonder diabetes (14,1%; odds ratio: 0,23, 95% BI: 0,07-0,73). We onderzochten vervolgens of deze onverwachte bevinding verklaard kon worden door differentiële selectie van mensen die hun huisarts bezochten voor een controle bezoek. Het aantal mensen dat voor een controle bezoek kwam, was inderdaad hoger bij de diabetes patiënten dan bij de mensen zonder diabetes (odds ratio: 2,79, 95% BI: 1,78-4,36). Tevens was het aantal mensen dat voor een controle bezoek kwam lager bij mensen met een depressie dan bij mensen zonder depressie (odds ratio: 0,64, 95% BI: 0,44-0,94). Dit onderzoek liet zien dat de lagere prevalentie van depressie bij diabetes patiënten voor een deel verklaard kon worden door differentiële selectie op basis van de reden van hun bezoek aan de huisarts.

In **hoofdstuk 3.5** bekeken we 150 gepubliceerde patiëntcontrole onderzoeken. Aan de hand van de selectie methode van patiënten en controles, bepaalden we hoe de berekende odds ratio's geïnterpreteerd konden worden. In 17 onderzoeken (11%) kon de odds ratio worden geïnterpreteerd als een risico ratio (cumulatieve incidentie ratio); in 16 van deze onderzoeken moest de uitkomst hiervoor zeldzaam zijn. In 57 onderzoeken (38%) kon de odds ratio geïnterpreteerd worden als een rate ratio (incidentie dichtheid ratio) op

voorwaarde dat de onderzochte populatie een stabiele populatie was, terwijl in 48 onderzoeken (32%) een rate ratio geschat werd zonder deze voorwaarde nodig te hebben. Een prevalentie odds ratio werd geschat in 12 onderzoeken (8%) en in 16 onderzoeken (11%) werd er onvoldoende informatie gerapporteerd om de odds ratio te kunnen interpreteren. De voorwaarde dat de uitkomst zeldzaam moest zijn, werd onnodig bediscussieerd in vier onderzoeken; geen van de onderzoeken besprak de noodzaak van een stabiele populatie. Deze studie liet zien dat de assumptie van een stabiele populatie een veel meer voorkomend aspect is van huidige patiëntcontrole onderzoeken dan de assumptie van een zeldzame uitkomst. Bovendien bespraken de auteurs zelf bijna nooit hoe de odds ratio in een patiëntcontrole onderzoek geïnterpreteerd kon worden.

Hoofdstuk 4 vat de beschikbare literatuur en studies beschreven in dit proefschrift samen over depressie als risico factor voor diabetes en over diabetes als risico factor voor depressie. Daarnaast werden verschillende methodologische aspecten van deze onderzoeken besproken.

De longitudinale onderzoeken waarin onderzocht werd of depressie een risico factor is voor diabetes, waren heterogeen met betrekking tot de studie karakteristieken en met betrekking tot de grootte van het effect. Alle onderzoeken hadden methodologische voor- en nadelen. Het is niet mogelijk om te zeggen welke potentiële vertekeningen van het effect het meest van invloed zijn op de resultaten en daarom kunnen we geen definitieve conclusie trekken of depressie een risico factor is voor diabetes en, als dat zo is, wat de grootte van het effect is. Toekomstige, goed opgezette, onderzoeken zijn nodig om een definitief antwoord te geven op de vraag of depressie een risico factor is voor diabetes.

De longitudinale onderzoeken waarin onderzocht werd of diabetes een risico factor is voor depressie, zijn heterogeen wat betreft studie karakteristieken maar niet wat betreft de grootte van het effect. Het lijkt er op dat het risico op een depressie niet verhoogd is bij diabetes patiënten na een periode van ten minste twee jaar na de diagnose. Het kan nog wel zo zijn dat het risico op depressie verhoogd is vlak na de diagnose van diabetes. Toekomstige onderzoeken zouden zich met name moeten richten op het bepalen van het tijdstip waarop het risico van een depressie verhoogd is na een diagnose van diabetes. Meer kennis hierover zou richting kunnen geven aan specifieke screening op depressie na een diagnose van diabetes.

Chapter 5

Op deze plek wil ik een aantal mensen bedanken voor hun bijdrage aan de totstandkoming van dit proefschrift.

Allereerst wil ik mijn beide promotoren en beide co-promotoren bedanken.
Prof. dr. D.E. Grobbee. Beste Rick, ik heb grote bewondering voor de manier waarop jij heel snel de essentie uit een artikel haalt. Bedankt voor je kritische blik bij het lezen van mijn artikelen en bedankt voor de mogelijkheid die je me geboden hebt om onderzoek te blijven doen op het Julius Centrum binnen het Escher project.
Prof. dr. A.C.G. Egberts. Beste Toine, bedankt voor je altijd snelle commentaar op mijn artikelen en e-mails en voor je oprechte interesse. Jouw passie voor je vak heeft mij erg gemotiveerd de afgelopen jaren. Ik hoop onze prettige samenwerking voort te zetten bij Escher; de s van drs is nu echt verdwenen.
Dr. M.I. Geerlings. Beste Mirjam, een jaar na het begin van mijn promotieproject werd jij mijn dagelijkse begeleider. Ik heb onder andere van je geleerd om dingen op te schrijven zoals ik ze bedoel. Bedankt dat je in oktober 2005 aan me vroeg of ik 'even' wilde uitzoeken of additieve interactie ook tussen twee continue determinanten berekend kan worden. Die vraag was het begin van het methodologische deel van dit proefschrift.
Dr. E.R. Heerdink. Beste Rob, bedankt voor je hulp bij het opschonen en analyseren van de immens grote databestanden met prescriptie gegevens. Jouw kennis van onderzoek met medicatie gegevens was van groot belang bij het schrijven van dit proefschrift.

Dr. H. Burger. Beste Huib, bedankt voor je hulp bij het opstarten van mijn promotie traject. Jammer dat we het traject niet samen voort konden zetten.
Dr. F. Pouwer. Beste Frans, onze prettige en efficiënte samenwerking is het bewijs dat communicatie via de e-mail ook erg goed kan verlopen.
Prof. dr. M. Egger. Dear Matthias, thank you for the opportunity you gave me to spend some time at the ISPM. Your commitment, both to your work and to your fellow researchers at the ISPM, inspired me. I want to thank you and the other people from ISPM for letting me feel at home in Bern.
Prof. dr. J.P. Vandenbroucke. Beste Jan, in korte tijd heb ik erg veel van je geleerd over verschillende aspecten van epidemiologisch onderzoek. Je bevlogenheid voor epidemiologisch onderzoek in het algemeen en voor case-control studies in het bijzonder vond ik erg motiverend.
De leden en adviseurs van UDES (Utrecht Diabetes Epidemiology Studies) wil ik bedanken voor hun input in het project: prof. dr. A. de Boer, drs. Ö. Bozkurt, dr. K.J. Gorter, dr. O.H.

Klungel, prof. dr. H.G.M. Leufkens, prof. dr. G.E.H.M. Rutten, prof. dr. R.P. Stolk.
Verder wil ik graag alle co-auteurs bedanken voor hun opbouwende commentaar op de artikelen in dit proefschrift.

Alle collega's van het Julius Centrum wil ik bedanken voor de uiteenlopende kennis die aanwezig is (en waarvan ik regelmatig gebruik maakte) en natuurlijk voor alle gezellige uitjes. In het bijzonder wil ik een aantal promovendi noemen die mijn werk op het Julius Centrum extra leuk maakten: Anne, Beate, Esther, Kristel, Marjolein, en Rolf.
Ondanks dat ik niet vaak aan de andere kant van de Uithof te vinden was, wil ik alle collega's van de afdeling farmacoepidemiologie en farmacotherapie bedanken. Özlem, Helga en Jeroen, succes met het afronden van jullie promotie. Patrick, bedankt voor al je hulp met de prescriptie databestanden.

Kristel, jouw belangstelling en betrokkenheid toen ik als pas afgestudeerde op het Julius Centrum kwam, heeft mij erg geholpen om me snel thuis te voelen. Bedankt dat je me gemotiveerd hebt om na te denken over wat ik na mijn promotie wilde doen. Leuk dat je mijn paranimf wilt zijn en leuk dat we samen op het Julius Centrum blijven werken.
Sandra, wij kunnen urenlang praten over promoveren, maar ook over heel veel andere dingen. Bedankt voor de vanzelfsprekendheid van onze vriendschap. Ik ben erg blij dat je tijdens mijn promotie achter mij staat als paranimf.

Vrienden en familie, hoewel ik met jullie niet veel over mijn promotie praatte, stel ik jullie belangstelling hiervoor erg op prijs. Vooral bedankt voor alle leuke en ontspannende activiteiten die ik buiten het werk om met jullie onderneem.
Papa en mama, bedankt voor jullie advies en steun als ik weer eens een keuze in mijn leven moest maken. Tot nu toe zijn het allemaal goede keuzes geweest.
Rolph, onze discussies over onderzoek waren door jouw niet-medische en niet-wetenschappelijke achtergrond vaak erg verhelderend voor mij. Bedankt voor deze bijdrage aan mijn proefschrift. Daarnaast wil ik je bedanken dat je er altijd voor mij bent.

Chapter 5

Summary

Samenvatting

Dankwoord

Curriculum vitae

Mirjam Knol was born in Amersfoort, the Netherlands, on June 22nd 1981. In 1999 she graduated from the Christelijk College Groevenbeek in Ermelo and started her studies in Health Sciences at the University Medical Center St. Radboud in Nijmegen. She obtained her Master of Science degree in 2004 with a specialization in Movement Sciences. During her studies she completed three research traineeships: 1) "Costs of multiple daily injections versus continuous subcutaneous insulin infusion in children with type 1 diabetes" at the Department of Public Health of the Erasmus Medical Center in Rotterdam, the Netherlands; 2) "Effect of exercise on the interstitial glucose concentration in human skeletal muscle" at the Department of Physiology and Pharmacology of the Karolinska Institute in Stockholm, Sweden; 3) "Gait, balance and falling in Huntington's disease" at the Department of Neurology of the University Medical Center St. Radboud in Nijmegen, the Netherlands.

In September 2004 she started her work described in this thesis at the Julius Center for Health Sciences and Primary Care of the University Medical Center Utrecht, the Netherlands, in collaboration with the Department of Pharmacoepidemiology and Pharmacotherapy of Utrecht University, the Netherlands, and under supervision of Professor D.E. Grobbee and Professor A.C.G. Egberts. She was member of the organizing committee of the European Congress of Epidemiology, which took place in June 2006 in Utrecht. In September 2006 she obtained her Master of Science degree in Clinical Epidemiology Cum Laude at Utrecht University, the Netherlands. She was one of the initiators and organizers of a one-day course on additive interaction, which took place in June 2007 in Leiden. From April to July 2007 she performed research at the Institute for Social and Preventive Medicine in Bern, Switzerland, under supervision of Professor M. Egger which resulted in two articles that are included in this thesis.

As from February 2008 she has a postdoctoral position in the Escher program of the Top Institute Pharma at the Julius Center for Health Sciences and Primary Care of the University Medical Center Utrecht, the Netherlands.