PYRY N. SIPILÄ
Dissecting Epidemiological Associations in Alcohol Drinking and Anorexia Nervosa
Dissecting epidemiological associations in alcohol drinking and anorexia nervosa

Pyry N. Sipilä

ACADEMIC DISSERTATION

To be presented with the permission of the Faculty of Medicine, University of Helsinki, for public examination in lecture hall PIII, Porthania, Yliopistonkatu 3, on 14 September 2018, at 12 noon.

Helsinki 2018
Dissertationes Scholae Doctoralis Ad Sanitatem Investigandam
Universitatis Helsinkiensis

Doctoral Programme in Population Health (DocPop)

Supervisors

Professor Jaakko Kaprio
Institute for Molecular Medicine Finland, University of Helsinki, and
Department of Public Health, University of Helsinki

Associate professor Anna Keski-Rahkonen
Department of Public Health, University of Helsinki

Reviewers

Research professor Pia Mäkelä
Alcohol and Drugs Unit, National Institute for Health and Welfare, Helsinki

Professor Henrik Larsson
Department of Medical Epidemiology and Biostatistics, Karolinska Institutet,
Stockholm, and
School of Medical Sciences, Örebro University

Opponent

Professor Sami Pirkola
Faculty of Social Sciences (Health Sciences), University of Tampere, and
Tampere University Hospital

© 2018 Pyry N. Sipilä
ISSN 2342-3161 (Print)
ISSN 2342-317X (Online)

Hansaprint Oy
Helsinki 2018
«And I applied my heart to seek and to search out by wisdom all that is done under heaven. It is an unhappy business that God has given to the children of man to be busy with.»

Ecclesiastes 1:13 (English Standard Version)
## CONTENTS

Abstract............................................................................................................... 7  
Tiivistelmä ........................................................................................................ 9  
List of original publications ............................................................................. 11  
Abbreviations .................................................................................................. 12  

1  Introduction .................................................................................................. 13  

2  Literature review ......................................................................................... 14  

2.1  Bias ......................................................................................................... 14  

2.1.1  Confounding ....................................................................................... 14  

2.1.2  Selection bias ...................................................................................... 15  

2.1.3  Information bias ................................................................................... 16  

2.1.4  Publication and funding bias ............................................................... 16  

2.2  Specific subject matter .......................................................................... 17  

2.2.1  Overview and epidemiology of anorexia nervosa ......................... 17  

2.2.2  Religiosity and anorexia nervosa ...................................................... 17  

2.2.3  Epidemiology of alcohol drinking .................................................... 19  

2.2.4  Definitions and measures of alcohol drinking ............................... 19  

  Total alcohol consumption ........................................................................... 20  

  Heavy drinking occasions (Binge drinking) ............................................. 20  

  Problem drinking ....................................................................................... 21  

  Alcohol use disorder .................................................................................. 21  

2.2.5  Parents’ and their children’s alcohol drinking ............................... 22  

  Genetic determinants of alcohol drinking .............................................. 22  

  Non-genetic familial determinants of alcohol drinking ............. 23  

  The question on causality ........................................................................ 23  

2.2.6  Alcohol drinking and health ............................................................... 24
Alcohol drinking and all-cause mortality ............................................24

2.3 Summary and open questions ......................................................26

3 Aims ..................................................................................28

4 Methods ..............................................................................29

4.1 Participants ........................................................................29

4.1.1 FinnTwin16 cohort ............................................................29

4.1.2 The Older Finnish Twin Cohort .......................................29

4.1.3 Ethical considerations ...................................................30

4.2 Measures ........................................................................30

4.2.1 Anorexia nervosa ..........................................................30

4.2.2 Religiosity ......................................................................31

4.2.3 Alcohol drinking ...........................................................32

Total alcohol consumption ........................................................32

Heavy drinking occasions .........................................................32

Alcohol-induced blackouts .......................................................32

Problem drinking ................................................................33

Drinking frequency ................................................................34

Grandparents’ drinking .........................................................35

4.2.4 Zygosity .......................................................................35

4.2.5 All-cause mortality .........................................................35

4.2.6 Covariates ....................................................................35

4.3 Statistical methods .............................................................37

4.3.1 Means ............................................................37

4.3.2 Correlations ...........................................................37

4.3.3 Multiple imputation ....................................................38

4.3.4 Risk of lifetime anorexia nervosa .................................39

4.3.5 Survival analysis .........................................................39
Abstract

Background. Biases threaten the validity of practically every epidemiological study. Hence, in this study, I tackled potential sources of bias in psychiatric epidemiology with systematic, population-based cohort studies in the context of anorexia nervosa and alcohol drinking. I used multiple imputation to reduce selection bias, and examined previously overlooked potential confounders with both traditional methods and using a natural experiment, the discordant-twin design.

Aims. One, to examine systematically individual and family religiosity as potential risk factors for anorexia nervosa on the population level. Two, to assess whether potential confounders identified from the literature can explain the association of parental problem drinking with problem drinking of their adult children. Three, to assess the potential confounding effects of genetic factors and childhood family environment in the association of alcohol drinking with all-cause mortality.

Methods. I used the population-based FinnTwin16 cohort (studies I and II) and the population-based Older Finnish Twin Cohort (study III). In study I (n = 2639), I examined the association of fathers’, mothers’ and women’s religiosity with lifetime anorexia nervosa (n = 91), reducing selection bias by multiple imputation. In study II (1235 men and 1461 women assessed in early adulthood), I examined the relation between parents’ and their adult children’s problem drinking with multiple linear regression. In study III (n = 14787), I examined the relationship between different dimensions of alcohol drinking and all-cause mortality (2203 deaths) using Cox proportional hazard models, and assessed the potential confounding effects of genetic factors and childhood family environment using the discordant-twin design.

Results. In study I, reducing selection bias with multiple imputation did not change the results: personal or family religiosity did not predict anorexia nervosa. In study II, area of residence, family structure, and fathers’ and mothers’ education, religiosity and one relevant dimension of personality were addressed as potential confounders. The previously overlooked potential confounders could not explain the association of parents’ problem drinking with problem drinking of their adult children. In study III, the confounding effects of genetic factors and shared childhood environment could not explain the associations of total alcohol consumption of at least 259 grams per month (more than about 5 drinks per week) and alcohol-induced blackouts (at least twice a year) with all-cause mortality. The findings for heavy drinking occasions were not statistically significant among monozygotic twin pairs.
Conclusions. I examined three potential sources of bias in psychiatric epidemiology. First, in a systematic study, in which I tried to minimize selection bias, religiosity did not seem to be a major risk factor for anorexia nervosa. This underscores the importance of systematic evidence as many case reports suggest the opposite. Second, the association between parents’ and their children’s problem drinking did not appear to be attributable to the proposed confounding factors. Nevertheless, causality cannot be inferred, as I was unable to exclude the effect of genetic predisposition to problem drinking. Third, the confounding effects of genetic factors, shared childhood environment, or traditionally assessed potential confounders could not explain the associations of total alcohol consumption and alcohol-induced blackouts with all-cause mortality.
Tiivistelmä


Tavoitteet. Tutkia systemaattisesti onko perheen tai yksilön itsensä uskonnollisuus laihuushäiriön riskitekijä väestötasolla. Tutkia selittävätkö kirjallisuudesta tunnistamani mahdolliset sekoittavat tekijät vanhempien ja heidän lastensa haitallisen alkoholinkäytön välillä olevan yhteyden. Tutkia sekoittavatko perimän ja lapsuuden ympäristö alkoholinkäytön ja kokonaiskuolleisuuden välistä yhteyttä.


Tulokset. Työssä I valikoitumisharhan vähentäminen moni-imputoinnilla ei muuttanut tuloksia: vanhempien tai yksilön itsensä uskonnollisuus ei ollut yhteydessä laihuushäiriöön. Työssä II huomioin asuinalueen, perherakenteen ja isien ja äitien koulutuksen, uskonnollisuuden ja yhden persoonallisuuden ulottuvuuden mahdollisina sekoittavina tekijöinä. Aiemmin vähälle huomiolle jääneiden mahdollisten sekoittavien tekijöiden huomioiminen ei selittänyt vanhempien ja heidän lastensa haitallisen alkoholinkäytön välistä yhteyttä. Työssä III perimän ja lapsuuden ympäristön sekoittavaa vaikutus ei selittänyt korkean alkoholin kokonaiskulutuksen (vähintään 259 g kuukaudessa eli enemmän kuin noin viisi annosta viikossa) tai sammutumisten (vähintään kahdesti vuodessa) yhteyttä kohonneeseen kokonaiskuolleisuuteen. Runsaan kertauomisen yhteys kokonaiskuolleisuuteen ei ollut tilastollisesti merkitsevää identtisillä kaksosilla.
List of original publications

This thesis is based on the following publications:


II  Sipilä PN, Keski-Rahkonen A, Lindbohm JV, Rose RJ, Kaprio J. Parental problem drinking and later problem drinking among their adult children. Submitted manuscript.


The publications are referred to in the text by their roman numerals and they are reprinted with a permission of their copyright holders.

*These authors contributed equally to the study.
Abbreviations

CI confidence interval
DSM-III-R Diagnostic and Statistical Manual of Mental Disorders, Third Edition - Revised
DSM-IV Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition
DSM-5 Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
DZ dizygotic
EDI-2 Eating Disorder Inventory 2
e.g. exempli gratia [for the sake of example]
HDO heavy drinking occasions
HR hazard ratio
ICD-10 International Statistical Classification of Diseases and Related Health Problems, 10th Revision
i.e. id est [that is]
IRB Institutional Review Board
MAST Michigan Alcoholism Screening Test
Mm-MAST Malmö-modified Michigan Alcoholism Screening Test
Mm-MAST-11 Malmö-modified Michigan Alcoholism Screening Test, extended, 11-item version
MMPI Minnesota Multiphasic Personality Inventory
MZ monozygotic
p p-value
Pd Pd or “Psychopathic deviate” scale of the Minnesota Multiphasic Personality Inventory
PhD philosophiae doctor [Doctor of Philosophy]
SCID Structured Clinical Interview for DSM-IV
SD standard deviation
SMC-FCS substantive-model compatible fully conditional specification
vs verse
1 Introduction

A major reason for wrong results in science is bias (Ioannidis, 2005). Biases threaten the validity of practically every epidemiological study (Rothman et al., 2008a). But despite the universality of the risk of bias, and its potentially serious consequences for the validity of epidemiological research, bias is still often addressed inadequately (Hemkens et al., 2018). Hence, epidemiological studies with vigorous efforts to control potential sources of bias are urgently needed.

In this study, I will tackle bias in psychiatric epidemiology – a field with a major role in public health. Psychiatric disorders and substance use are among the top causes of the global burden of disease. The conventional estimate places them on the 5th place among all disorders (Whiteford et al., 2013), but some authors even argue that the global burden of disease associated with psychiatric and substance use problems is second only to cardiovascular and circulatory disorders (Vigo et al., 2016).

Within psychiatric disorders, two broad groups can be identified: internalizing and externalizing disorders (Achenbach, 1966; Krueger et al., 1998). In this study, I will focus on one internalizing and one externalizing problem. Anorexia nervosa is an internalizing disorder that causes a high burden of disease and substantial mortality, especially among women (Harris & Barraclough, 1998; Forbush et al., 2010; Hoek, 2016; Keski-Rahkonen & Mustelin, 2016). In contrast, risky alcohol drinking is an externalizing problem that is more common in men (Halme et al., 2008; Grant et al., 2017), and causes an enormous burden of disease (Lim et al., 2012; Whiteford et al., 2013).

The aetiology of anorexia nervosa and alcohol use disorder, and of psychiatric problems in general, is complex and remains poorly understood (Kendler, 2008, 2014; Zipfel et al., 2015). Anorexia nervosa, alcohol drinking, and alcohol use disorder typically have their onsets in adolescence or early adulthood (Hingson et al., 2006; Volpe et al., 2016). Therefore, research on risk factors stemming from the family environment is critical to efforts to improve prevention and early detection of psychiatric disorders.

Specifically, in this study, I will examine religiosity and parental problem drinking as potential risk factors from family environment in the epidemiology of anorexia nervosa and alcohol drinking. Towards the end of this study, I will also broaden my scope, and look at the causal role of alcohol drinking in the ultimate adverse outcome – death. I will address potential sources of bias by conducting systematic, population-based cohort studies, by using multiple imputation to reduce selection bias (Sterne et al., 2009; Hernán & Robins, 2018), and by examining potential confounders overlooked in earlier research using both traditional methods and a natural experiment, discordant-twin design.
2 Literature review

I will start this literature review by introducing the concept of bias in epidemiology. Then, I will review the branches of psychiatric epidemiology relevant to this study: religiosity and parental problem drinking as potential risk factors from family environment in the epidemiology of anorexia nervosa and alcohol drinking, and the role of alcohol drinking in all-cause mortality. Within these branches, I will review the current state of knowledge, and identify potential sources of bias that may arise from gaps in the current body of literature.

2.1 Bias

In scientific research, there are two kinds of errors: random and systematic (Rothman et al., 2008b). Epidemiologist call systematic errors biases. While random variation of results is inherent to the nature, biases will distort the results of the study beyond the effects of random variation (Egger et al., 1998; Lindley, 2014).

Bias can arise in numerous ways. For example, a recent glossary of the most important biases lists 77 different biases (Delgado-Rodríguez & Llorca, 2004). Fortunately, most biases can be grouped into three main categories: confounding, selection bias and information bias (Delgado-Rodríguez & Llorca, 2004; Rothman et al., 2008b), although the distinctions between these groups are not sharp. Especially, the definitions of selection bias and confounding vary from author to author (Glymour & Greenland, 2008; Haneuse, 2016; Hernán & Robins, 2018).

In the following paragraphs, I will outline the basic characteristics of each of the three main categories: confounding, selection bias and information bias. In addition, I will briefly discuss publication bias and funding bias. I will use the classification of Rothman et al. (2008a). What really matters, however, are not the exact classification and definitions of the biases, but adequate treatment of them in epidemiological research.

2.1.1 Confounding

Confounding, or confounding bias, is probably the most serious threat to the validity of observational research (Haneuse, 2016; Hernán & Robins, 2018; Hemkens et al., 2018). This is partly because a researcher can rarely, if ever, be sure that all possible sources of confounding have been adequately taken into account (Weiss, 2008).

Confounding occurs when the effect of the exposure on the outcome is biased by the effect of a third factor, a confounder (Pearl, 2009). Think of the...
association of alcohol drinking with lung cancer, for example. People who drink more tend also to smoke more, and smoking increases the risk of lung cancer. Therefore, people who drink alcohol will have an increased risk of lung cancer because of a third factor: smoking. Smoking confounds the association between alcohol drinking and lung cancer unless its effects are properly controlled for (Breslow & Day, 1980; Djoussé et al., 2002).

Formally, the necessary (but not sufficient) criteria for a confounding factor are: 1) it is a risk factor for the outcome under study, 2) it is associated with the exposure under study, and 3) it is not on the causal pathway that leads from the exposure to the outcome (Greenland et al., 1999).

Confounding can distort the true association to any direction. It can exaggerate and hide true effects, create spurious effects when the real association is null, and even turn positive associations to negative and vice versa (Rothman et al., 2008b).

Interesting special cases of confounding include healthy worker effect and confounding by indication. Healthy worker effect means that those who are able to work are healthier than the general population. Therefore, bias will arise if e.g. an occupational hazard is studied by comparing exposed workers to the whole population (Hernán et al., 2004). Comparative biases can occur if the selection of study participants is conditional on health status or a correlate of health (such as ability to travel to a study site). Confounding by indication can bias studies that compare different treatments. The severity of the disease or other patient-related characteristics can influence the selection of the treatment which may distort the results of the study (Miettinen, 1983). For example, prescription of antipsychotic drugs may seem to worsen the prognosis of psychiatric patients if they received those drugs precisely because they were worse off in the first place.

### 2.1.2 Selection bias

Selection bias arises when the association under study differs between the source population of the study (those who are supposed to be studied) and the actual study population (those who are studied). Selection bias may arise from factors that affect the selection of participants to the study or from factors that affect study participation (Heckman, 1979).

A few examples of selection bias include Berksonian bias, self-selection bias and missing data. Berksonian bias arises when both exposure and outcome affect the probability of inclusion to the study. It is of special concern in hospital-based studies, and it may either exaggerate or mask the effects of the exposure (Berkson, 1946).

Self-selection bias arises when those volunteering to participate the study have a different chance for the outcome than those who do not volunteer (Greenland, 1977). For example, a screening study may exaggerate the positive effects of screening on survival if those volunteering for the study are more health conscious, and thus healthier, than the general population (Rothman,
A similar selection bias may arise when loss to follow-up is not random (Greenland, 1977).

Missing data is still one important source of selection bias (Hernán & Robins, 2018). It affects selection to the study if only those with complete data are analysed. This may bias the results unless the data are missing completely at random. In contrast, data that are missing completely at random will only lead to loss of precision (Sterne et al., 2009).

### 2.1.3 Information bias

Information bias arises from measurement errors (Rothman et al., 2008b). Both researcher dependent and study participant dependent reasons can cause measurement errors (Szklo & Nieto, 2014). Recall bias is a special case of information bias. It causes most problems in case-control studies, in which information on exposure is collected after the occurrence of the outcome (Szklo & Nieto, 2014).

In the instance of categorical variables, measurement error is often called misclassification. While non-differential misclassification is independent of any variables in the study, differential misclassification is not. Differential misclassification is especially malicious; it can bias the results to any direction (Rothman et al., 2008b). In contrast, non-differential misclassification of an inherently binary variable will usually bias the results towards null (Copeland et al., 1977; Rothman et al., 2008b). Nevertheless, when a continuous variable, or a categorical variable with more than two categories, is collapsed into a binary variable, non-differential misclassification of the original variable may lead to differential misclassification of the binary variable (Wacholder et al., 1991; Flegal et al., 1991). Further, non-differential misclassification of variables with more than two categories may bias the results to any direction (Dosemeci et al., 1990). Finally, misclassification of a confounder will cause residual confounding (Fewell et al., 2007).

### 2.1.4 Publication and funding bias

Publication bias occurs when the results of a study affect its probability to be published (Dickersin, 1990). The most well known form of publication bias is significance bias: studies with statistically significant results are more likely to be published (Sterling, 1959; Dickersin, 1990; Easterbrook et al., 1991; Dickersin et al., 1992). Publication bias is a problem, because it can distort the scientific evidence and the conclusions that are drawn from the evidence.

Another related bias is funding bias. Studies funded by the industry tend to favour the products of the sponsoring industry more often that studies that did not receive industry funding (Lexchin et al., 2003; Lundh et al., 2017). Possible reasons for this include biased selection of control interventions, biased interpretation of the results and publication bias: unfavourable results may be suppressed from publication (Lexchin et al., 2003; Lundh et al., 2017).
2.2 Specific subject matter

Addressing bias is crucial for psychiatric epidemiology; it improves the validity of research which is needed for efficient interventions to prevent and treat psychiatric disorders (Kendler, 2017).

Next, I will review the specific topics of psychiatric epidemiology within which I will identify and examine potential sources of bias. I will briefly introduce anorexia nervosa, the first target outcome of this thesis, and review religiosity as a potential family-environment-related risk factor for anorexia nervosa. I will then introduce alcohol drinking, the second specific target of this thesis, and review parents’ alcohol drinking as a risk factor for their children’s alcohol drinking. I will also review the role of alcohol drinking in all-cause mortality. I will discuss the current state of knowledge, and the gaps within it, in order to identify potential sources of bias.

2.2.1 Overview and epidemiology of anorexia nervosa

Anorexia nervosa is a mental disorder characterized by “restriction of energy intake relative to requirements, leading to significantly low body weight”, “intense fear of gaining weight” and disturbed perception of body shape or weight (American Psychiatric Association, 2013).

Anorexia nervosa is relatively common in European women, with lifetime prevalence up to 4%, and it is associated with a high burden of disease and substantial mortality (Harris & Barraclough, 1998; Hoek, 2016; Keski-Rahkonen & Mustelin, 2016). Albeit less commonly, anorexia nervosa also occurs in non-Western countries and in men (Raevuori et al., 2009; Hoek, 2016).

2.2.2 Religiosity and anorexia nervosa

Religiosity is a multidimensional phenomenon that has nuanced relationships with health (Koenig et al., 2012). In psychiatric epidemiology, strongest evidence is available for depression, substance abuse and suicide: religious involvement seems to protect from them (Bonelli & Koenig, 2013; VanderWeele et al., 2016).

The aetiology of anorexia nervosa is not yet clear (Zipfel et al., 2015). Many case reports and series suggest that religiosity may be one factor contributing to the onset of anorexia nervosa. This evidence spans from the Middle Ages to modern days (Bell, 1987; Bynum, 1987; Morgan et al., 2000; Bennett et al., 2004; Marsden et al., 2007; Kaluski et al., 2008; Abraham & Birmingham, 2008; Moga et al., 2009; Espi Forcen, 2013; Davis & Nguyen, 2014; Harris, 2014; Akgül et al., 2014). Although case reports and series cannot be used to estimate absolute or relative risks, or to confirm hypotheses, they can be very useful in suggesting new explanations that can later be confirmed or refuted in more systematic studies (Vandenbroucke, 2001; Dekkers et al., 2012).
Religious asceticism and fasting could explain the possible association of religiosity with anorexia nervosa (Huline-Dickens, 2000). Another potential mechanism could be tension between religious parents and their children. Some studies have suggested that this kind of tension between young immigrant women, who adopt Western values, and their religious families may predispose these women to eating disorders (Ahmad et al., 1994; Furnham & Husain, 1999; Gordon, 2000). In these studies, however, it is difficult to distinguish between the effects of religion and culture, because the religious families also belong to an ethnic minority. A Canadian study among female adolescents emphasizes this mixing of religiosity with ethnic minority status. It found that Jewish girls had more disordered eating than non-Jewish girls, but among the Jewish religious observance was not associated with disordered eating (Pinhas et al., 2008).

Few systematic studies have examined the potential role of religiosity in anorexia nervosa (Bonelli & Koenig, 2013). A couple of studies have found associations between religiosity and disordered eating behaviours (Gates & Pritchard, 2009; Thomas et al., 2018). In other studies, the associations have been complex. That is, different aspects of religiosity have shown either positive, negative or no associations (Smith et al., 2004; Kim, 2006, 2007; Castellini et al., 2014; Akrawi et al., 2015). Two studies from clinical settings looking directly into anorexia nervosa have suggested an association between anorexia nervosa and religiosity: Wilbur and Colligan (1981) observed that female patients with anorexia nervosa had higher scores on the Religious fundamentalism content scale of the Minnesota Multiphasic Personality Inventory than did patients treated for another nonpsychotic psychiatric illness. Sykes et al. (1988) found that among those who were referred to treatment for anorexia nervosa, there were less Protestants than there were in the general population on the same metropolitan area. For Catholics they observed no difference. Yet, some studies have found no associations (Feinson & Meir, 2012a, 2014). Moreover, one study even reported that religiosity may protect from body dissatisfaction and disordered eating (Gluck & Geliebter, 2002). A possible explanation for this finding is that religiosity may protect from the Western sociocultural pressure to be thin (Platte et al., 2000; Gluck & Geliebter, 2002; Homan & Boyatzis, 2010). Albeit this sociocultural pressure to be thin is not necessary for the development of anorexia nervosa, many authors think it is an important risk factor for eating disorders (Nasser, 1986, 1988; Hoek et al., 1998; Gordon, 2000; Bhugra et al., 2003; Homan & Boyatzis, 2010; Zipfel et al., 2015). Further, the potential protective role of religiosity in eating disorders is not limited to the onset of them. During the course of the illness, religiosity may provide tools for coping (Jacobs-Pilipski et al., 2005) and protect mental health (Henderson & Ellison, 2015).

To my knowledge, only two population studies have studied religiosity in the context of disordered eating. Neither of them has looked directly into the association of religiosity with anorexia nervosa. Boisvert and Harrel (2013) found no direct association between religiosity and eating disorder
symptomatology, but religiosity was associated with existential well-being, which in turn was negatively associated with eating disorder symptomatology. Henderson and Ellison (2015) found that religiosity may protect mental health among those who are affected by eating disorders. To summarize, the question whether or not religiosity is associated with anorexia nervosa remains open.

2.2.3 Epidemiology of alcohol drinking
Alcoholic beverages have been consumed for thousands of years (Michel et al., 1992; McGovern et al., 2004, 2017), and their popularity prevails. With the exception of some countries with a large Muslim population, alcoholic beverages are popular around the globe (World Health Organization, 2014).

A high proportion of adult population drink alcohol in Western countries. In 2010, 66% of European adult population (those who were at least 15 years old) were current drinkers, while the global percentage was only 38. In the United States of America, a recent nationally representative survey found that 73% of adult population (defined as 18-year-old and older people) drank alcohol during the past year (Grant et al., 2017). Drinking was a bit more common in men than in women (77% vs 69%). In Finland, alcohol consumption rose strongly after selling middle strength beer in grocery stores was legalized in 1969 (Mäkelä & Österberg, 2016). Nowadays Finns drink alcohol in similar amounts to most Western countries (World Health Organization, 2014). In 2016, adult per capita alcohol consumption was 10.8 l of pure alcohol among Finns who were at least 15 years old (National Institute for Health and Welfare, 2017). In 2013, among 15–64-year-old Finns, 88% of men, 85% of women and 87% of both sexes combined drank alcohol during the last year (Mäkelä & Härkönen, 2017).

The prevalence of alcohol use disorder was 8% in Europe and 4% globally in 2010, according to the criteria of the International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10) (World Health Organization, 2014). In the United States of America, alcohol use disorder was likewise common: 17% of men, 9% of women and 13% of total population were affected during the past year according to the criteria of the Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) (Grant et al., 2017). In addition, the prevalence of both alcohol drinking and alcohol use disorder had increased during the last eleven years (Grant et al., 2017). In Finland, the latest figures are from 2000–2001. At that time, the prevalence of ICD-10 alcohol use disorder was 8.9% in men, 1.9% in women and 5.4% in both sexes combined (Halme et al., 2008).

2.2.4 Definitions and measures of alcohol drinking
Alcohol drinking is a multidimensional phenomenon. Different aspects of it can be characterized in terms of frequency and typical amount of drinking,
total alcohol consumption, heavy drinking occasions and problem drinking (Rehm, 1998; Rehm et al., 2017).

National Institutes of Health recommend measuring alcohol drinking at least in three dimensions: frequency of drinking, drinking amount on a typical drinking day and frequency of binge drinking, defined as drinking “5 or more (males) or 4 or more (females) drinks containing any kind of alcohol in within a two-hour period” (National Institutes of Health, 2003). A drink – or a standard drink – typically refers to a bottle or a can of beer, a glass of wine or a shot of liquor or spirits (National Institutes of Health, 2003). World Health Organization recommends a definition of 10 g for a standard drink, but in practice the definition varies by country from 8 g in the United Kingdom to 12 g in Finland, 14 g in the United States and 20 g in Austria (Mongan & Long, 2015; Kalinowski & Humphreys, 2016).

Total alcohol consumption

Total alcohol consumption is simply defined as the total amount of alcohol drunk by a person in a certain period. In practice, an estimate of total alcohol consumption can easily be calculated from self-reported frequency and typical amount of drinking (Rehm, 1998), but the reference period in questionnaires can vary from last 12 months to last month, last week and simply average drinking without an explicit timeframe (Kaprio et al., 1987; National Institutes of Health, 2003; Boniface & Shelton, 2013). Whatever the reference period, the estimated total alcohol consumption can be expressed in any convenient unit. Typical choices are litres per year and grams per day (World Health Organization, 2014), but some researchers prefer grams per week or grams per month (Kaprio et al., 1987; Bagnardi et al., 2008).

Heavy drinking occasions (Binge drinking)

Heavy drinking occasions (HDO) mean drinking large amounts of alcohol on a single occasion (Rehm et al., 2017). Sometimes the occasion is more definitely defined to be two hours (National Institutes of Health, 2003). Heavy drinking occasions can also be called risky single-occasion drinking (Gmel et al., 2011), heavy episodic drinking (World Health Organization, 2014) or binge drinking (National Institutes of Health, 2003). The term binge drinking has two different meanings. Traditionally it has been used to refer to “[a] pattern of heavy drinking that occurs in an extended period set aside for the purpose” (World Health Organization, 1994), but the modern usage is synonymous to heavy drinking occasions (National Institutes of Health, 2003; Gmel et al., 2011).

There is no universal agreement on how much exactly one needs to drink to have a heavy drinking occasion, but the most commonly used cut-off is 60 g of pure alcohol (Gmel et al., 2011; World Health Organization, 2014; Rehm
et al., 2017). The National Institute on Alcohol Abuse and Alcoholism defines “a binge” as a blood alcohol concentration of 0.08% or more, which according to them corresponds to 5 drinks or more for men and 4 drinks or more for women within two hours (Department of Health and Human Services, National Institutes of Health, 2004). Nevertheless, the actual blood alcohol concentrations after 5 or 4 drinks (for men and women, respectively) vary a lot (Gmel et al., 2011).

**Problem drinking**

Problem drinking is alcohol drinking that leads to problems or risk of problems, but does not fulfil the criteria of alcohol use disorder. These problems may be health-related or social (Kahan, 1996; Aronson, 2017).

Problem drinking can be measured with numerous self-report-based scales and screening tests (Gibbs, 1983; White & Labouvie, 1989; Cherpitel, 1997; Allen et al., 1997; Bush et al., 1998; Fiellin et al., 2000; Hodgson et al., 2002; Miller et al., 2007; O’Brien, 2008). They are useful in detecting problem drinking and alcohol use disorders in clinical settings (Fiellin et al., 2000; Allen et al., 2001; Reinert & Allen, 2002; Dhall & Kopec, 2007). They are also used in research settings to measure problem drinking and alcohol use disorders (Seppä et al., 1999; Pitkänen et al., 2005; Dick et al., 2011a; Bloomfield et al., 2013). Nonetheless, they have their limitations: their sensitivity and specificity is far from perfect, and their performance may vary by sex and dimension of harmful drinking (Gottesman, 1989; Allen et al., 1997, 2001; Fiellin et al., 2000; Reinert & Allen, 2002; Dhall & Kopec, 2007).

**Alcohol use disorder**

Clinical **alcohol use disorder** is the most severe expression of alcohol drinking. The diagnosis of alcohol use disorder in the present, fifth, edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) is characterized by eleven symptoms. They are recurrent or persistent 1) drinking more or longer than was intended, 2) desire or futile efforts to reduce drinking, 3) spending a lot of time with obtaining or drinking alcohol or recovering from drinking, 4) strong desire to drink, 5) failure to fulfil obligations in consequence of drinking, 6) continuing drinking despite social harm from drinking, 7) giving up important activities due to drinking, 8) drinking in situations where it is dangerous, 9) continuing drinking despite knowledge of harm caused by drinking, 10) tolerance and 11) withdrawal. Severity of alcohol use disorder is defined by the number of existing symptoms: mild (2–3 symptoms), moderate (4–5), and severe (6 or more) (American Psychiatric Association, 2013). In contrast, in the previous edition, DSM-IV, severity of the condition was taken into account by dividing alcohol use disorder into two diagnoses, abuse and dependence. Dependence was the

2.2.5 Parents’ and their children’s alcohol drinking

It has been known for decades that parental alcohol use disorder is a risk factor for alcohol use disorder in their offspring (Cotton, 1979; Johnson & Leff, 1999). In addition, newer evidence shows that the full range of parents’ alcohol drinking is associated with drinking of their offspring (Rossow et al., 2016b, 2016a). Both genetic and environmental effects seem to contribute to these associations (Dick et al., 2009; Verhulst et al., 2015; Dick, 2016).

Genetic determinants of alcohol drinking

Twin and adoption studies demonstrate that alcohol drinking and alcohol use disorder have a substantial genetic component to them. The heritability of the frequency of alcohol drinking has been estimated to be 0.27 [95% confidence interval (CI) 0.05–0.50] and 0.27 (95% CI 0.14–0.43) in 14 years old Finnish male and female twins, respectively (Dick et al., 2009). Other studies into adolescent twins and adoptees have discovered heritability estimates for various measures of alcohol drinking that range from 0.2 to 0.72. A study from the Netherlands is an exception: adolescent and young adult women with religious upbringing had a zero heritability for alcohol use initiation (Koopmans et al., 1999; Hopfer et al., 2003; Unger et al., 2011).

In a recent meta-analysis, the heritability of alcohol use disorder was 0.51 (95% CI 0.45–0.56) when twin studies were combined and 0.36 (95% CI 0.22–0.50) when adoption studies were combined (Verhulst et al., 2015). These heritability estimates did not differ significantly between men and women in either of these comparisons. It is crucial to note, however, that heritability is not a universal constant, but specific to environment, place and time (Koopmans et al., 1999; Hopfer et al., 2003).

Strong evidence links genetic polymorphisms in major alcohol metabolizing enzymes alcohol dehydrogenases and aldehyde dehydrogenases to the risk of alcohol drinking and dependency (Bierut et al., 2012; Li et al., 2012; Gelernter et al., 2014; Quillen et al., 2014). Polymorphisms in γ-aminobutyric acid type A receptor α2 subunit have also been found to be associated with the risk of alcohol dependence (Edenberg et al., 2004; Covault et al., 2004). Despite these findings and high heritability estimates in twin and adoption studies, in overall, molecular genetic studies have found few specific genes contributing to the heritability of alcohol drinking and alcohol use disorder. In a recent study, a polygenic risk score explained only some 0.6% of the variance in alcohol problems (Salvatore et al., 2014). For alcohol dependence (9% or 30%, depending on the study) and alcohol use (13%),
polygenetic risk scores have been more successful (Palmer et al., 2015; Clarke et al., 2017; Walters et al., 2018).

**Non-genetic familial determinants of alcohol drinking**

Twin studies have shown that common environmental factors, which the twins share, largely determine drinking initiation in adolescence. After initiation, only 8–15% of variance in problem drinking and 10% of variance in alcohol use disorder is explained by common environmental factors in twin and adoption studies (Pagan et al., 2006; Verhulst et al., 2015). Moreover, across adolescence till early adulthood, the share of variance in alcohol drinking habits explained by genetic effects increases with increasing age (Viken et al., 1999; Rose et al., 2001, 2001; Hopfer et al., 2003; Pagan et al., 2006).

Several factors may contribute to the familial environment that predisposes to alcohol drinking and problems. Parental alcohol drinking may lead to social learning (Rossow et al., 2016b). It may also affect parenting, such as monitoring and discipline (Latendresse et al., 2008), and increase stress in the family (Leonard & Eiden, 2007). These childhood adversities may further activate genetic predisposition to alcohol drinking: a gene–environment interaction (Jacob et al., 2003; Rossow et al., 2016b).

**The question on causality**

Despite strong evidence for the familial aggregation of alcohol drinking and drinking problems, and for knowledge about both genetic and environmental effects contributing to these associations, it is unknown whether parental alcohol drinking has causal effects on alcohol drinking of their offspring (Rossow et al., 2016b, 2016a).

A recent systematic review suggests that potential confounders that could have caused spurious associations to appear between parental and offspring alcohol drinking are local environment, cultural and religious factors, and parental comorbidity and temperament (Rossow et al., 2016b). Other potential confounders include socioeconomic status and childhood family structure (Kestila et al., 2008).

As reviewed above, genetic predisposition to alcohol drinking could also explain the association of parental alcohol drinking with offspring alcohol drinking. Existing studies on children of twins suggests that both genetic and environmental effects contribute to the familial aggregation of alcohol use disorder, but their statistical power is not sufficient to draw clear conclusions (Jacob et al., 2003; Duncan et al., 2006; Slutske et al., 2008; McAdams et al., 2014). Recent studies on adopted children and triparental families, however, imply that the association of parents’ alcohol use disorder with offspring alcohol use disorder is not likely to be fully explained by genetic predisposition.
to alcohol abuse that is inherited from the parents to their children (Kendler et al., 2015a, 2015b). This leaves the possibility of causal effects open.

### 2.2.6 Alcohol drinking and health

Excessive alcohol drinking is associated with a myriad of diseases. These include injuries, diseases of the liver and pancreas, neuropsychiatric disorders, infectious diseases, cardiovascular diseases and many types of cancers (Rehm et al., 2017; Topiwala et al., 2017). Despite these numerous associations, the overall relationship between alcohol and health is not that clear; observational studies have consistently found a reduced risk of cardiovascular diseases among moderate drinkers (Roerecke & Rehm, 2012; Bell et al., 2017).

This postulated cardioprotective effect of moderate alcohol drinking is highly controversial (Fernández-Solà, 2015). A recent Mendelian randomization study found no evidence for a protective effect of alcohol drinking on coronary heart disease (Holmes et al., 2014), but a reanalysis of a subset of the same data partly disagreed. Focusing on nonlinearities, it indicated that light alcohol drinking might have small beneficial effects on some cardiovascular risk factors: systolic blood pressure, non-high-density lipoprotein cholesterol, body mass index, waist circumference and C-reactive protein (Silverwood et al., 2014).

Given the possibility of both harmful and beneficial effects, how to assess the overall effect of alcohol drinking on health? One approach is to calculate global burden of disease estimates (Lim et al., 2012). The problem is that these estimates are dependent on both several assumptions and the quality of the underlying meta-analyses and individual studies (Polinder et al., 2012). Another approach is to look at all-cause mortality. All-cause mortality is admittedly a crude measure of the entire spectrum of harm and possible benefits caused by alcohol drinking. On the other hand, it can be measured objectively and accurately, and it sums together both the harm and the possible benefits.

### Alcohol drinking and all-cause mortality

It has been a while since F. G. P. Neison and Raymond Pearl observed that excessive alcohol drinking is associated with increased mortality (Neison, 1851; Pearl, 1923). Meta-analyses of modern observational studies have confirmed their findings (Di Castelnuovo et al., 2006; Jayasekara et al., 2014) and new observational cohort studies keep coming (Zaridze et al., 2014; Smyth et al., 2015; Goulden, 2016; Wood et al., 2018). A notable exception to the general picture is a recent British study: it found no associations between alcohol drinking and all-cause mortality among young people and elderly men; only elderly women seemed to be protected by moderate drinking (Knott et al., 2015).
The present evidence indicates that high total alcohol consumption (Di Castelnuovo et al., 2006; Jayasekara et al., 2014), heavy drinking occasions (Kauhanen et al., 1997; Rehm et al., 2001; Laatikainen et al., 2003; Boyle et al., 2008; Molokhia et al., 2011; Holahan et al., 2014; Smyth et al., 2015), and alcohol use disorder (Roerecke & Rehm, 2013; Kendler et al., 2016) are associated with increased all-cause mortality. Nevertheless, this evidence is from observational studies that are prone to confounding and other types of bias (Rothman et al., 2008a; Rehm et al., 2010). Randomized evidence on alcohol drinking and all-cause mortality is limited: in a handful of small studies, brief interventions aimed at reducing alcohol drinking have shown some ability to reduce mortality (Cuijpers et al., 2004; McQueen et al., 2011).

Many observational studies on alcohol drinking and mortality have found what they call a J-shaped or a U-shaped curve (Di Castelnuovo et al., 2006). This means that moderate drinking has been associated with lowest all-cause mortality. Evidence on biological mechanisms supports these epidemiological findings. Alcoholic beverages contain many carcinogenic substances, ethanol itself having the largest impact (Lachenmeier et al., 2012), and alcohol damages the liver through multiple mechanisms (Orman et al., 2013). The biochemical effects of alcohol on cardiovascular health are complex (Fernández-Solà, 2015). Alcohol has a toxic effect on myocardium and promotes cardiac arrhythmias. On the other hand, the direction of some effects is dose-dependent. Low alcohol doses relax blood vessels and have anti-inflammatory effects, but high doses increase blood pressure and inflammation. Low alcohol doses also have a beneficial effect on glucose metabolism, whereas high doses do not. Finally, alcohol has favourable effects on blood lipids, which are important risk factors of coronary heart disease, and on blood clotting (Fernández-Solà, 2015).

While both observational evidence and knowledge about biological mechanisms support the notion of a J-shaped relation between alcohol drinking and mortality, many authors doubt the causality of this J-shaped relation. Instead, they believe that the apparent beneficial effect of moderate drinking is caused by bias due to misclassification error or residual confounding (Knott et al., 2015; Goulden, 2016; Stockwell et al., 2016). They have a point when insisting caution in the interpretation of observational findings. The history of epidemiology provides a good reminder of this: Both numerous observational studies and strong mechanistic evidence lead researchers to believe erroneously that postmenopausal hormone replacement therapy would prevent coronary heart disease. Only after the surprise contradictory findings from randomized trials were they able to detect and correct the bias in the observational studies (Hulley et al., 1998; Manson et al., 2003; Hernán et al., 2008). Thus, even though naive acceptance of observational findings often leads astray, carefully planned and analysed observational studies can produce reliable results (Vandenbroucke, 2004, 2009).
When randomized studies are not possible, natural experiments can be used to strengthen observational evidence for causality (Rutter, 2007). Ideally, triangulation will be used, whereby evidence from multiple different study settings, each of them with different sources of potential bias, will be combined (Lawlor et al., 2016). One powerful tool is Mendelian randomization studies (Davey Smith & Ebrahim, 2003). Another useful tool is discordant-twin design which enables control for the confounding effects of genes and shared family environment (Gesell, 1942; Kujala et al., 2002; McGue et al., 2010). Both genetic predisposition and family environment affect alcohol drinking and, thus, are potential confounders in the association of alcohol drinking with mortality (Leonard & Eiden, 2007; Saraceno et al., 2009; Verhulst et al., 2015). The likelihood of this potential confounding is increased by the fact that alcohol drinking shares familial risk factors with externalizing and internalizing disorders that increase mortality (Kendler et al., 2003; Jokela et al., 2009).

The few discordant-twin studies that have assessed the association of alcohol drinking with mortality have not been conclusive. A seminal study from the United States of America focused on moderate drinking and abstinence (Carmelli et al., 1995). A recent study from a subsample of the same cohort found no association between alcohol drinking and all-cause mortality among monozygotic drinking-discordant twin pairs (Dai et al., 2015). A Finnish study by Kujala et al. (2002) also found no association among monozygotic twin pairs, but their results were based on only 13 monozygotic twin pairs that were discordant for both alcohol drinking and death.

2.3 Summary and open questions

In this literature review, I have reviewed the concept of bias in epidemiology and the current knowledge about the fields of psychiatric epidemiology that are relevant to this study. I have identified three important potential sources of bias in the present body of literature on the risk factors and adverse outcomes of anorexia nervosa and alcohol drinking.

First, the notion of religiosity as a risk factor for anorexia nervosa is largely based on case reports and series, and systematic evidence is sparse (Vandenbroucke, 2001; Dekkers et al., 2012; Bonelli & Koenig, 2013).

Second, parents’ alcohol drinking is a well-known risk factor for alcohol drinking of their children, but it is unknown whether this association is causal (Rossow et al., 2016b, 2016a). Potential confounders that could have caused spurious associations to appear between drinking of parents and their children are local environment, cultural and religious factors, and parental comorbidity and temperament (Rossow et al., 2016b).

Third, there is strong observational evidence that excessive alcohol drinking is associated with increased all-cause mortality (Di Castelnuovo et al., 2006; Zaridze et al., 2014; Jayasekara et al., 2014; Smyth et al., 2015;
Goulden, 2016; Wood et al., 2018), but this evidence mostly does not take into account the potential confounding effects of genetic background and family environment (Kendler et al., 2003; Leonard & Eiden, 2007; Jokela et al., 2009; Saraceno et al., 2009; Verhulst et al., 2015). Discordant-twin studies could adjust for these factors, but evidence from them is sparse (Carmelli et al., 1995; Kujala et al., 2002; McGue et al., 2010; Dai et al., 2015).
3 Aims

I aimed to study whether the potential sources of bias that I have identified from the literature of psychiatric epidemiology will affect the observed associations. Specifically, I aimed to study the following three questions about the associations between potential risk factors and adverse outcomes in the epidemiology of anorexia nervosa and alcohol drinking:

I  Is individual or family religiosity a risk factor for anorexia nervosa on the population level?

II  Can the confounding effects of area of residence, family structure, and fathers’ and mothers’ education, religiosity and personality explain the association of parents’ problem drinking with problem drinking of their adult children?

III  Do the potential confounding effects of genetic background and shared family environment affect the associations of different dimensions of alcohol drinking with all-cause mortality?
4 Methods

This study is based on two prospective, population-based cohorts. The cohorts and the data analysis are described below.

4.1 Participants

4.1.1 FinnTwin16 cohort
The FinnTwin16 cohort was established by identifying from the Finnish Population Information System all twins who were born in Finland in 1975–1979 (Kaprio et al., 2002). A questionnaire was sent to the families when the twins were 16 years old. Returning the family questionnaire implied informed consent to contact the children. Since then, five study waves have been conducted. In wave 1, questionnaires were sent both to the twins (adolescence, age 16 years) and to their fathers and mothers. Follow-up questionnaires (waves 2–4) were sent to the twins at ages 17, 18.5, and 21–28 years (early adulthood, 99.8% of the twins were 22–27 years old). In wave 5, an electronic follow-up questionnaire was used at age 31–37 years (mid-thirties). (Kaprio et al., 2002; Kaprio, 2006, 2013)

Study I is based on adult women (female twins) in waves 1 and 4, and study II is based on the adolescent twins in waves 1, 4 and 5. Information on fathers and mothers was obtained from wave 1 (when the twins were 16-year-old adolescents) both in studies I and II. Response rates were 89.9% in wave 1, 84.5% in wave 4 and 71.9% in wave 5. In wave 1, the father responded in 76.4% and the mother responded in 84.5% of those families whose twins were invited to the study.

4.1.2 The Older Finnish Twin Cohort
The Older Finnish Twin Cohort was established by identifying from the Finnish Population Information System all same-sex twin pairs in Finland who were alive in 1967 and who were born before 1958 (Kaprio et al., 1978). Opposite-sex twin pairs were added to the study in 1996 (Kaprio & Koskenvuo, 2002). Questionnaires were mailed to the twins in 1975, 1981, 1990 and 2011–2012 (Kaprio & Koskenvuo, 2002; Kaprio, 2013).

In study III, I included same-sex twin pairs who answered the 1975 and 1981 questionnaires. Response rates were 89% and 84%, respectively. To enable comparison with an earlier study on alcohol drinking and mortality in the same cohort (Kujala et al., 2002), I used the same inclusion criteria: I studied twins who were 24–60 years old at the end of 1981. To reduce confounding by baseline morbidity, twins who had chronic diseases at the
baseline were excluded. This was done on the basis of the questionnaires and medical register information as of 1 January 1983 (Kujala et al., 2002).

4.1.3 Ethical considerations
In the questionnaires, the twins and their parents were provided with information on the study. Returning the questionnaire implied informed consent. In FinnTwin16 cohort (studies I and II), the twins were minor at the start of the study. Therefore, family questionnaires were first sent to the parents of the twins, and the twins were contacted only after the parents had returned the family questionnaire.

In the FinnTwin16 cohort (studies I and II), the ethics committee of the Department of Public Health of University of Helsinki and the Institutional Review Board (IRB) of Indiana University approved the data collection and analysis. The ethical committee of the Hospital District of Helsinki and Uusimaa approved the data collection in wave 4, and the ethical committee of the Hospital District of Central Finland in wave 5. In the Older Finnish Twin Cohort (study III), the ethics committee of the Department of Public Health, University of Helsinki approved record linkage.

4.2 Measures

4.2.1 Anorexia nervosa
In study I, I defined anorexia nervosa according to the DSM-5 criteria (American Psychiatric Association, 2013). I considered women who fulfilled the diagnostic criteria at the time of, or at any time before, the diagnostic interviews in early adulthood (age 21–28 years) to have lifetime anorexia nervosa.

There was a self-report screen for eating disorder symptoms in the wave 4 questionnaire of the FinnTwin16 cohort (Keski-Rahkonen et al., 2006; Mustelin et al., 2015). The screen included three subscales of the Eating Disorder Inventory 2 (EDI-2) (Garner, 1991): Bulimia, Body Dissatisfaction and Drive for Thinness. 2825 women completed the screen in early adulthood (age 21–28 years).

All screen-positive women (N = 292), their female co-twins (N = 130) and 210 randomly selected women were invited to diagnostic telephone interviews. Four medical doctors and one registered nurse from the Eating Disorder Unit of Helsinki University Central Hospital interviewed 86.7% of the invited women using Structured Clinical Interview for DSM-IV (SCID) (First et al., 2002; Keski-Rahkonen et al., 2006; Mustelin et al., 2015).

Age of symptoms onset was determined on the basis of the interviews (Keski-Rahkonen et al., 2007; Mustelin et al., 2016). After publication of the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5)
in 2013, four experienced medical doctors recoded the interviews and established consensus lifetime DSM-5 anorexia nervosa diagnoses which were used in study I (Mustelin et al., 2016). I considered healthy the women who were not diagnosed with any lifetime eating disorder and who did not have a twin sister with a lifetime eating disorder.

### 4.2.2 Religiosity

The religiosity of fathers and mothers (studies I and II), and of adolescent women (age 16 years, study I) was measured with the Religious fundamentalism content scale of the Minnesota Multiphasic Personality Inventory (MMPI) (Wiggins, 1966; Winter et al., 1999) (Table 1). It emphasizes Christian tenets and measures religious behaviour and beliefs with 12 yes–no items. The scores range from 0 to 12; higher scores reflect higher religiosity. Cronbach’s alphas were 0.82 for mothers, 0.85 for fathers and 0.82 for adolescent women. In study I, I used multiple imputation to impute the religiosity scores for those with missing items. In study II, I included respondents answering to at least nine items, and substituted for the missing items the mean of that respondent’s available items.

Table 1. Items of the Religious fundamentalism content scale of the Minnesota Multiphasic Personality Inventory (MMPI). Wording from (Winter et al., 1999).

1. Everything is turning out just like the prophets of the Bible said it would
2. I go to church almost every week
3. I believe in the second coming of Christ
4. I believe in a life hereafter
5. I am very religious (more than most people)
6. I believe there is a Devil and Hell in the afterlife
7. I believe there is a God
8. I feel sure that there is only one true religion
9. Christ performed miracles such as changing water into wine
10. I pray several times a week
11. I read the Bible several times a week
12. I have no patience with people who believe there is only one true religion

For items 1–11, endorsing ‘True’ yields score 1 and endorsing ‘False’ yields score 0. For item 12 “True” yields 0 and “False” yields 1. Summing scores across items 1–12 yields the Religious fundamentalism content scale.

Religiosity in early adulthood (age 21–28 years, study I) was measured with a multiple-choice item asking “How important do you think religion is in your life?” The available options were: 1) very important, 2) important, 3) not very important, 4) not at all important and 5) cannot tell. Few respondents chose options 1 and 5. Hence, I combined categories 1 and 2, and regarded answers that embraced option 5 as missing information. This yielded a three-category variable: religious, not very religious and not at all religious.

In multiple imputation, I also used two auxiliary variables from the wave 4 questionnaire in early adulthood that were related to religiosity. I analysed
church going frequency in four categories: 1) once a week or once a month, 2) once a year, 3) less often and 4) not at all. Likewise, I analysed the importance of the religiosity of the partner and peers in four categories: 1) very important, 2) important, 3) not very important and 4) not at all important.

4.2.3 Alcohol drinking
The twins (studies II and III) and their parents (study II) reported their alcohol drinking and drinking-related problems in the questionnaires.

Total alcohol consumption
I used self-reports in the 1975 and 1981 questionnaires to calculate average total alcohol consumption in the Older Finnish Twin Cohort (study III). The respondents reported their average weekly consumption of beer (in bottles) and wine or other mild alcoholic beverages (in glasses). They also reported their average monthly consumption of spirits (in bottles) (Kaprio et al., 1987). I estimated one drink (a 330-ml bottle of beer, a 12-cl glass of wine or a 4-cl portion of spirits) to contain 12 g of pure alcohol, and converted the reported total alcohol consumption into monthly alcohol consumption in grams.

Heavy drinking occasions
In study III, I defined heavy drinking occasions as consuming within one occasion more than five bottles of beer, a bottle of wine, or half a bottle of spirits (or a similar amount). This amounts to consuming more than five standard drinks (> 60 g of pure alcohol) on a single occasion. I assessed heavy drinking occasions with a single item whether (or not) the respondents had heavy drinking occasions at least once a month. From the answers in the 1975 and 1981 questionnaires, I formed a three-category variable: no heavy drinking occasions (neither 1975 nor 1981), non-persistent heavy drinking occasions (either 1975 or 1981) and persistent heavy drinking occasions (both 1975 and 1981).

Alcohol-induced blackouts
The 1981 questionnaire in the Older Finnish Twin Cohort also included a multiple-choice question on alcohol-induced loss of consciousness (Kaprio et al., 1987; Paljärvi et al., 2011, 2012), hereafter called blackouts (study III). Pass-out would be a more accurate English translation for the Finnish term sammuminen used in the original questionnaire (Paljärvi et al., 2009), but to be consistent with study III, blackout is used throughout this dissertation. The respondents reported their frequency of blackouts during the past year and I analysed the answers in three categories: 1) none, 2) one and 3) two or more.
**Problem drinking**

I assessed problem drinking with the **Malmö-modified Michigan Alcoholism Screening Test (Mm-MAST)** (Study II). It is a modification of Brief MAST, an abridged version of the Michigan Alcoholism Screening Test, and measures harmful alcohol consumption with nine yes–no items (Kristenson & Trell, 1982). It is associated with total alcohol consumption, intoxication frequency, heavy drinking and problem drinking (Kristenson & Trell, 1982; Seppa *et al.*, 1990, 1992; Nyström *et al.*, 1993). An extended version of Mm-MAST also exists with two additional items intended to improve the scales similarity with alcohol abuse and dependence criteria of the Diagnostic and Statistical Manual of Mental Disorders, 3rd edition – Revised (DSM-III-R), and 4th edition (DSM-IV) (Kaprio *et al.*, 2002).

In study II, fathers’ and mothers’ problem drinking were assessed in wave 1 of the FinnTwin16 cohort with the 9-item Mm-MAST. It yields scores ranging from 0 to 9 (Table 2a). Cronbach’s alphas were 0.69 for fathers’ Mm-MAST and 0.66 for mothers’ Mm-MAST. Cut-offs from 2 to 4 have been suggested for Mm-MAST to identify problem drinking (Kristenson & Trell, 1982; Seppa *et al.*, 1990, 1992; Nyström *et al.*, 1993; Rose *et al.*, 1999a). For this reason, I analysed both fathers’ and mothers’ Mm-MAST in five categories (scores 0, 1, 2, 3 and ≥ 4). In interaction tests, I used a dichotomy (≤ 1 vs ≥ 2) to increase statistical power. Because I found no interactions, I constructed a variable of parental Mm-MAST (with scores ranging from 0 to 8) by summing the categorical scores of fathers’ Mm-MAST (from 0 to 4) and mothers’ Mm-MAST (from 0 to 4).

**Table 2a. Items of Malmö-modified Michigan Alcoholism Screening Test (Mm-MAST)**

1. Do you take a drink before going to a party?
2. Do you usually drink a bottle of wine or corresponding amount of beer or other alcoholic beverages over the weekend?
3. Do you drink a couple of drinks (or beers) a day to relax?
4. Do you tolerate more alcohol now than you did ten years ago?
5. Have you difficulties not drinking more than your friends?
6. Do you fall asleep after moderate drinking without knowing how you got to bed?
7. Do you have a bad conscience after drinking?
8. Do you take a drink (the day after a party) for your hang-over?
9. Do you try to avoid alcoholic beverages for a determined period of time – e.g., a week?

**Table 2b. Items of Malmö-modified Michigan Alcoholism Screening Test, extended, 11-item, lifetime version (Mm-MAST-11)**

1. Do/did you take a drink before going to a party?
2. Do/did you usually drink a bottle of wine or corresponding amount of beer or other alcoholic beverages over the weekend?
3. Do/did you drink a couple of drinks (or beers) a day to relax?
4. Do/did you tolerate more alcohol now than before?
5. Have/had you difficulties not drinking more than your friends?
6. Do/did you fall asleep after moderate drinking without knowing how you got to bed?
7. Do/did you have a bad conscience after drinking?
8. Do/did you take a drink (the day after a party) for your hang-over?
9. Do/did you try to avoid alcoholic beverages for a determined period of time – e.g., a week?
10. *After you have/had taken a drink, do you find it hard to stop?
11. *Have/had you ever felt that anyone close to you thinks that you should drink less?

*This item is not included in Mm-MAST.

I used the extended, 11-item, version of Mm-MAST (Mm-MAST-11, Table 2b) to assess lifetime problem drinking of the adult twins (sons and daughters) in early adulthood (wave 4, age 21–28 years) and during their mid-thirties (wave 5, age 31–37 years) (study II). The two additional items increased the internal consistency of the scale. Cronbach’s alphas were 0.78 and 0.75 for the Mm-MAST-11 of sons and daughters in early adulthood and 0.78 and 0.77 for the Mm-MAST-11 of sons and daughters at mid-thirties, respectively.

I included all responses with no more than two missing items, and substituted for the missing items the mean of that respondent’s available items. This applied for both Mm-MAST and Mm-MAST-11. Further, in wave 1, fathers and mothers were instructed to skip the entire scale if they did not drink at all. Therefore, they received score zero if all items were missing and they did not drink alcohol during the past year. The determinants of abstinence, however, are different from the determinants of drinking habits among those who drink (Rose et al., 1999a; Maes et al., 1999; Rose et al., 2001; Viken et al., 2007). Hence, in study II, I excluded all fathers, mothers and offspring sons and daughters from the analyses if they were lifetime abstainers. Consequently, the analyses consisted of sons and daughters who themselves, and whose both parents, had drunk alcohol during their lifetime. I, however, tested the effects of these exclusions by conducting sensitivity analyses in which lifetime abstainers were included.

**Drinking frequency**

I assessed drinking frequency of the offspring children (twins) in wave 1 of the FinnTwin16 cohort (baseline of study II). At that time, the children were 16-year-old adolescents. The multiple-choice item included nine categories: 1) daily, 2) a couple of times a week, 3) once a week, 4) a couple of times a month, 5) about once a month, 6) about once every two months, 7) 3–4 times a year, 8) once a year or less and 9) I don’t drink any alcohol. To avoid categories with few respondents, I analysed this variable in six categories: 1) I do not drink alcohol at all, 2) once a year or less, 3) 3–4 times a year or about once every two months, 4) about once a month, 5) a couple of times a month and 6) once a week or more often. Those who reported never having drunk beer, wine, long drinks or spirits were considered lifetime abstainers and excluded.
Grandparents’ drinking
Fathers and mothers of the offspring sons and daughters reported drinking habits of their parents (grandparents of the sons and daughters) (study II). I regarded any grandparent as a regular drinker if he or she had drunk alcohol quite regularly as an adult. All other answers (no drinking or drinking very rarely, drinking occasionally and cannot say) implied low drinking. I validated this measure of grandparents’ drinking by testing that it was positively associated with problem drinking (Mm-MAST scores) of their children (the parents of the offspring sons and daughters).

4.2.4 Zygosity
The zygosity of the twin pairs was determined with a validated questionnaire method (Sarna et al., 1978) (study III). The twin pairs were classified to monozygotic (MZ) pairs, dizygotic (DZ) pairs and pairs with unknown zygosity.

4.2.5 All-cause mortality
I followed all-cause mortality until 31 December 2011 (study III). Dates of death were retrieved from the Finnish Population Information System on 29 February 2012.

4.2.6 Covariates
Age of study participants was available from the Finnish Population Information system (studies I–III). In study III, I used time since birth as the time-variable in the survival models to achieve exact adjustment for age.

I used education as a proxy for socioeconomic status (studies I–III). I analysed women’s (study I, wave 4), and fathers’ and mothers’ (study II, wave 1) self-reported education as a dichotomy: academic (at least completed high school) vs non-academic (Lajunen et al., 2012). In study III, I had information on self-reported education from the 1975 questionnaire. I grouped the answers in three categories to approximate primary, secondary and tertiary education.

In study III, I also had information on social class from the 1975 questionnaire. Self-reported occupations were classified according to the classification of Statistics Finland. The categories were 1) upper white-collar workers and comparable entrepreneurs, 2) lower white-collar workers and small entrepreneurs, 3) skilled workers, 4) non-skilled workers, 5) farmers and fishermen and 6) unknown, including students, homemakers and pensioners (Statistics Finland, Department of Population Statistics, 1972).

Marital status was available from the 1975 questionnaire in study III. I analysed it dichotomously: single, divorced or separated vs married; divorced and remarried; cohabiting; or widowed.
The sons and daughters reported their family structure in adolescence (study II, wave 1, age 16 years). I analysed family structure as a dichotomy: living vs not living with both biological parents.

**Area of residence** was retrieved from the Finnish Population Information System in the last year of wave 1 (studies I and II). I classified it according to the European Union’s Nomenclature of Territorial Units (Statistics Finland, 1998; Rose *et al.*, 1999). In study I, I used the same two categories than a previous study on Finnish religiosity (Winter *et al.*, 1999). I contrasted the less religious and more densely populated Southern Finland with the more religious and less densely populated Northern Finland and excluded the Åland Islands. In study II, I analysed area of residence in three categories relevant to alcohol drinking and drinking problems: the capital area, Mid-Finland (comprising the former provinces of Vaasa and Central Finland), and the rest of Finland. The capital area is characterized by high population density, high urbanization, a high proportion of highly educated people, high average alcohol consumption and high availability of alcoholic beverages. The former province of Vaasa, on the Western coast in Mid-Finland, is characterized by lower average alcohol consumption, higher religiosity and a higher proportion of Swedish speaking people with their own cultural features (Simpura & Lahti, 1988).

I assessed one dimension of fathers’ and mothers’ personality that is relevant to alcohol drinking in study II. The social deviance scale (Pd or “Psychopathic deviate” scale of the MMPI) measures social deviance with 50 yes–no items and is associated with alcohol drinking and problems (Mustanski *et al.*, 2003; Viken *et al.*, 2007). Because externalizing disorders have a partly common genetic background, it also captures part of the genetic risk for externalizing disorders and drinking problems in the family (Kendler *et al.*, 2003). I included the respondents who answered at least 40 items, and substituted for the missing items the mean score of that respondent’s available items.

The respondents reported their cigarette smoking in the 1981 questionnaire of the Older Finnish Twin Cohort (study III). I analysed cigarette smoking in nine categories: 1) never-smoker, 2) ex-smoker, 3) occasional smoker or a current smoker smoking less than five cigarettes per day and 4) current smokers smoking 5–9, 5) 10–14, 6) 15–19, 7) 20–24, 8) 25–39 and 9) ≥ 40 cigarettes per day.

The respondents reported how vigorous was they leisure-time physical activity in both the 1975 and 1981 questionnaires of the Older Finnish Twin Cohort (Kujala *et al.*, 2002) (study III). I considered active those whose leisure-time physical activity was as strenuous as “alternately walking and jogging”. This yielded a variable with three categories: persistently non-active (active in neither 1975 nor 1981), non-persistently active (active in either 1975 or 1981) and persistently active (active in both 1975 and 1981).

The respondents reported their height and weight in the 1981 questionnaire of the Older Finnish Twin Cohort (study III). According to the classification of
World Health Organization (World Health Organization, 2000), I defined **obesity** as a body mass index (body mass in kg / (length in m)²) ≥ 30 kg/m².

I assessed **life satisfaction** with four Likert scale items asking whether the respondent perceived his or her life “at present” as interesting or boring, happy or sad, easy or hard and lonely or not (study III). This yielded a score ranging from 4 to 20. I grouped the score into three categories: satisfied (score 4–6), intermediate (7–11) and dissatisfied (12–20) (Koivumaa-Honkanen et al., 2000).

### 4.3 Statistical methods

I analysed the data with the Stata statistical software package (StataCorp LP, College Station, TX). I used release 12 in study III and release 13 in studies I and II. Because the data consist of twin pairs, the twin individuals in a pair are not statistically independent observations. In other words, the value of one twin’s variable is informative of some fraction of the value of the other twin’s variable, reflected as an overall correlation between twins. Therefore, I adjusted confidence intervals and p-values for clustering within twin pairs using a generalization of heteroskedasticity-consistent standard errors, also known as Hubert–White standard errors and robust standard errors (Rogers, 1994; Williams, 2000). Throughout this study, I report 95% confidence intervals and two-sided p-values.

#### 4.3.1 Means

I estimated mean Mm-MAST-11 with multiple linear regression (study II). I used robust variance estimators to get unbiased confidence intervals despite heteroskedasticity.

#### 4.3.2 Correlations

I used Spearman’s rank correlation coefficients (rho) to estimate correlations (studies I and II). In study II, I also calculated polychoric correlations between ordinal variables (Kolevnikov & Angeles, 2004). I assessed the assumption of underlying bivariate normality with Pearson’s chi-squared tests. P-values less than 0.05 suggest violation of the assumption and, hence, call for caution in interpretation of the estimated polychoric correlations. Sons’ and daughters’ Mm-MAST-11 (study II) was assessed continuously. Therefore, I estimated polyserial correlations between it and ordinal variables (Kolevnikov & Angeles, 2004).
4.3.3 Multiple imputation

To reduce the risk of selection bias, I imputed missing values with multiple imputation in study I (Sterne et al., 2009; Hernán & Robins, 2018). I used a multiple imputation method called substantive-model compatible fully conditional specification (SMC-FCS) (Bartlett & Morris, 2015; Bartlett et al., 2015). It is an updated version of multiple imputation with chained equations that is modified to ensure compatibility between the imputation model and the substantive model (i.e. the model with which the data are analysed).

Twin individuals are not independent of each other, and cluster correction was not available for multiple imputation in standard software packages at the time of the data analysis. Therefore, I divided each twin pair to separate datasets. This yielded two datasets (A and B) that consisted of independent observations. These datasets, however, most definitely were not independent of each other. Fathers and mothers of complete twin pairs were even included in both of them.

For the single question on personal religiosity in early adulthood, I had a high proportion (94–95%) of complete information. Proportion of complete information was lower for religiosity in adolescence and for parents’ religiosity (59–70%). Complete information for all variables that I used in multiple imputation was available for 22–28% of women in the datasets.

The substantive model compared women with lifetime anorexia nervosa to healthy women. Therefore, before multiple imputation, I excluded women who did not have lifetime anorexia nervosa but were not healthy either. Consequently, I excluded women who had other lifetime eating disorders than anorexia nervosa. I also excluded twin sisters of the women with a lifetime eating disorder diagnosis, because they often had some eating disorder symptoms. Finally, to be consistent with an earlier study on Finnish religiosity, I excluded the few women living on the Åland Islands (Winter et al., 1999).

After these exclusions, the final dataset A consisted of 1312 women of whom 38 had a diagnosis of lifetime anorexia nervosa. The final dataset B consisted of 1327 women of whom 53 had a diagnosis of lifetime anorexia nervosa. Of the women with lifetime anorexia nervosa, 23 and 29 had the onset of their disease after the baseline measurement of religiosity (at the age of 16 years) in datasets A and B, respectively.

I included personal, maternal and paternal religiosity, place of residence, education and lifetime anorexia nervosa status in the multiple imputation models. In addition, I included interactions between religiosity variables in wave 1 and place of residence in wave 1. I used two auxiliary religiosity variables from wave 4 to improve the quality of the imputed values: church going frequency and the importance of the religiosity of partner and peers.
I did multiple imputation separately for each dataset. I based analyses on the two different outcomes (lifetime anorexia nervosa and anorexia nervosa with onset after the age of 16 years) on separate multiple imputation models with separate substantive models to ensure compatibility between the imputation model and the substantive model. I used 100 burn in iterations for each imputation to ensure proper convergence of the chained equations and 200 imputations per multiple imputation model to achieve low Monte Carlo errors. Finally, I ensured the quality of the imputations by graphically evaluating the convergence of the burn in iterations, by comparing the distributions of the observed and imputed values, and by examining the Monte Carlo errors as recommended by White et al. (2011).

4.3.4 Risk of lifetime anorexia nervosa
I modelled the risk of lifetime anorexia nervosa with logistic regression (study I). We used Rubin's rules to combine odds ratios and their 95% confidence intervals across the imputed datasets (Rubin, 1987; White et al., 2011).

4.3.5 Survival analysis
I used the Cox proportional hazards model in survival analysis (study III). I assessed the proportional hazards assumption with scaled Schoenfeld residuals, link test and graphical methods (UCLA: Statistical Consulting Group, n.d.; Pregibon, 1979).

Comparisons within twin pairs
Co-twins typically share their childhood environment. Dizygotic twins share on average 50% of their segregating genes and monozygotic twins are genetically identical at the sequence level (apart from sporadic mutations). Therefore, comparisons within all twin pairs adjust for shared family environment and partly for genetic effects, and comparisons within monozygotic twin pairs adjust both for shared family environment and genetic effects (Gesell, 1942; Kujala et al., 2002; McGue et al., 2010).

I used pairwise analyses within all twin pairs and within monozygotic twin pairs to examine whether adjustments for shared family background and genetic factors would change the effect estimates (study III). I applied dichotomous drinking measures, and defined those twin pairs as drinking-discordant in which the co-twins were in different drinking categories. I used Cox proportional hazard models stratified by family (i.e. twin pair) to analyse whether the more drinking (exposed) or the less-drinking (unexposed) co-twin was more likely to die first.
4.3.6 Adjustments for covariates

In study I, I analysed women only, because there were not enough men with anorexia nervosa to enable analyses in men (Raevuori et al., 2009). In supplementary analyses, I tested whether adjustments for education would affect the results.

In study II, I analysed men and women separately. Partial adjustment for age was inherent to data collection, which was done in waves at defined ages. I also did additional adjustments for fathers’, mothers’, sons’ and daughters’ age, but this had only a very small effect on the results. In multiple linear regression models, I adjusted our results for area of residence, family structure, and fathers’ and mothers’ education, religiosity and one relevant dimension of personality (the social deviance scale). I modelled religiosity and personality with restricted cubic splines using three knots at 10th, 50th and 90th percentiles (Harrell, 2015). This took account of nonlinear relations and, thus, reduced residual confounding. In supplementary analyses, I also did additional adjustments for grandparents’ drinking and sons’ and daughters’ drinking frequency in adolescence (at the age of 16 years).

In study III, I adjusted all results for age by using time since birth as the time variable in survival models. Exact adjustment for age was inherent to pairwise twin comparisons, because co-twins are born at the same time. When building the model (UCLA: Statistical Consulting Group, n.d.; Greenland & Rothman, 2008), I included smoking, sex, obesity, physical activity, social class, education, marital status and life satisfaction as potential confounders. I used likelihood-ratio tests with p < 0.10 as a cut-off. Consequently, I dropped life satisfaction from the final models. I also tested for interactions. Sex had no statistically significant interactions. Therefore, to increase statistical power, I analysed men and women together with adjustments for sex.
5 Results

5.1 Descriptive results

In adolescence (at the age of 16 years), mean religiosity of the women on the Religious fundamentalism content scale of the MMPI was 4.5 and 4.4 [standard deviation (SD) = 2.9] in datasets A and B, respectively. In early adulthood (at the age of 21–28 years), 29–30% of the women were religious (study I, Table 3).

Mean Mm-MAST-11 of the sons was 4.6 (SD = 2.8) in early adulthood and 4.4 (SD = 2.9) at mid-thirties (age 31–37 years). For daughters, the means were 3.3 (SD = 2.4) and 3.0 (SD = 2.5), respectively (study II, Table 4).

In study III, men drank more than women (Table 5). Almost half (49%) of men and 12% of women had a total alcohol consumption over mean (≥ 259 g/month, more than about 5 drinks per week), 32% of men and 5% of women were persistent heavy drinkers, and 13% of men and 2% of women had alcohol-induced blackouts at least twice a year.

Table 3. Religiosity of the women in the FinnTwin16 cohort (study I).

<table>
<thead>
<tr>
<th></th>
<th>Dataset A</th>
<th>Dataset B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>Religiosity in adolescence (age 16 years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1312</td>
<td>4.5</td>
</tr>
<tr>
<td>Fathers’ religiosity</td>
<td>1312</td>
<td>3.9</td>
</tr>
<tr>
<td>Mothers’ religiosity</td>
<td>1312</td>
<td>5.0</td>
</tr>
<tr>
<td>Religiosity in early adulthood (age 21–28 years)</td>
<td>N*</td>
<td>Proportion (of total N 1312)</td>
</tr>
<tr>
<td>-not at all religious</td>
<td>302</td>
<td>0.23</td>
</tr>
<tr>
<td>-not very religious</td>
<td>620</td>
<td>0.47</td>
</tr>
<tr>
<td>-religious</td>
<td>389</td>
<td>0.30</td>
</tr>
</tbody>
</table>

Religiosity in adolescence, mother’s religiosity and father’s religiosity measured with the Religious fundamentalism content scale of the Minnesota Multiphasic Personality Inventory when the women were 16-year-old adolescents. 95% CI, 95% confidence interval; SD, standard deviation. * N for categories of religiosity in early adulthood estimated from multiple imputation and rounded to nearest integer.
Table 4. Sons and daughters in the FinnTwin16 cohort (study II).

<table>
<thead>
<tr>
<th>Sons</th>
<th>Daughters</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N</strong></td>
<td><strong>Mean (95% CI)</strong></td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-------------------------------------------</td>
</tr>
<tr>
<td>Problem drinking (Mm-MAST-11) in early adulthood (age 21–28 years)</td>
<td>Problem drinking (Mm-MAST-11) at mid-thirties (age 31–37 years)</td>
</tr>
<tr>
<td>2309</td>
<td>4.6 (4.5–4.7)</td>
</tr>
<tr>
<td>Problem drinking (Mm-MAST-11) at mid-thirties (age 31–37 years)</td>
<td>1874</td>
</tr>
<tr>
<td>Heavy drinking occasions (per year) in early adulthood (age 21–28 years)</td>
<td>Heavy drinking occasions (per year) at mid-thirties (age 31–37 years)</td>
</tr>
<tr>
<td>2316</td>
<td>28.0 (26.7–29.4)</td>
</tr>
<tr>
<td>1891</td>
<td>24.5 (22.9–26.4)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Father’s problem drinking (Mm-MAST)</th>
<th>Mother’s problem drinking (Mm-MAST)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N</strong></td>
<td><strong>%</strong></td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>1818</td>
<td>2110</td>
</tr>
<tr>
<td>0</td>
<td>497</td>
</tr>
<tr>
<td>1</td>
<td>370</td>
</tr>
<tr>
<td>2</td>
<td>309</td>
</tr>
<tr>
<td>3</td>
<td>253</td>
</tr>
<tr>
<td>≥4</td>
<td>389</td>
</tr>
<tr>
<td>1856</td>
<td>2189</td>
</tr>
<tr>
<td>0</td>
<td>847</td>
</tr>
<tr>
<td>1</td>
<td>440</td>
</tr>
<tr>
<td>2</td>
<td>265</td>
</tr>
<tr>
<td>3</td>
<td>140</td>
</tr>
<tr>
<td>≥4</td>
<td>164</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Father’s heavy drinking occasions</th>
<th>Mother’s heavy drinking occasions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1854</td>
<td>2148</td>
</tr>
<tr>
<td>-never</td>
<td>308</td>
</tr>
<tr>
<td>-once a year or less often</td>
<td>246</td>
</tr>
<tr>
<td>-a few times a year</td>
<td>569</td>
</tr>
<tr>
<td>-about once a month</td>
<td>336</td>
</tr>
<tr>
<td>-about once a week or more often</td>
<td>395</td>
</tr>
<tr>
<td>1902</td>
<td>2230</td>
</tr>
<tr>
<td>-never</td>
<td>927</td>
</tr>
<tr>
<td>-once a year or less often</td>
<td>327</td>
</tr>
<tr>
<td>-a few times a year</td>
<td>387</td>
</tr>
<tr>
<td>-about once a month</td>
<td>163</td>
</tr>
<tr>
<td>-about once a week or more often</td>
<td>98</td>
</tr>
</tbody>
</table>

Lifetime abstainers excluded. CI, confidence interval; SD, standard deviation; Mm-MAST, Malmö-modified Michigan Alcoholism Screening Test (original 9-item version); Mm-MAST-11, Malmö-modified Michigan Alcoholism Screening Test (extended 11-item version).
Table 5. Alcohol drinking in the Older Finnish Twin Cohort (study III).

<table>
<thead>
<tr>
<th>Monthly alcohol consumption in grams (mean of 1975 and 1981)</th>
<th>Men</th>
<th>Women</th>
<th>Total</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>376</td>
<td>928</td>
<td>1334</td>
<td>9.0</td>
</tr>
<tr>
<td>1–69</td>
<td>671</td>
<td>1885</td>
<td>2556</td>
<td>17.3</td>
</tr>
<tr>
<td>70–139</td>
<td>1173</td>
<td>2273</td>
<td>3446</td>
<td>23.3</td>
</tr>
<tr>
<td>140–209</td>
<td>1144</td>
<td>1104</td>
<td>2248</td>
<td>15.2</td>
</tr>
<tr>
<td>210–419</td>
<td>1610</td>
<td>787</td>
<td>2397</td>
<td>16.2</td>
</tr>
<tr>
<td>420–839</td>
<td>1573</td>
<td>317</td>
<td>1890</td>
<td>12.8</td>
</tr>
<tr>
<td>840–1199</td>
<td>487</td>
<td>35</td>
<td>522</td>
<td>3.5</td>
</tr>
<tr>
<td>≥ 1200</td>
<td>369</td>
<td>25</td>
<td>394</td>
<td>2.7</td>
</tr>
<tr>
<td>0–258</td>
<td>3811</td>
<td>6502</td>
<td>10313</td>
<td>69.7</td>
</tr>
<tr>
<td>≥ 259</td>
<td>3592</td>
<td>882</td>
<td>4474</td>
<td>30.3</td>
</tr>
<tr>
<td>Heavy drinking occasions at least monthly</td>
<td>7403</td>
<td>7384</td>
<td>14787</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>3333</td>
<td>6222</td>
<td>9555</td>
<td>64.6</td>
</tr>
<tr>
<td>1975 or 1981</td>
<td>1720</td>
<td>814</td>
<td>2534</td>
<td>17.1</td>
</tr>
<tr>
<td>1975 and 1981</td>
<td>2350</td>
<td>348</td>
<td>2698</td>
<td>18.2</td>
</tr>
<tr>
<td>Alcohol–induced blackouts (per year)</td>
<td>7403</td>
<td>7384</td>
<td>14787</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>5732</td>
<td>6892</td>
<td>12624</td>
<td>85.4</td>
</tr>
<tr>
<td>1</td>
<td>700</td>
<td>319</td>
<td>1019</td>
<td>6.9</td>
</tr>
<tr>
<td>≥ 2</td>
<td>971</td>
<td>173</td>
<td>1144</td>
<td>7.7</td>
</tr>
</tbody>
</table>

5.2 Religiosity and anorexia nervosa (study I)

The analyses of religiosity and lifetime anorexia nervosa were based on 1312 women of whom 38 had lifetime anorexia nervosa in dataset A and 1327 women of whom 53 had lifetime anorexia nervosa in dataset B (study I). Of the women with lifetime anorexia nervosa, 23 (dataset A) and 29 (dataset B) had the onset of their symptoms after baseline measurement of religiosity in adolescence (at the age of 16 years).

5.2.1 Assessment of selection bias

To assess the potential of missing data to introduce selection bias into the results, I compared the results from the imputed datasets to those of the complete case analyses. These results were very similar; this decreases the probability that missing data distorted the results. Therefore, in the following paragraphs, I will present the results from the imputed datasets only.
5.2.2 Fathers’ and mothers’ religiosity

Fathers’ and mothers’ religiosity were not associated with the risk of lifetime anorexia nervosa among their female daughters (Figure 1).

**Figure 1.** The associations of fathers’, mothers’ and personal religiosity in adolescence (at the age of 16 years) with the risk of lifetime anorexia nervosa among Finnish women. Religiosity was measured with the Religious fundamentalism content scale of the Minnesota Multiphasic Personality Inventory (MMPI). Lifetime anorexia nervosa was diagnosed according to the DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, 5th edition). N = 1312 with 38 lifetime anorexia nervosa diagnoses in dataset A, and N = 1327 with 53 lifetime anorexia nervosa diagnoses in dataset B.

5.2.3 Personal religiosity

Personal religiosity in adolescence (at the age of 16 years) was not associated with lifetime anorexia nervosa (Figure 1). Personal religiosity in early adulthood (at the age of 21–28 years) showed a U-shaped association with lifetime anorexia nervosa. In other words, both low and high religiosity seemed to be associated with lifetime anorexia nervosa, but these associations were not statistically significant (Figure 2).

I also conducted a subgroup analysis where I looked at those women whose symptoms of anorexia nervosa had started after the baseline measurement of religiosity in adolescence (at the age of 16 years). I discovered that personal religiosity in adolescence was not associated with an onset of anorexia nervosa later in life. The odds ratios for unit increase in religiosity were 0.94 (95%
confidence interval (CI) 0.79–1.13, p = 0.52) in dataset A and 1.07 (95% CI 0.93–1.24, p = 0.33) in dataset B.

Finally, I tested whether adjustment for education would affect the results, but they remained similar.

**Figure 2.** The associations of personal religiosity in early adulthood (age 21–28 years) with the risk of lifetime anorexia nervosa among Finnish women. Religiosity was measured with a single item assessing the subjective importance of religion in the respondent’s life. Lifetime anorexia nervosa was diagnosed according to the DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, 5th edition). N = 1312 with 38 lifetime anorexia nervosa diagnoses in dataset A, and N = 1327 with 53 lifetime anorexia nervosa diagnoses in dataset B. * Reference category.

5.3 Parents’ and their children’s problem drinking (study II)

High combined parental problem drinking (high parents’ Mm-MAST) showed a clear association with problem drinking (high Mm-MAST-11) of their adult children (study II). This was observable in both sons and daughters, and both in early adulthood and at mid-thirties. Statistical evidence for a linear trend was strong (p for linear trend < 0.001). I assessed area of residence, family structure, and fathers’ and mothers’ education, religiosity and one relevant dimension of personality as potential confounders. Adjustments for these potential confounders had but a very small effect on the results. Consequently, I report the fully adjusted estimates with the number of complete cases in each category of parental problem drinking in Figures 3 and 4. I also conducted
sensitivity analyses in which lifetime abstainers were included: in them, the observed associations were somewhat stronger. Thus, if exclusion of lifetime abstainers from the main analyses biased the results, the bias was towards null.

Figure 3. The effect of parents’ problem drinking (Mm-MAST) on problem drinking (mean Mm-MAST-11) of their sons in early adulthood (age 21–28 years) and at mid-thirties (age 31-37 years). Parents’ problem drinking was measured at baseline when the sons were 16-year-old adolescents. Results adjusted for area of residence, family structure, and fathers’ and mothers’ education, religiosity and one relevant dimension of personality. Total N with complete information was 1235 in early adulthood and 991 at mid-thirties. * Reference category.
5.3.1 Drinking frequency in adolescence – a partial mediator

To study mediation, I did an additional adjustment for sons’ and daughters’ drinking frequency in adolescence (at baseline at the age of 16 years). This reduced, but did not remove, the associations of parental problem drinking with sons’ and daughters’ problem drinking in early adulthood and at mid-thirties (Table 6).

Parental problem drinking was associated with both sons’ and daughters’ drinking frequency in adolescence. Polychoric correlation between parental problem drinking (Mm-MAST) and sons’ drinking frequency in adolescence was 0.17 (Pearson’s chi squared test p = 0.25, Spearman’s rho = 0.17). Polychoric correlation between parental problem drinking (Mm-MAST) and daughters’ drinking frequency in adolescence was likewise 0.17 (Pearson’s chi squared test p = 0.02, Spearman’s rho = 0.16).

Sons’ and daughters’ drinking frequency in adolescence predicted their problem drinking (Mm-MAST-11) in early adulthood and at mid-thirties. For sons, polyserial correlations were 0.26 (Spearman’s rho = 0.29) between drinking frequency in adolescence and problem drinking in early adulthood,
and 0.20 (Spearman’s rho = 0.22) between drinking frequency in adolescence and problem drinking at mid-thirties. For daughters, they were 0.30 (Spearman’s rho = 0.26) and 0.26 (Spearman’s rho = 0.25), respectively.

To sum this evidence up, the effect of parental problem drinking on their children’s problem drinking in adulthood seems to be partly mediated by the children’s drinking patterns during adolescence (Baron & Kenny, 1986).

**Table 6.** Combined parental problem drinking (Mm-MAST) as a predictor of problem drinking (Mm-MAST-11) in early adulthood (age 21–28 years) and at mid-thirties (age 31–37 years) with and without adjustment for drinking frequency at age 16.

<table>
<thead>
<tr>
<th></th>
<th>Sons’ Mm-MAST-11 in early adulthood (Model A)</th>
<th>Sons’ Mm-MAST-11 in early adulthood (Model B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Adjusted mean</td>
</tr>
<tr>
<td><strong>Parental Mm-MAST</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0*</td>
<td>168</td>
<td>4.45</td>
</tr>
<tr>
<td>1</td>
<td>146</td>
<td>4.71</td>
</tr>
<tr>
<td>2</td>
<td>169</td>
<td>4.75</td>
</tr>
<tr>
<td>3</td>
<td>136</td>
<td>5.13</td>
</tr>
<tr>
<td>4</td>
<td>158</td>
<td>5.09</td>
</tr>
<tr>
<td>5</td>
<td>84</td>
<td>5.15</td>
</tr>
<tr>
<td>6</td>
<td>88</td>
<td>5.96</td>
</tr>
<tr>
<td>7</td>
<td>61</td>
<td>5.71</td>
</tr>
<tr>
<td>8</td>
<td>40</td>
<td>5.80</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Daughters’ Mm-MAST-11 in early adulthood (Model A)</th>
<th>Daughters’ Mm-MAST-11 in early adulthood (Model B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Adjusted mean</td>
</tr>
<tr>
<td><strong>Parental Mm-MAST</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0*</td>
<td>206</td>
<td>3.32</td>
</tr>
<tr>
<td>1</td>
<td>190</td>
<td>3.43</td>
</tr>
<tr>
<td>2</td>
<td>173</td>
<td>3.59</td>
</tr>
<tr>
<td>3</td>
<td>177</td>
<td>3.64</td>
</tr>
<tr>
<td>4</td>
<td>200</td>
<td>3.53</td>
</tr>
<tr>
<td>5</td>
<td>120</td>
<td>3.98</td>
</tr>
<tr>
<td>6</td>
<td>99</td>
<td>3.62</td>
</tr>
<tr>
<td>7</td>
<td>55</td>
<td>4.68</td>
</tr>
<tr>
<td>8</td>
<td>45</td>
<td>4.09</td>
</tr>
</tbody>
</table>
5.3.2 Grandparents’ drinking – a proxy for genetic predisposition

In supplementary analyses, I also adjusted the results for grandparents’ alcohol drinking – a proxy for genetic predisposition to alcohol drinking and problem drinking in the family. This had but a negligible effect on the results (results not shown).
5.4 Alcohol drinking and all-cause mortality (study III)

5.4.1 Mortality follow-up
A total number of 15607 twin individuals were available for the mortality follow-up in study III. They were 24–60 years old at the end of 1981. I had complete data on 14787 of them (95%). These data comprised 1569 monozygotic twin pairs, 3130 dizygotic twin pairs and 299 twin pairs of uncertain zygosity. In total, I had 4998 same-sex twin pairs and 4791 twin individuals without data on their co-twins.

The follow-up of all-cause mortality started when the 1981 questionnaires were returned. I continued it until censoring (N = 106, due to moving abroad), death (N = 2203) or 31 December 2011 (N = 12478). This yielded a follow-up of 421153 person-years with a median length of 30.2 years (range 2 days – 30.2 years).

5.4.2 Individual-level analyses
When I analysed the twins as individuals, total alcohol consumption showed a clear dose-dependent association with all-cause mortality. Already those who drank 210–419 grams of pure alcohol per month had a higher mortality risk than the lightest drinkers who drank 1–69 g/month. This corresponds to drinking about 4–8 drinks per week vs about 0–1 drinks per week. Higher levels of drinking were consistently associated with higher risks. Abstainers’ risk was similar to the risk of lightest drinkers. We report detailed results with adjustments for age, sex, smoking, physical activity, obesity, education, social class and marital status in Figure 5.

Persistent heavy drinking occasions increased all-cause mortality (Figure 6). So did alcohol-induced blackouts at least twice a year (Figure 6). The hazards associated with blackouts were not entirely proportional as they diminished somewhat towards the end of the follow-up. This may be because of the tendency of heavy drinking people to decrease their drinking with increasing age (Knott et al., 2018). I also found statistical evidence for effect measure modification; among the few non-smokers who had blackouts at least twice a year, blackouts yielded higher hazard ratios than among the entire cohort.
Figure 5. The effect of total alcohol consumption on the risk of all-cause mortality in individual-level analyses in a 30-year follow-up. Results adjusted for age, sex, smoking, physical activity, obesity, education, social class and marital status. N = 14787 with 2203 deaths during the follow-up. * Reference category.
Figure 6. Different dimensions of alcohol drinking and the risk of all-cause mortality in a 30-year follow-up. Comparisons between analyses on the individual level, among all drinking-discordant twin pairs and among drinking-discordant monozygotic twin pairs. Results adjusted for age, sex, smoking, physical activity, obesity, education, social class and marital status. N = 14787 with 2203 deaths during the follow-up.

5.4.3 Comparisons within twin pairs

Analyses within twin pairs yielded hazard ratios that were comparable to the hazard ratios from individual-level analyses (Figure 6). This was especially true for total alcohol consumption over mean (≥ 259 g/month, more than about 5 drinks per week) vs below mean (0–258 g/month, about 0–5 drinks/week), and for alcohol-induced blackouts (≥ 2/year vs. 0/year). For persistent heavy drinking occasions (both 1975 and 1981 vs neither 1975 nor 1981), the hazard ratio within all twin pairs was comparable to the hazard ratio in individual-level analyses, but the hazard ratio within monozygotic twin pairs was not statistically significant despite a comparable point estimate. This may be because there were only 30 monozygotic twin pairs who were discordant for persistent heavy drinking occasions and in which at least one of the co-twins had died during the follow-up. Similarly with individual-level analyses, I adjusted the pairwise analyses for age, sex, smoking, physical activity, obesity, education, social class and marital status.
6 Discussion

6.1 Main findings

I have identified some potential sources of bias from the literature of psychiatric epidemiology, and examined whether they affect the observed associations. Specifically, I have examined three associations between potential risk factors and adverse outcomes in the epidemiology of anorexia nervosa and alcohol drinking.

This led to three main findings: First (study I), in a systematic study using multiple imputation to reduce selection bias, religiosity did not seem to be a major risk factor contributing to the risk of anorexia nervosa in modern Finland. Given the many case reports that suggest the opposite, this underscores the importance of systematic evidence. Second (study II), potential confounders, that I identified from literature, could not explain the association of parents’ problem drinking with problem drinking of their adult children. Third (study III), the potential confounding effects of childhood family environment and genetic factors could not explain the associations of total alcohol consumption of at least 259 grams per month (more than about 5 drinks per week) and alcohol-induced blackouts with all-cause mortality.

In the following sections, I will discuss these three main findings in detail. I will also discuss the strengths and limitations of my study and propose directions for further research.

6.2 Religiosity and anorexia nervosa

In study I, I studied the associations of personal and family religiosity with lifetime anorexia nervosa. I conducted a population-based cohort study to acquire systematic evidence. I also imputed missing data with multiple imputation to reduce the risk of selection bias (Sterne et al., 2009; Hernán & Robins, 2018). I did not find evidence that either higher parental religiosity or personal religiosity is significantly associated with lifetime anorexia nervosa. Religiosity in early adulthood had a U-shaped association with lifetime anorexia nervosa, but this finding was not statistically significant. Further, religiosity in adolescence (at the age of 16 years) neither was associated with lifetime anorexia nervosa nor increased the risk of anorexia nervosa onset in the future.

If a true U-shaped association between religiosity in early adolescence and lifetime anorexia nervosa were to exist, this could reflect three possibilities. For one thing, both high and low religiosity could increase the risk of anorexia nervosa. Second, a severe illness like anorexia nervosa may have an effect on religiosity (Joughin et al., 1992; Rogers et al., 2002). Third, some underlying
factor that affects both religiosity and the risk of anorexia nervosa could cause a U-shaped association to appear between religiosity and anorexia nervosa.

My negative results fit together with the evidence from earlier studies. Two modern population studies exist. In Canada, religiosity and eating disorder symptoms did not have a direct association (Boisvert & Harrell, 2013). In the United States of America, a big study reported beneficial effects of religiosity on mental health. More specifically, the negative effects of eating disorders on self-esteem were mitigated by religious attendance and prayer (Henderson & Ellison, 2015).

Evidence from other recent studies is diverse: some argue that religiosity may protect from body dissatisfaction and disordered eating (Gluck & Geliebter, 2002; Homan & Boyatzis, 2010; Feinson & Meir, 2012b), others are of the opposite mind (Wilbur & Colligan, 1981; Gates & Pritchard, 2009; Thomas et al., 2018). Still some studies could not find an association between religiosity and disordered eating (Pinhas et al., 2008; Feinson & Meir, 2012a, 2014), and some studies found differences between different denominations and intrinsic motivations for religiosity (Sykes et al., 1988; Smith et al., 2004; Kim, 2006, 2007; Castellini et al., 2014; Akrawi et al., 2015).

Although case reports of religious women with anorexia nervosa or anorexia-like symptoms are an effective tool to generate new hypotheses for the aetiology of anorexia nervosa (Vandenbroucke, 2001), it is important to bear in mind that evidence from case reports is anecdotal rather than systematic. It does not allow for the calculation of the risk of anorexia nervosa among religious women (Dekkers et al., 2012). This equally applies to both modern and medieval reports. Nevertheless, if causal effects between medieval religiosity and self-starvation existed, their direction is hard to determine. The aetiology of “holy anorexia” may as well have been self-starvation that was expressed religiously than religiosity that was expressed as self-starvation. Further, the effects of religiosity are difficult to dissect from the effects of other factors. For example, the “holy anorexia” of Saint Catherine of Siena onset simultaneously with her attempt to stay away from a marriage that was arranged against her will (Bell, 1987).

The religious, cultural, and social context of my study should be borne in mind. Finland is ethnically relatively homogenous (Alesina et al., 2003) and mostly Protestant, but has secularized quickly (Palmu et al., 2012). Over 70% of the Finnish population belong to the Lutheran church (Kirkkokalliitus, 2016), but only 6% of the population attend a church service at least monthly (Palmu et al., 2012). Furthermore, only some 0.3% of Finns are Catholic (Catholic Information Centre, 2017), and fasting is unusual in Finland. Therefore, my results mostly reflect moderate Protestant Christian religiosity in a present-day, highly secular, Western context. They may or may not generalize to other contexts.
6.3 Parents’ and their children’s problem drinking

In study II, I examined how potential confounders that I identified from the literature affect the association of parents’ problem drinking with their adult children’s problem drinking. I discovered that the association of parents’ problem drinking (measured with Mm-MAST) with their adult children’s problem drinking (measured with Mm-MAST-11) could not be explained by confounding due to area of residence, family structure, and fathers’ and mothers’ education, religiosity and one relevant dimension of personality. My results are compatible with earlier research as both alcohol drinking (Rossow et al., 2016b, 2016a) and alcohol use disorder (Cotton, 1979; Kendler et al., 2015a, 2015b) are known to cluster into families.

The authors of a recent systematic review judged that four earlier studies had some potential for causal inference (Rossow et al., 2016b); the rest had less potential. I contrasted my results to the results of those four studies: The more heavily drinking parent was assessed in another Finnish twin cohort with 4731 adolescent twins and their parents (Latendresse et al., 2008). His or her alcohol use, Mm-MAST-11 and intoxication frequency were associated with alcohol use and intoxication frequency of his or her 14- and 17.5-year-old children. Adjusted standardized betas ranged from 0.052 to 0.178. A study from the United States of America examined a score that combined frequency and volume of alcohol drinking with harmful alcohol use in 103 men and their fathers and mothers (Pears et al., 2007). Mean of fathers’ and mothers’ score correlated with the score of their 16–18-year-old sons (correlation coefficient = 0.27), and had a standardized beta of 0.22 in an adjusted path model.

Alati et al. (2014) studied in Australia how fathers’ and mothers’ drinking on a semi-ordinal Likert scale affects their 13.5–17.5-year-old adolescents’ drinking trajectories. Both fathers’ and mothers’ drinking were associated with an increased risk for their children to be on a high-drinking trajectory. A unit increase in fathers’ drinking had an adjusted odds ratio of 1.40 and a unit increase in mothers’ drinking had a considerable adjusted odds ratio of 2.77. Another study by Mares et al. (2011) on 13- and 15-year-old teenagers from the Netherlands confirmed these findings on fathers’ drinking but were in conflict in regard of mothers’ drinking. The associations of fathers’ alcohol use and alcohol related problems with excessive drinking and alcohol related problems in their children were predominantly positive although standardized betas from individual associations ranged from −0.15 to 0.17. The associations of mothers’ alcohol use and alcohol related problems with excessive drinking and alcohol related problems in their children were close to zero or slightly negative. Standardized betas ranged from −0.10 to 0.02. These inconsistencies between studies may be attributable to random variation. There were 751 parent-offspring pairs in the study of Alati et al. (2014) and 428 families with 856 parent-offspring pairs in the study of Mares et al. (2011).

I aimed to control for local environment, cultural and religious factors and parental comorbidity and temperament. They were not assessed in earlier
studies but were proposed in the systematic review of Rossow et al. (2016b) to be able to induce spurious associations to appear between alcohol use of parents’ and their children. I adjusted my analyses for area of residence, family structure and fathers’ and mothers’ education and religiosity. I did not have direct measurements on parental comorbidity or temperament, but I assessed a risk-relevant dimension of fathers’ and mothers’ personality with a social deviance scale (Pd scale). The scale is associated with alcohol drinking and problems (Mustanski et al., 2003; Viken et al., 2007). It also captures part of the genetic risk for externalizing disorders and drinking problems in the family, because externalizing disorders share part of their genetic background (Kendler et al., 2003). Given these facts, I argue that the social deviance scale reflects parental comorbidity and temperament relevant to alcohol use. Despite these adjustments, however, the association between parents’ and their children’s problem drinking may have been caused by residual confounding, which may arise from compromised validity or accuracy of the covariate measurements in this study (Fewell et al., 2007). Unmeasured or unknown confounders may also exist.

It is easy to think of several causal mechanisms that might link parents’ problem drinking with their children’s problem drinking. Parents’ drinking problems may lead to social learning, increase stress in the family and affect parenting practices (Leonard & Eiden, 2007; Rossow et al., 2016b). This may lead to gene–environment interactions by activating children’s genetic predisposition to alcohol use (Jacob et al., 2003; Rossow et al., 2016b). As the results of this study suggest, children’s drinking in adolescence may mediate these effects at least partly: the mechanisms described above may increase children’s drinking in adolescence, which may affect their later problem drinking in adulthood.

Genetic predisposition to problem drinking inherited from parents to their children may also explain the association of parents’ problem drinking with their children’s problem drinking (Hopfer et al., 2003; Verhulst et al., 2015). I used grandparents’ alcohol drinking as a proxy for the genetic predisposition to alcohol use in the family, which is transmitted from the grandparents to their children (parents of this study) and grandchildren (sons and daughters of this study). Additional adjustment for grandparents’ drinking as a proxy for this genetic predisposition did not affect the results. My measure of grandparents’ alcohol drinking, however, was far from perfect. It did not assess total alcohol consumption nor problem drinking, but only regularity of drinking. It was also reported by the next generation (the fathers and mothers of my study), and thus susceptible to reporting errors. On top of that, even a perfect measure of grandparents’ alcohol drinking would capture only part of their genetic predisposition to alcohol use (Dick et al., 2011b). The validity of grandparents’ alcohol drinking as a proxy for genetic predisposition to drinking may also be limited because the overall alcohol consumption in Finland was low before the 1960s (National Institute for Health and Welfare, 2017).
Future studies should differentiate between genetic and environmental effects in the association of parents’ problem drinking with problem drinking of their children. In regard of alcohol use disorder, a good attempt has already been made in an adoption setting. Both alcohol use disorder among the biological parents and alcohol use disorder among the adoptive parents increased the risk of alcohol use disorder among the adoptee children (Kendler et al., 2015a).

6.4 Alcohol drinking and all-cause mortality

In study III, I assessed the potential confounding effects of childhood family environment and genetic factors in the associations of different dimensions of alcohol drinking with all-cause mortality. There were three main results. First, I replicated the findings of extensive earlier research that high alcohol consumption and heavy drinking occasions are associated with increased all-cause mortality (Kauhanen et al., 1997; Laatikainen et al., 2003; Di Castelnuovo et al., 2006; Zaridze et al., 2014; Holahan et al., 2014; Smyth et al., 2015; Goulden, 2016; Wood et al., 2018). Second, these associations remained similar when I controlled for shared family background in discordant-twin analyses among all twin pairs. Thus, confounding by family background does not seem to explain these associations. Third, the associations of total alcohol consumption and alcohol-induced blackouts with all-cause mortality remained similar even after controlling for both shared family background and genetic factors in discordant-twin analyses among monozygotic twin pairs. Thus, confounding by genetic factors does not seem to explain these associations.

Heavy drinking occasions had a hazard ratio of 1.40 for all-cause mortality in discordant-twin analyses among monozygotic twin pairs, but this association was not statistically significant. This may be either due to some confounding by genetic factors or simply due to small statistical power. There were only 30 monozygotic twin pairs, who were discordant for heavy drinking occasions, and of whom at least one had died during the follow-up.

Three earlier studies have assessed all-cause mortality among drinking-discordant twin pairs. The oldest of them focused on abstainers and light to moderate drinkers (Carmelli et al., 1995). A new study on a subsample of the same cohort could not find an association between alcohol consumption and all-cause mortality (Dai et al., 2015). Among monozygotic twin pairs, hazard ratio for all-cause mortality per 10 g/day increment in alcohol consumption was 0.98 (95% CI 0.95–1.04). Finally, an earlier study from the Older Finnish Twin Cohort found a clear association between heavy drinking occasions and all-cause mortality among all twin pairs (odds ratio 3.00, 95% CI 1.70–5.82) (Kujala et al., 2002). Among monozygotic twins, however, only 13 informative pairs were available yielding an inconclusive result (odds ratio 0.86, 95% CI 0.29–2.55).
To my knowledge, this is the first study to demonstrate that high total alcohol consumption ($\geq 259$ g/month corresponding to more than about 5 drinks per week) and alcohol-induced blackouts (at least twice a year) are associated with increased all-cause mortality even when childhood family environment and genetic factors are adjusted for in discordant-twin analyses among monozygotic twin pairs. In regard of heavy drinking occasions, further studies with more drinking-discordant monozygotic twin pairs are needed to reach firm conclusions.

The current Finnish nutrition guidelines recommend a maximum alcohol consumption of 20 grams per day (140 g/week or 600 g/month) for men and 10 grams per day (70 g/week or 300 g/month) for women, and similar or higher guidelines are used in several Western countries (Fogelhom et al., 2014; Kalinowski & Humphreys, 2016). According to this study, these recommended drinking limits should be lower, especially for men. A recent multi-cohort study also supports this conclusion. It found that alcohol consumption in excess of 100 grams per week ($\approx 400$ grams per month) was associated with increased all-cause mortality, and called for lower limits of recommended alcohol consumption (Wood et al., 2018).

6.5 Strengths

This study has important strengths that help tackle bias. First, I used two population-based cohorts with high response rates (studies I–III). Second, in study I, I had actual diagnoses of lifetime anorexia nervosa from structured clinical interviews, and to my knowledge, my study was the first one to assess the direct association between religiosity and anorexia nervosa in a population-based sample. I also used multiple imputation to decrease the risk of selection bias due to missing information (Sterne et al., 2009; Hernán & Robins, 2018). Third, I studied outcomes associated with alcohol drinking in a prospective follow-up setting (studies II and III), which reduces the risk of reverse causation (Ioannidis, 2016). Fourth, Mm-MAST and heavy drinking occasions are validated measures for alcohol drinking (Kristenson & Trell, 1982; Seppa et al., 1990, 1992; Nyström et al., 1993; Gmel et al., 2011; Rehm et al., 2017) (studies II and III), and the Religious fundamentalism content scale of the Minnesota Multiphasic Personality Inventory is a validated measure for religiosity (Wiggins, 1966) (study I). Fifth, in study II, I assessed potential confounders that were proposed but overlooked in earlier studies (Rossow et al., 2016). Sixth, in study III, I had a thirty-year follow-up that enabled me to assess the long-term consequences of alcohol drinking, and I used discordant-twin design which permits adjustments for childhood family environment and genetic factors that are otherwise hard to control for (McGue et al., 2010).
6.6 Limitations

Listing the strengths of this study is not to say that it is without limitations. In study I, I used multiple imputation to impute missing information. Nevertheless, bias may have occurred if there is information missing not at random (Sterne et al., 2009). Additionally, analysing the data in two separate datasets compromised statistical power to some extent. In study III, selection bias due to missing data is unlikely, because only 5.3% of the individuals in the study had missing items. But in study II, despite high overall response rates, missing information on individual variables may have caused selection bias (Hernán & Robins, 2018). Further, in study II, I imputed missing items in problem drinking, religiosity and personality scales with a single imputation method that, despite often considered acceptable, may nonetheless lead to bias (Graham, 2009; Mazza et al., 2015).

Reporting errors may have affected my results as my measures were mostly self-reported. In study II, I measured parents’ problem drinking only once. In study III, I had one measurement of alcohol-induced blackouts and two measurements of total alcohol consumption and heavy drinking occasions, but they all were conducted before the start of the follow-up. Most people seem to report their alcohol use with a reasonable reliability (Chermack et al., 1998; Babor et al., 2000; Del Boca & Darkes, 2003), but heavy drinkers may underreport their drinking (Northcote & Livingston, 2011). In addition, many people change their alcohol drinking habits over time (Seppä et al., 1999; Maggs & Schulenberg, 2004; Sher et al., 2011; Britton et al., 2015; Knott et al., 2018). This is especially of concern in study III, in which substantial changes in alcohol drinking may have occurred during the thirty-year follow-up. Therefore, misclassification errors may have introduced information bias to my results (Rothman et al., 2008b).

The questions on offspring Mm-MAST-11 assessed lifetime problem drinking, and could not differentiate between current and past drinking problems (study II). In study I, lifetime anorexia nervosa was diagnosed in retrospect. This may have led to recall bias. My measures of religiosity did not allow for direct comparison between study waves. In adolescence, and among fathers and mothers, I had a validated 12-item religiosity scale, but it was insensitive to religiosity that does not adhere to Christian tenets (Wiggins, 1966; Winter et al., 1999). In contrast, in early adulthood, my measure of religiosity had only one item. It measured general religiosity rather than adherence to any particular religion. It also had a constrained distribution with very few highly religious women. I also could not analyse motives of religiosity which may be important (Smith et al., 2004; Castellini et al., 2014; Akrawi et al., 2015). I did not have a genuine prospective setting to study the risk of incident anorexia nervosa as the information on the age of the onset of anorexia nervosa was collected in retrospect. Finally, there were in overall relatively few women with lifetime anorexia nervosa. Therefore, I cannot
interpret my results as evidence for no association between religiosity and lifetime anorexia nervosa.

I cannot exclude the possibility of reverse causation in the association between parents’ and their children’s problem drinking (study II). Sons’ and daughters’ problems in adolescence (drinking problems or otherwise) may have changed their parents’ alcohol use by increasing stress in the family (Leonard & Eiden, 2007). Reverse causation is also possible in study I. Daughters’ anorexia nervosa may have affected both their personal religiosity and their parents’ religiosity.

Finally, no honest epidemiologist can report his or her findings without acknowledging the possibility of residual confounding (Davey Smith & Phillips, 1992). Inaccurately measured, unmeasured and unknown confounders can affect my results. First, my measures of socioeconomic status were limited by the lack of information on income (Braveman et al., 2005) (studies I–III). Second, although my measure of social deviance (Pd scale) arguably measures a dimension of personality that is relevant to problem drinking, I could not adjust for other dimensions of personality (McCrae, 2011) (study II). Third, in study II, I assessed the potential mediating effect of children’s alcohol drinking in adolescence. Measurement error in a mediating variable may lead to biased estimates of the mediation effect. In addition, I assessed only one dimension of children’s alcohol drinking in adolescence. The observed mediation effects would have been larger if my measures had captured other dimensions of alcohol drinking as well. Fourth, despite its strengths, the discordant-twin design cannot control for environmental confounders that are not shared by the co-twins (Frisell et al., 2012) (study III). To overcome this limitation, future studies should synthesize evidence from different study designs with different potential sources of bias (Lawlor et al., 2016). For example, Mendelian randomization studies on alcohol drinking and all-cause mortality could be conducted, and in respect of moderate drinking, even randomized controlled trials may be feasible (Holmes et al., 2014; Rabin, 2018).

6.7 Conclusions and future directions

I have identified some potential sources of bias from the literature of psychiatric epidemiology, and examined whether they affect the observed associations. Specifically, I have studied associations between potential risk factors and adverse outcomes in the epidemiology of anorexia nervosa and alcohol drinking.

In summary, I can draw three main conclusions from these studies. First, despite intriguing case reports, religiosity does not seem to be a major risk factor for anorexia nervosa in modern Finland. Other factors are probably more important in the aetiology of this severe disorder. Additionally, addressing selection bias by multiple imputation did not change the results.
Second, my results indicate that the association of parents’ problem drinking with problem drinking of their adult children cannot be explained by the confounding effects of area of residence, family structure, and fathers’ and mothers’ education, religiosity and personality. Parents’ problem drinking appears to be a robust risk factor for their children’s problem drinking. Future studies should investigate to which extent this association reflects genetic predisposition to problem drinking that is inherited from generation to generation. They should also investigate to which extent, if any, this association reflects causal effects of parents’ problem drinking beyond the effects of genes predisposing to problem drinking.

Third, the associations of total alcohol consumption and alcohol-induced blackouts with all-cause mortality were robust to adjustments for the assessed potential confounders: childhood family environment, genetic factors, age, sex, smoking, physical activity, obesity, education, social class and marital status. My findings support the causal interpretation of these associations. Concerning heavy drinking occasions, future studies with adequate statistical power are needed before firm conclusions can be drawn.

If we wish the scientific research to be in any way a meaningful exercise, it should produce results that are rather right than wrong. Therefore, the possibility of bias should be thoroughly examined in every epidemiological study. I have examined three potential sources of bias, but a lot remains to be done. Bias is ubiquitous.
I planned to write this dissertation on causal inference. Gradually, however, I realized that my plans were overly self-confident. While wrestling with epidemiological problems, I started to see some of the many uncertainties in scientific knowledge in general – and in my studies in particular. I started to learn.

In all I have learned I owe a lot to others. The Doctoral Programme in Population Health, University of Helsinki, paid my salary for two years. In addition, the Medical Faculty employed me as a research assistant during my undergraduate studies, and the Finnish Foundation for Alcohol Studies (Alkoholitutkimussäätiö in Finnish) supported my research with a one-year grant. Still, despite their zeal to back me, the funders did not intervene my work in any other way.

I am very thankful to my supervisors Professor Jaakko Kaprio and Associate professor Anna Keski-Rahkonen. Jaakko taught me the basics of research, and guided and supported me during these years. In addition, he arranged the necessary funds to attend useful courses in epidemiology and to pay my salary during a gap in funding. Anna has been enormously influential during these first steps of my scientific career. She taught me conceptual and critical thinking (I wonder how I may have been thinking beforehand!). She even taught me to read and write. She built onto our department a warm community of young epidemiologists amongst whom it has been a pleasure to work. She also helped me a lot in funding acquisition, and mentored and supported me in the moments of despair. There were quite a few of them. I would also like to thank the members of my thesis committee, Professor Ossi Rahkonen and Karri Silventoinen, for supporting my journey. Ossi deserves a special mention: he has been like a godfather looking after my progression.

Besides those in official supervising positions, many colleagues have provided me with instrumental help. Richard J. Rose worked hard to supervise and help me with the paper on drinking and mortality, and was involved in the other original studies of this thesis as well. Gulnara Harrasova and Linda Mustelin were important contributors to the paper on religiosity and anorexia nervosa, and Joni V. Lindbohm has been a wonderful room-mate ever ready to help me. Yasmina Silén helped me to draft a successful press release on drinking and mortality. Finally, the preliminary examiners, Research professors Pia Mäkelä and Professor Henrik Larsson, helped ensure that all my work was condensed to a dissertation with at least some scientific coherence.

I have mentioned some of the many wonderful colleagues I have been blessed to work with. Still, I think I have been even more blessed outside my office. First of all, my parents gave their all to raise me and my siblings. They also taught me to embrace education, which I have now been doing for twenty-
four years. Other important building blocks of my life are my dear sister, brother, friends, parents-in-law and brothers-in-law. Mikael shared with me the moments of doctoral student’s frustration. Reverend Juuso Mäkinen helped me to cite the Scriptures correctly.

Most importantly, I am so thankful of my family. My wife Elina and my two wonderful sons, you bring immense joy to my every single day. Elina, your love has been the same in both good and bad days. When we got married, I had already started this PhD project. We promised that our marriage would last until death. I am glad this PhD project did not.

Helsinki, 22 August 2018

Pyry Sipilä
References


Catholic Information Centre (2017). Rekisteröityjen katolilaisten määrä kasvaa tasaisesti [Number of registered Catholics is in steady rise]. Fides (3), 4.


First MB, Spitzer RL, Gibbon M & Williams JBW (2002). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient*


The original publications are printed in the paperback version of this dissertation. They are also available on the publishers’ websites on the World Wide Web.
PYRY N. SIPILÄ
Dissecting Epidemiological Associations in Alcohol Drinking and Anorexia Nervosa