Celiac Disease and Infections
To my family
Celiac Disease and Infections
Abstract

Anna Röckert Tjernberg (2019): Celiac Disease and Infections.
Örebro Studies in Medicine 188.

Background: Celiac disease (CD) is a chronic immune-mediated enteropathy affecting about 1% of the population worldwide. CD is triggered by ingestion of gluten in genetically predisposed individuals but additional factors (e.g. infections) are required for the disease to develop. CD also seems to be associated with infectious complications.

Aim: The main objective of this thesis was to increase the knowledge about the associations between CD and infections.

Methods: Epidemiological and laboratory approaches. Studies I-III used a data set consisting of small intestinal biopsy reports. The biopsies were taken in 1969-2008 and collected in 2006-2008. A total of 29,096 individuals with CD, 13,306 with inflammation and 3,719 with potential CD were identified. Each individual was matched with up to 5 controls from the general population (n= 228,632). Through linkage of the data to the Patient Register study I examined the risk of hospital visits due to respiratory syncytial virus (RSV) in children <2 years prior to onset of CD. Study II used the Patient Register and Cause of Death Register to assess whether CD affects the outcome in sepsis. Study III linked the data to microbiological data bases and the Public Health Agency to estimate risk of invasive pneumococcal disease (IPD) in CD. In study IV children with CD and controls were recruited from Kalmar County Hospital. Complement activation (C3a and sC5b-9) in plasma were analysed after incubation with pneumococci.

Results: Study I found that children with CD were more likely than controls to have attended hospital due to RSV infection prior to diagnosis (odds ratio 1.46; 95% confidence interval (CI)=1.02-2.07). CD did not seem to influence survival in sepsis (adjusted hazard ratio (HR) 1.10 95%CI=0.72-1.69) (study II). Study III indicated a 46% risk increase for individuals with CD to acquire IPD (HR 1.46; 95%CI=1.05-2.03) but study IV did not reveal any differences in complement response in regard to CD status (p=0.497and p=0.724), explaining this excess risk.

Conclusion: This thesis supports associations between CD and infections preceding and complicating diagnosis. However, CD does not seem to influence the outcome in a severe infection like sepsis and altered complement function is unlikely to be responsible for the excess IPD risk in CD.

Keywords: celiac disease, small intestinal, infection, respiratory syncytial virus, sepsis, streptococcus pneumoniae, complement, cohort, register

Anna Röckert Tjernberg, School of Medical Sciences
Örebro University, SE-701 82 Örebro, Sweden, anna.rockert@regionkalmar.se
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGA</td>
<td>Anti-Gliadin Antibodies</td>
</tr>
<tr>
<td>CD</td>
<td>Celiac Disease</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>DGP</td>
<td>Deamidated Gliadin Peptide (Antibodies)</td>
</tr>
<tr>
<td>EMA</td>
<td>Endomysial Antibodies</td>
</tr>
<tr>
<td>ESPGHAN</td>
<td>European Society for Pediatric Gastroenterology, Hepatology and Nutrition</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme-linked Immunosorbent Assay</td>
</tr>
<tr>
<td>GFD</td>
<td>Gluten Free Diet</td>
</tr>
<tr>
<td>HLA</td>
<td>Human Leukocyte Antigen</td>
</tr>
<tr>
<td>HR</td>
<td>Hazard Ratio</td>
</tr>
<tr>
<td>ICD</td>
<td>International Statistical Classification of Diseases and Related Health Problems</td>
</tr>
<tr>
<td>IPD</td>
<td>Invasive Pneumococcal Disease</td>
</tr>
<tr>
<td>OR</td>
<td>Odds Ratio</td>
</tr>
<tr>
<td>PIN</td>
<td>Personal Identity Number</td>
</tr>
<tr>
<td>RSV</td>
<td>Respiratory Syncytial Virus</td>
</tr>
<tr>
<td>T1DM</td>
<td>Type 1 Diabetes Mellitus</td>
</tr>
<tr>
<td>TPR</td>
<td>Total Population Register</td>
</tr>
<tr>
<td>tTG</td>
<td>Tissue Transglutaminase (including antibodies)</td>
</tr>
<tr>
<td>VA</td>
<td>Villous Atrophy</td>
</tr>
</tbody>
</table>
List of papers

I.  **Röckert Tjernberg A.** Ludvigsson JF.
Children with celiac disease are more likely to have attended hospital for prior respiratory syncytial virus infection. *Dig Dis Sci.* 2014 Jul;59(7):102-8.

II. **Röckert Tjernberg A.** Bonnedahl J, Ludvigsson JF.


Celiac disease and complement activation in response to *Streptococcus pneumoniae*. Submitted.

---

Paper I is reproduced with permission from *Dig Dis Sci*, Copyright© Springer Science + Business Media New York 2014
Paper II is reproduced with permission from *PLoS One*
Paper III is reproduced with permission from *Epidemiol Infect*, Copyright© Cambridge University Press 2017

*Contributed equally*
# Table of Contents

## INTRODUCTION

INTRODUCTION ........................................................................................................... 13

## BACKGROUND

BACKGROUND ........................................................................................................... 14

### Celiac disease

**Celiac disease** ........................................................................................................ 14

- History ................................................................................................................... 14
- Descriptive epidemiology ..................................................................................... 15
- Pathogenesis .......................................................................................................... 16
- **Multifactorial etiology** ..................................................................................... 16
- **Gluten** ............................................................................................................... 16
- **Genetics** ............................................................................................................ 16

### Environmental and lifestyle risk factors

- Infections as risk factors ....................................................................................... 17
- Microbiota .............................................................................................................. 19

### Immunology

**Immunology** ........................................................................................................ 20

### Clinical presentation

**Clinical presentation** ........................................................................................... 22

### Children

**Children** ............................................................................................................... 22

### Adults

**Adults** .................................................................................................................. 23

### Diagnostics

**Diagnostics** .......................................................................................................... 24

### Serologic testing

**Serologic testing** ................................................................................................... 25

### Genetic testing

**Genetic testing** .................................................................................................... 26

### Small intestinal biopsy (histopathology)

**Small intestinal biopsy (histopathology)** ........................................................... 26

### Risk groups (subjects for regular surveillance)

**Risk groups (subjects for regular surveillance)** .................................................. 28

### Treatment

**Treatment** ............................................................................................................. 29

### Gluten-free diet

**Gluten-free diet** .................................................................................................... 29

### Alternative treatments

**Alternative treatments** ....................................................................................... 30

### Associated conditions and complications

**Associated conditions and complications** ......................................................... 31

### Refractory celiac disease

**Refractory celiac disease** .................................................................................... 33

### Dermatitis herpetiformis and gluten ataxia

**Dermatitis herpetiformis and gluten ataxia** ....................................................... 33

### Infections following celiac diagnosis

**Infections following celiac diagnosis** .................................................................. 33

### Hyposplenism

**Hyposplenism** ....................................................................................................... 34

### Increased intestinal permeability

**Increased intestinal permeability** ....................................................................... 34

### Mortality

**Mortality** ............................................................................................................... 35

### Infections

**Infections** ............................................................................................................. 36

- Respiratory syncytial virus infections .................................................................. 36
- Sepsis ..................................................................................................................... 36
- Pneumococcal infections ...................................................................................... 37
- The complement system ....................................................................................... 37
The complement system in celiac disease ................................................................. 39
Rationale for this thesis .................................................................................................... 40
OBJECTIVES ................................................................................................................ 41
Specific objectives .......................................................................................................... 41
METHODS ......................................................................................................................... 43
Setting and personal identity number ........................................................................ 43
Registers/data sources .................................................................................................. 43
Statistics Sweden ......................................................................................................... 43
The Swedish Patient Register ....................................................................................... 44
The Swedish Medical Birth Register ............................................................................ 44
The Swedish Cause of Death Register ......................................................................... 44
The Swedish Public Health Agency ............................................................................ 45
Study populations and designs ..................................................................................... 45
Study I Respiratory syncytial virus and later celiac disease ....................................... 47
Study II Celiac disease and survival in sepsis ............................................................. 47
Study III Celiac disease and invasive pneumococcal disease ..................................... 48
Study IV Celiac disease and complement response to pneumococci ......................... 49
Statistics ......................................................................................................................... 51
Data characteristics ....................................................................................................... 51
Logistic regression ......................................................................................................... 52
Cox proportional hazards regression model ................................................................... 52
Independent sample t-test, Mann-Whitney U-test and Wilcoxon signed rank test .......... 53
Confidence intervals and p-values ................................................................................. 53
Ethical considerations .................................................................................................... 54
RESULTS ......................................................................................................................... 55
Study I Respiratory syncytial virus and later celiac disease ....................................... 55
Respiratory syncytial virus ............................................................................................ 55
Unspecified viral bronchiolitis ....................................................................................... 55
Study II Celiac disease and survival in sepsis ............................................................. 55
All-cause mortality ......................................................................................................... 56
Mortality from sepsis ....................................................................................................... 56
Study III Celiac disease and invasive pneumococcal disease ..................................... 56
Study IV Celiac disease and complement response to pneumococci ......................... 57
Complement activity (main outcome) .......................................................................... 57
Mannan-binding lectin .................................................................................................. 59
Pneumococcal serology ................................................................................................. 59
DISCUSSION

Findings and implications

Study I Respiratory syncytial virus and later celiac disease
Study II Celiac disease and survival in sepsis
Study III Celiac disease and invasive pneumococcal disease
Study IV Celiac disease and complement response to pneumococci

Methodological considerations

Strengths and limitations in general
Study designs

Validity

Internal validity
Selection bias
Misclassification
Celiac disease
Respiratory syncytial virus
Sepsis
Invasive pneumococcal disease
Causes of death
Recall bias
Surveillance bias
Confounding bias
Chance, precision and power
External validity

Causality

CONCLUDING REMARKS AND FUTURE RESEARCH

SAMMANFATTNING PÅ SVENSKA

ACKNOWLEDGEMENTS

REFERENCES

PAPERS I-IV
Introduction

This thesis is about celiac disease, a common gastrointestinal disorder, and its associations with infections. Both celiac disease and infections are diagnoses I come across almost daily in my work as a pediatrician, making them excellent topics to explore. The journey this thesis has taken me on has been very joyful and stimulating. I have learnt a lot and come to realise that there is so much left to learn. Obviously this has only been the beginning of a never-ending journey.

Figure 1. Vincent van Gogh “Wheat Field with Cypresses” 1889. (Wikimedia Commons)
Background

Celiac disease

A broad consensus regarding the definition and nomenclature related to celiac disease (CD) was for a long time lacking. However, in 2012, researchers from seven countries made an effort to reach agreement in this issue. The panel work resulted in the so called “Oslo definitions for celiac disease and related terms”. The paper defines CD as “a chronic small intestinal immune-mediated enteropathy precipitated by exposure to dietary gluten in genetically predisposed individuals” [1]. This definition still describes the condition well, is generally accepted and since the publishing of the work commonly used.

CD is a rather common chronic autoimmune disorder affecting people of all ages [2, 3]. Like for many other autoimmune diseases the etiology is considered multifactorial [4]. CD is associated with other autoimmune diseases [5] like type 1 diabetes mellitus (T1DM) [6, 7] and autoimmune thyroiditis [8] but also with malignancies [9, 10] and infections [11, 12]. Patients often present themselves with abdominal complaints, growth retardation (children) or weight loss but symptoms can be far more subtle [13, 14]. So far, treatment consists of a lifelong gluten-free diet (GFD) [15].

History

As early as in the second century AD, Aretaeus the Cappadocian, a Greek physician gave the first known description of CD (koiliakos meaning abdominal in Greek) [16]. The disease again received attention in 1887 through Dr Samuel Jones Gee who was the first to recognize that this “chronic indigestion” with steatorrhea could affect people of all ages [16, 17]. Several physicians then showed an interest in this “most obscure” disease and a few of them believed, like Dr Gee, that it could be cured by diet, for example with mussels [16]. In the first half of the twentieth century Dr Sydney Haas promoted a banana diet and was hesitant when the Dutch pediatrician Willem Karel Dicke in the 1950s suggested an association between wheat and CD. Dicke believed wheat to be the causative agent already in the 1930s. He was, however, further convinced when he observed that children suffering from CD improved during the bread shortage of World War II and seemed to relapse once the cereal stock was restored [16, 17]. He later confirmed this observation experimentally [18]. Still, about 75 years later, the withdrawal of wheat and related cereals from the diet is the only working treatment for CD [15].
**Descriptive epidemiology**

CD occurs in about 1% of the population worldwide [3, 4]. Nevertheless, there are variations in prevalence. Geographical differences as well as differences according to age, sex and ethnicity are seen [2-4, 14, 19]. However, in addition to sometimes small sample sizes, some studies base their prevalence numbers on serological markers while others base them on histopathological findings, hampering comparisons.

While the CD prevalence is still unknown in several parts of the world, one of the highest estimates of prevalence, 5.6%, has been identified in Saharawi children living in a desert area in Algeria [20]. In contrast, in a study from 2012, no positive tests for endomysial antibodies (EMA) were seen in 860 Sub-Saharan African-derived Brazilians [21] whereas Brazilians claiming to have European ancestry had a considerably higher prevalence of CD [22]. In most studies the CD prevalence is lower in Sub-Saharan Africa and the Orient, probably partly due to a lower wheat consumption and human leukocyte antigen (HLA) DQ2 frequency [19, 23]. Ethnical differences are also seen within the United States, with the highest prevalence observed amongst non-Hispanic whites ([24] and in people with ancestry from northern India [4]. CD affects people of all ages and given the rather low CD related mortality the prevalence is surprisingly similar in different age groups although there are variations [2, 3]. Similarly to many other autoimmune diseases [25] a female predominance is seen [3, 14, 26]. Considering the rather wide variety in CD prevalence seen for example between different countries in Europe (Finland 2.4% versus Germany 0.3%) [27] as well as within some countries themselves [2, 4] it is obvious that the disparities in disease frequency cannot fully be explained by the today known genetic and environmental factors nor by different awareness amongst the population and medical personnel. This observation of course encourages further research.

All around the world the incidence of CD appears to be increasing [2, 19, 26, 28, 29]. This is partly due to a greater knowledge and more available and improved diagnostic tools but also seems to reflect a true increase [19, 30]. Moreover, despite the improved diagnostic possibilities the current opinion is that a non-negligible part of the celiac population still is unrecognized and undiagnosed [4, 31, 32]. This is often referred to as the “celiac iceberg” with the tip of the iceberg representing the symptomatic CD with classic or non-classic symptoms [33].

Worth mentioning is that we, in Sweden, experienced a “celiac epidemic” during the years 1984 to 1996. The CD incidence rate in children below 2 years of age
suddenly increased 4-fold. The phenomenon was partly explained by changes in infant feeding patterns and the increase in incidence rate declined rather abruptly after about decade [34]. However, a follow-up study on 12-year-old children born during the epidemic still showed a rather impressive CD prevalence of 3% [31].

**Pathogenesis**

**Multifactorial etiology**

The etiology of CD is considered multifactorial [4]. A genetic predisposition is a prerequisite [35] but far from the only factor contributing to the disease. Unlike many other autoimmune diseases CD has a known trigger: gluten, a protein complex contained in wheat, rye and barely [1]. However, several other factors also have to coincide for the disease to develop.

**Gluten**

Gluten, the Latin word for glue, is the sticky (glutinous) protein part that remains after washing wheat flour with water. Due to the elastic properties it is very favourable in baking [4, 35]. The gluten complex can be further divided in to the alcohol-soluble gliadins and glutenins [1]. Related storage proteins (prolamins) are found in rye (secalins) and barley (hordeins) [1, 36]. Hence, strictly speaking, the term gluten refers only to the wheat proteins but is now commonly used to describe all the dietary proteins involved in CD pathogenesis [4, 35].

**Genetics**

The genetic predisposition which is a prerequisite for developing CD is mainly attributed to genes encoding for human leukocyte antigens (HLA) [35]. The first observations showing that HLA molecules affect the risk of CD came already in the 1970s [37]. It is now known that about 90% of individuals with CD carry genes encoding for the HLA molecule DQ2.5. The remaining patients instead express HLA molecules DQ2.2 or DQ8 associated with a somewhat lower risk of CD. Only very few patients express, the also CD associated, HLA DQ7.5 [35].

Considering the importance of these certain HLA molecules in disease pathogenesis it is not surprising that the prevalence of CD is increased in both first- and second-degree relatives (pooled prevalence 7.5% and 2.3% respectively). The highest risk seems to be found in siblings, particularly sisters [38]. However, expression of HLA DQ2 or DQ8 is common in the general population. The prevalence is about 40% among people in Europe, North and South America and Southeast Asia and most individuals carrying these HLA molecules will never develop CD [4, 39].
Bearing that in mind and also considering twin studies showing much higher CD concordance rates in monozygotic twins (>80%) compared to dizygotic twins (=15-20%) [40, 41] it is clear that other genes than those encoding for HLA molecules also are of importance in CD [35, 42]. During the last decades genome-wide association studies have contributed with a lot of information. In 2016 as many as 42 non-HLA CD related loci had been identified. All contributing to the CD pathogenesis but compared to the effect size of the HLA molecules their individual effect sizes are very small [35]. However, their combined effect size is not negligible [35, 42]. It is estimated that the HLA alleles account for about 40% of the genetic CD risk whereas the combination of 37 non-HLA loci contributes with about 14% [35].

Environmental and lifestyle risk factors
A genetic predisposition and exposure to dietary gluten are prerequisites for developing CD but still not sufficient to cause disease [4]. Several additional factors have been proposed to be involved in disease pathogenesis [43-45].

The previously mentioned “Swedish celiac epidemic” gave opportunities to explore environmental factors. Studies showed that the steep rise in incidence seemed to coincide with changes in infant feeding patterns. During this period the national recommendation was to introduce gluten, in rather large amounts, at 6 months of age (when many infants had already weaned) [34, 46]. However, research on breast-feeding and gluten introduction and risk of CD has been contradictory [47-49]. Several efforts have been made to compile the existing knowledge and in 2016, an ESPGHAN (European Society for Pediatric Gastroenterology, Hepatology and Nutrition) position paper on this issue was published [43]. The paper acknowledges the benefits of breast-feeding in general but states that there is not enough evidence to conclude that breast-feeding during gluten introduction protects against CD. Likewise, it states that gluten can be introduced in the child’s diet at any time between 4 and 12 months of age, the timing does not appear to affect the absolute risk of CD [43]. However, in children at high risk of CD (i.e. HLA DQ2.5 homozygous) an early gluten introduction seems to be associated with an earlier development of CD autoimmunity and CD but still the cumulative incidence is not affected [43, 47]. Some studies show that the amount of ingested gluten can affect the risk of CD, particularly in at-risk children [34, 50] and the position paper recommends that large quantities of gluten should be avoided the first weeks after gluten introduction but emphasizes that the optimal amounts remain to be established [43].
CD has also been linked to perinatal and socioeconomic factors, however, evidence can still be considered circumstantial and the clinical significance of these findings is difficult to determine. A few studies have shown a weak association between being born small for gestational age [51] or having a low birth weight [51, 52] as well as being the second child [51, 53] but other studies have failed to confirm these findings [54]. Decreasing maternal age has been associated with CD in some studies [51] whereas others only find differences between strata and no clear temporal trends [52, 53]. Research is not consistent when it comes to the effects of education and socioeconomic position [52-54]. However, in general the prevalence of CD is somewhat higher in areas considered to be less socioeconomically deprived [55-57]. The latter is, most probably, mainly due to a greater awareness of the disease among people with higher socioeconomic position, although other explanations like the “hygiene hypothesis” cannot be excluded [55].

Infections as risk factors

In CD, like in many other autoimmune conditions, infections have been suggested to be involved in the pathogenesis [44, 51, 58]. It appears as infections can induce expression of specific cell surface antigens on the enterocytes, enhancing epithelial destruction [59]. In addition, viruses might initiate loss of oral tolerance by inducing T-helper (TH1) immunity to gluten [60] and can probably also induce autoimmunity though molecular mimicry [61]. Furthermore, it is possible that infections increase intestinal permeability, facilitating the crossing of gliadin peptides into the lamina propria [59, 61].

During the last decades efforts have been made to determine which microbe or microbes that can be of importance. A number of studies have investigated the role of a gastrointestinal infection preceding CD diagnosis [44, 52, 62]. In particular rotavirus has received attention. In 2006, Stene et al showed that increasing levels of antibodies against rotavirus in children carrying HLA risk alleles seemed to be associated with the risk of developing CD [63]. The finding was not statistically significant but later studies showing that vaccination against rotavirus might reduce the risk of CD (particularly in at-risk children) support this observation [62, 64] as does the finding that a subset of transglutaminase antibodies recognize a rotavirus protein [61].

Also respiratory tract infections have been subject to interest. Studies have shown an increased onset of CD after both upper and lower respiratory tract infections including influenza [44, 58, 65]. One of the most pronounced effects was seen in an Italian study where at-risk children who experienced respiratory tract infections
in the second year of life had a more than twofold increase in the risk of developing CD [66].

A large Norwegian prospective cohort study also showed that the risk of CD increased with the number of reported infections before 18 months of age [58], confirming findings from a previous Swedish study [67]. There are also indications that an infection at time of gluten introduction might act synergistically and increase the risk of CD in small children (odds ratio (OR) 1.8; 95% confidence interval (CI)=0.9-3.6) [68], findings supported by observations that children born in summer (believed to wean and introduce gluten during winter when viral infections are more frequent) have a somewhat higher risk of CD [69, 70]. However, evidence is not strong (ORs 1.17-1.4).

Data regarding infections as risk factors for CD are not consistent. However, taken together, evidence supports hypotheses regarding the involvement of infections in CD pathogenesis but further research is needed.

Microbiota
As described above, infections seem to play a role in CD development. One possible mechanism for this could be through the intestinal microbiota. Studies have shown that individuals with CD have alterations in their microbiota [4, 71] both regarding bacterial composition and diversity [45, 72] but also metabolic activity [73]. Differences have been demonstrated in both the small and large intestine as well as in fecal samples, with for example increased proportions of potentially pro-inflammatory Gram-negative bacteria in active CD although bacterial findings are a bit inconsistent [45, 71]. Despite results being conflicting [74] the theories are indirectly supported by studies observing an increased risk of CD following early antibiotic exposure [52, 75] as well as in children born with elective caesarean section [53, 76]. In addition, Spanish researchers have made the interesting observation that HLA haplotypes seem to influence the early gut microbiota composition [77].
Immunology

The complex immune response in CD involves both the adaptive and the innate immune system [59].

The gluten complex is poorly and incompletely digested in the gastrointestinal system and the immunogenic 33-amino acid long α2-gliadin peptides are able to cross the upper small intestine’s epithelial barrier rather intact [36, 59, 78]. The peptides enter the lamina propria either through a transcellular or a paracellular route [78, 79]. In the lamina propria, the enzyme tissue transglutaminase (the major autoantigen in CD) acts by deamidating the gliadin peptides, thereby (through altered charges) increasing their affinity to the HLA DQ2 and DQ8 molecules expressed on the surface of antigen-presenting cells [35, 36, 59]. The HLA-gliadin complex in turn activates CD4+ T-cells which start releasing pro-inflammatory cytokines, particularly interferon-γ. The subsequent inflammatory cascade (including the release of metalloproteinases) results in crypt hyperplasia and villous atrophy (VA), characteristic features of CD, but also in the activation of B-cells that produce (disease-specific) antibodies [4, 35, 59]. Through the innate immune system the gliadin peptides also induce changes in the epithelium [59]. Damaged epithelial cells (enterocytes) overexpress interleukin-15 leading to recruitment and activation of cytotoxic intraepithelial lymphocytes (natural-killer (NK)-cells). Gluten, infections or similar stressors seem to induce expression of specific cell-surface antigens on the enterocytes. These antigens are recognized by the NK-cells and hereby the destruction of the epithelial cells is enhanced [4, 59].

While both the adaptive immune system (in the lamina propria) and the innate system (in the epithelium) are involved in CD pathogenesis it is still not fully clear how the two systems interact [59].
Figure 2. CD pathogenesis.
Clinical presentation

The wide variation in clinical presentation of CD can confuse the clinician. In addition, symptoms and signs of CD have shifted over the years and fewer patients now present with classical CD (Table 1) [14, 28]. This of course further hampers diagnosing.

Table 1. Definitions of CD related terms, proposed terminology.
Adapted from Ludvigsson et al [1].

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition/presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classical CD</td>
<td>Symptoms and signs of malabsorption (i.e. diarrhea, steatorrhea, weight loss or growth failure).</td>
</tr>
<tr>
<td>Non-classical CD</td>
<td>Symptoms associated with CD (e.g. abdominal pain, bloating or constipation) but no signs of malabsorption.</td>
</tr>
<tr>
<td>Symptomatic CD</td>
<td>Clinically evident symptoms attributable to gluten ingestion.</td>
</tr>
<tr>
<td>Asymptomatic CD</td>
<td>No symptoms commonly associated with CD, nor even when specifically asked for. No clinical response to GFD.</td>
</tr>
<tr>
<td>Subclinical CD</td>
<td>Below the threshold of clinical detection. No obvious clinical symptoms that indicate need for CD testing. (The term has also been used for extraintestinal manifestations that can be seen in CD i.e. clinical or laboratory signs of anaemia, elevated liver enzymes, osteoporosis etc.).</td>
</tr>
<tr>
<td>Potential CD</td>
<td>Positive CD serology but normal small intestinal histology.</td>
</tr>
<tr>
<td>CD autoimmunity</td>
<td>Positive serology on at least two occasions and no knowledge of histological finding.</td>
</tr>
<tr>
<td>Refractory CD</td>
<td>Malabsorptive symptoms and villous atrophy (VA) despite GFD for &gt;12 months.</td>
</tr>
</tbody>
</table>

Children

The clinical picture has shifted and today children more commonly present with diffuse gastrointestinal (e.g. recurrent abdominal pain) or extraintestinal (e.g. fatigue) complaints rather than diarrhea, steatorrhea and growth restriction (i.e. features of classical CD) [4]. A recent American study found that 43% of the pediatric patients presented with non-classical symptoms, 34% had a classical CD and as many as 23% were screening detected. Most patients had a normal body mass index (BMI) [28]. The results confirmed previous Swedish findings [13]. Features of classical CD are, however, probably slightly more common in younger ages [13, 80].
Fig 3. Growth chart from a girl diagnosed with CD at about 8.5 years of age.

Adults
Also among adults the clinical presentation has changed over the years. This is probably partly due to a greater awareness and improved diagnostic tools resulting in earlier diagnosis [4, 71]. Recently published Irish data showed that diarrhea occurred in about 70% of patients before 1986 but in less than 40% after 2010 [14]. Subclinical CD is now more common and the disease can be unmasked when investigating patients with for example anaemia (probably one of the most common modes of presentation [81]), osteoporosis or depression [4, 14].
Diagnostics

The diagnostic tools have improved over the years and in particular for children the diagnostic work-up has been simplified. For a long time biopsy-proven VA has been considered the gold standard for diagnosing CD [82-84]. The evolving knowledge and development of reliable serological markers [85] have, however, come to question this standard. The progress in diagnostic work-up is well illustrated by the guidelines published by ESPGHAN (Table 2) [82, 86].

Table 2. Summary of diagnostic criteria for CD* proposed by ESPGHAN [82, 86].

<table>
<thead>
<tr>
<th>Year</th>
<th>Proposed diagnostic work-up</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1974</td>
<td>Initial biopsy*</td>
<td>Structurally abnormal mucosa (i.e. VA, crypt hyperplasia)</td>
</tr>
<tr>
<td></td>
<td>Biopsy on GFD (healing biopsy)</td>
<td>Improvement of mucosal changes</td>
</tr>
<tr>
<td></td>
<td>Biopsy after gluten challenge</td>
<td>Recurrence of mucosal deterioration</td>
</tr>
<tr>
<td>1990</td>
<td>Initial biopsy*</td>
<td>Structurally abnormal mucosa</td>
</tr>
<tr>
<td></td>
<td>Clinical follow-up</td>
<td>Clinical improvement on GFD</td>
</tr>
<tr>
<td></td>
<td>Serological markers support diagnosis</td>
<td>Elevated/positive at diagnosis, normalized on GFD</td>
</tr>
<tr>
<td>2012</td>
<td>Alternative 1</td>
<td>Structurally abnormal mucosa</td>
</tr>
<tr>
<td></td>
<td>Biopsy</td>
<td>Improvement on GFD</td>
</tr>
<tr>
<td></td>
<td>Clinical follow-up</td>
<td>Elevated at diagnosis and normalized on GFD</td>
</tr>
<tr>
<td></td>
<td>Serological markers support diagnosis</td>
<td>Compatible with CD</td>
</tr>
<tr>
<td></td>
<td></td>
<td>tTG elevated &gt;10x cut-off (on two occasions or, preferably, in combination with EMA positivity)</td>
</tr>
<tr>
<td></td>
<td>Genetic testing</td>
<td>HLA DQ2 or DQ8 positive</td>
</tr>
<tr>
<td></td>
<td>Clinical and serological follow-up</td>
<td>Clinical and serological improvement on GFD</td>
</tr>
</tbody>
</table>

*For children

On a gluten-containing diet

VA=villous atrophy, GFD=gluten-free diet, tTG=tissue transglutaminase, EMA=endomysial antibodies
The proposed diagnostic procedure from ESPGHAN (where biopsy can be omitted) still only applies for children where validation studies have strengthened the recommendation [87]. In adults, biopsy with findings compatible with CD are required for diagnosis [4, 71, 83].

Serologic testing

The development of serologic testing for CD has revolutionised the diagnostic work-up. At present, the serological markers used in CD outperform almost all antibody tests used for other autoimmune and inflammatory conditions [85].

The first serologic test for CD was launched in the 1980s and measured anti-gliadin antibodies (AGA) [88]. The AGA test was a welcomed tool in the limited diagnostic arsenal but unfortunately both sensitivity and specificity turned out to be quite low (80-90%) yielding a positive predictive value only around 30% [85]. Following the AGA, the endomysial antibody (EMA) test was developed [89]. EMA showed a much higher sensitivity and specificity than AGA (in average 95 and 99% respectively), particularly in patients with total VA [85]. Unfortunately EMA tests are (in contrast to other CD antibody tests which are performed on enzyme-linked immunosorbent assays (ELISAs)) based on immunofluorescence and need expensive substrates in form of monkey oesophagus or human umbilical cord. The test also depends on individual reading, consequently affecting reproducibility and hampering standardization. Altogether this has led to a decreased use of EMA in clinical practice [85]. Despite the importance of the enzymatic activity of tissue transglutaminase (tTG) in disease pathogenesis, German researchers, in 1997, proved it also to be the major autoantigen in CD [90]. Measurement of IgA antibodies against tTG is now, in most parts of the world, the test of choice when diagnosing and monitoring CD [71]. Like EMA, tTG antibodies have a high diagnostic performance with both sensitivity and specificity around 98% [85]. In general the specificity is considered to be slightly lower than for EMA but given the advantages of the test this small difference is usually considered negligible [85, 91]. In addition, tests quantifying antibodies against deamidated gliadin peptides (DGP) are available. It is the deamidated gliadin peptide that is involved in CD pathogenesis and therefore DGP is considered more specific than AGA [85]. Studies, however, disagree regarding the usefulness of DGP [92] but some suggest that it can be a helpful tool in children below 2 years of age where the serological markers in general have less sensitivity [93]. Likewise DGP can be useful in individuals suffering from IgA deficiency [85]. Considering the increased prevalence of CD in people with IgA deficiency [94] there is of course a need for serological markers...
not based on IgA antibodies and here IgG DGP is commonly used in combination with IgG tTG [85].

It is the high performance of the tTG tests [85] as well as the observation that the level of antibodies seem to correlate to the grade of mucosal damage [95] that has led to the revision of the ESPGHAN guidelines [86]. It is, however, worth pointing out that tTG antibodies can be temporarily elevated during an infectious episode [96] emphasizing the need for repeated serologic testing before diagnosing CD [86].

The serological markers have also become useful tools for monitoring CD patients. An increase in antibody levels usually reflects a deviation from dietary adherence [71], caution is, however, advocated since negative serology does not exclude continued gluten intake and incomplete mucosal healing [83, 85, 97].

Genetic testing
As described earlier a genetic predisposition is a prerequisite for CD [35]. While the HLA molecules linked to the disease are common in the general population the absence of them can basically rule out a CD diagnosis [4, 35, 86]. Consequently genetic testing is now routinely used in the diagnostic work-up [86, 98]. The negative predictive value of the absence of the disease associated HLA DQ2 and DQ8 is almost 100% [4]. Using HLA analyses as a tool also reduces the need for repeated testing in risk groups since individuals without risk alleles can be excluded from these surveillance programs (see page 28) [86].

Small intestinal biopsy (histopathology)
The histological findings seen in the upper small intestine in CD has been thoroughly described by Michael Marsh [99]. His grading of the enteropathy was later modified by Oberhuber and their classification is still commonly used [100]. However, additional efforts to classify the histological findings and reduce the inter-observer variability have been made by for example Corazza and Villanacci [101, 102]. In Sweden, biopsy findings are classified according to the Systematic Nomenclature of Medicine (SnoMed) clinical terms [84, 103]. An overview of the relationships between the different classifications are presented in Table 3.
The main histopathological features of CD are characterized by an increasing number of intraepithelial lymphocytes, shortened villi and crypt hyperplasia (Fig 4) [99, 100]. The mucosal changes can be patchy [104, 105] wherefore current guidelines recommend that biopsy samples are taken from several sites [83, 86]. It is important to be aware that the described histopathological findings also can be seen in other diseases (e.g. giardiasis, Crohn disease) and are thus not pathognomonic for CD [83, 106]. However, in the validated Swedish biopsy material used in this thesis, diagnoses other than CD were uncommon [84].

<table>
<thead>
<tr>
<th>Table 3. Small intestinal histopathology classifications – a comparison [84, 99-103].</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Classification</strong></td>
</tr>
<tr>
<td><strong>Marsh-Oberhuber classification</strong></td>
</tr>
<tr>
<td><strong>Marsh-Oberhuber description</strong></td>
</tr>
<tr>
<td><strong>Corazza et al. classification</strong></td>
</tr>
<tr>
<td><strong>SnoMed Codes</strong></td>
</tr>
<tr>
<td><strong>KVAST/Alexander classification</strong></td>
</tr>
<tr>
<td><strong>Characteristics</strong></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

*Marsh type 4 is not included in this overview because such lesion are very rare and cannot be identified through SnoMed codes.

*Codes and classifications used in the studies included in this thesis.

#Increased intraepithelial lymphocyte count (usually >25-30/100 epithelial cells).
Risk groups (subjects for regular surveillance)

From the previous section on genetics it is easy to conclude that first- and second-degree relatives are at increased risk of developing CD [38]. International guidelines therefore recommend screening in first-degree relatives [71, 86]. The risk of CD is also elevated in people with Down syndrome (OR=6) [107] as well as Turner (OR=3) [108] and Williams (CD prevalence 9.5%) [109] syndromes but also in individuals with IgA deficiency (prevalence increased 35-fold) [94]. CD is also rather strongly associated with T1DM, at least partly due to shared genetics [110]. Studies show a prevalence of CD around 6-8% in patients with T1DM [7]. In Sweden and many other countries individuals with T1DM but also Down syndrome are regularly screened for tTG antibodies [83, 86, 111, 112]. It is also recommended that other risk groups like the above mentioned syndromes and individuals with thyroid disease, autoimmune hepatitis and inflammatory bowel disease are evaluated [71, 86, 111]. All the latter autoimmune diseases associated with CD [5, 81].

Figure 4. Small intestinal biopsies.
a) Normal mucosa  b) Increased number of IEL  c) Partial VA  d) Total VA
(Photomicrographs by courtesy of Dr H. Nobin)
Treatment

Gluten-free diet

Ever since Willem Karel Dicke made the observation that wheat seemed to cause CD [18], a gluten-free diet (GFD) has been the only available treatment [15]. Since the disease is chronic this dietary restriction has to be life-long [71]. Improvement of the gastrointestinal symptoms is most often seen within weeks and usually precedes improvement of celiac serology and mucosal healing [15, 113]. Data on normalization rate of CD serology vary but elevated levels can persist for years [113, 114]. Likewise, studies regarding the frequency of mucosal healing are very inconsistent (8-81%!? [115-117]. Even though different follow-up times can account for some of the disparities further investigations are probably needed.

Gluten is found in wheat (including triticale, semolina, spelt and Khorasan wheat), rye and barley (malt) [1, 71]. Several studies have by now established that non-contaminated oats can be considered safe to include in the diet [118, 119]. Efforts have been made to establish the amounts of gluten tolerated in the diet and results vary (10-100 mg/day) but a general recommendation is an intake of less than 10-20 mg/day [71, 83, 120]. The global Codex Alimentarius (part of the World Health Organization, WHO) states that foods that shall be allowed to be labelled gluten-free have to be naturally free from gluten (a measured gluten level of ≤20 mg/kg) or processed to <100 mg/kg [1]. Despite labelling, caution is needed since there are many hidden sources of gluten (e.g. sauces, shared food preparation equipment etc.) [71]. Consequently all patients should be referred to a dietician with extensive knowledge regarding all these “dietary pitfalls” as well as other problems that can arise while adhering to a GFD [71]. Data indicate that a GFD tend to include less amounts of fibres than a “normal” diet [71, 121]. It can also be deficient in vitamins and minerals why patients might be advised to add a multivitamin supplement to their diet [121].

The compliance with the GFD vary [118, 122] and having to adhere to a diet so strictly can have adverse effects [122]. The consequences can be both psychosocial [122, 123] and economical [124, 125]. The rather high compliance seen in Swedish children [118] might of course be attributed to the privilege of receiving parts of the gluten-free products on prescription (free of charge). However, not to forget is that a GFD usually results in an improvement of symptoms and thereby often also an improved quality of life (QoL) [126].
Alternative treatments

Although symptoms and QoL often improve after introduction of a GFD, patients still ask for alternative treatments [126]. Furthermore, mucosal healing takes time, is not always complete [115] and some individuals with CD do not respond at all to the GFD. Patients with this disease, termed refractory CD, risk severe complications and are in great need of an alternative treatment [4]. As of today these patients are often treated with systemic immunosuppressants but topical steroids have also been tried [127]. However, treatments with a minimum of systemic side effects that could be used in all CD patients are of course preferred. There is a lot of ongoing research targeting different steps in CD pathogenesis. One area of interest is that of gluten detoxification where oral enzymatic therapies have been tried. Studies on proteases breaking down gluten and hereby making the peptides less immunogenic have been somewhat promising [128]. Another possible drug target is that of intestinal permeability where Larazotide acetate (a tight junction modulator) seem to improve symptoms in patients on a GFD [128, 129]. There are also ongoing experiments aiming to affect antigen presentation and immune response, here tTG inhibitors [130] and HLA blockers [128] might be effective but also drugs affecting T-cell infiltration and IL-15 expression [128]. As a sequel to the successful allergen-specific immunotherapies there are attempts to induce oral tolerance even in CD. Intradermal injections with vaccines containing gluten peptides are under investigation [128]. As a curiosity, hookworm therapy also seems to have some benefits. Hookworms down regulate the immune response to promote their own survival and by doing so they suppress the immunologic reaction to gluten, thereby inducing tolerance. However, there might be side effects to consider [128, 131]

Figure 5. Overview of pharmacological approaches in celiac disease. McCarville et al 2015 [128]. Reproduced with permission from Current Opinion in Pharmacology. Copyright © Elsevier Ltd.
In conclusion there are many promising trials but still many of the tested drugs seem to work best as adjuvants to a GFD.

**Associated conditions and complications**

CD is associated with a number of conditions [5, 132]. This is probably partly due to shared genetics but most likely also to environmental factors [110, 133, 134]. Several of the associated conditions were presented in the section about risk groups. Worth re-mentioning is the association with other autoimmune conditions [5, 135] and in particular T1DM [6]. The prevalence of CD in the T1DM population is around 6-8% [7, 133] and an interesting observation is that it is seems more common with a T1DM diagnosis preceding CD than the other way around [136]. Given the shared genetics [110, 134], the relative risk of 2.4 (95%CI=1.9–3.0), for CD patients to later develop T1DM, must be considered quite low [136].

CD is also associated with malignancies and in particular lymphoproliferative malignancies (standardized incidence ratios varying from 1.9 to >5, higher for some subtypes) [9, 132, 137, 138]. Studies indicate that the highest risk of lymphoproliferative malignancies is found in patients with persistent VA [10].

Other complications seen in CD can be attributed to malabsorption and inflammation and include anaemia, osteoporosis and vitamin deficiencies [81]. Likewise there is psychiatric comorbidity, perhaps a direct impact of the disease as well as a consequence of the burden of living with a chronic condition [122, 139, 140].

Determining what is an associated condition and what is a complication is challenging (but of some interest since they need to be approached differently). Although no distinct lines can be drawn, an attempt (not claiming to be complete) is presented in Table 4.
Table 4. CD associated conditions and complications.
Not arranged according to frequency.
Based on references [5, 7, 8, 11, 12, 81, 86, 94, 107-109, 132-135, 137, 140-146].

<table>
<thead>
<tr>
<th>Associated conditions</th>
<th>Complications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gluten-related disorders:</td>
<td></td>
</tr>
<tr>
<td>Anaemia</td>
<td></td>
</tr>
<tr>
<td>Dermatitis herpetiformis</td>
<td>Osteoporosis/osteopenia</td>
</tr>
<tr>
<td>Gluten ataxia</td>
<td>Vitamin/mineral deficiencies</td>
</tr>
<tr>
<td>Chromosomal anomalies:</td>
<td></td>
</tr>
<tr>
<td>Down syndrome</td>
<td></td>
</tr>
<tr>
<td>Turner syndrome</td>
<td></td>
</tr>
<tr>
<td>Williams syndrome</td>
<td></td>
</tr>
<tr>
<td>Autoimmune/inflammatory diseases:</td>
<td></td>
</tr>
<tr>
<td>T1DM&lt;sup&gt;iii&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Autoimmune thyroiditis&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Autoimmune hepatitis&lt;sup&gt;iii&lt;/sup&gt;</td>
<td>Elevated liver enzymes</td>
</tr>
<tr>
<td>Inflammatory bowel disease</td>
<td></td>
</tr>
<tr>
<td>Arthritis</td>
<td></td>
</tr>
<tr>
<td>Psoriasis</td>
<td></td>
</tr>
<tr>
<td>Hyposplenism</td>
<td>Infections</td>
</tr>
<tr>
<td>Aphtous ulcers</td>
<td>Aphtous ulcers</td>
</tr>
<tr>
<td>IgA deficiency</td>
<td></td>
</tr>
<tr>
<td>Neuropathy</td>
<td>Neuropathy</td>
</tr>
<tr>
<td>Delayed puberty</td>
<td>Delayed puberty</td>
</tr>
<tr>
<td>Short stature</td>
<td></td>
</tr>
<tr>
<td>Malignancies?</td>
<td>Malignancies</td>
</tr>
<tr>
<td>Cardiovascular diseases&lt;sup&gt;iii&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Pulmonary disorders&lt;sup&gt;ii&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Renal diseases&lt;sup&gt;ii&lt;/sup&gt;</td>
<td>Depression</td>
</tr>
<tr>
<td>Epilepsy</td>
<td></td>
</tr>
<tr>
<td>Migraine</td>
<td>Headache</td>
</tr>
</tbody>
</table>

<sup>a</sup> Comorbidities controlled for in study II
<sup>ii</sup> Comorbidities controlled for in study III
Refractory celiac disease

A special but rare entity is the refractory CD (RCD) [4, 147]. In the “Oslo definitions” it is defined as “persistent or recurrent malabsorptive symptoms and signs with VA, despite a strict GFD for more than 12 months” [1]. Generally most patients with RCD test negative for CD antibodies and if not, dietary adherence should be questioned. Naturally a thorough assessment regarding the diet always has to be performed before establishing the diagnosis [1]. RCD can be divided into two types; in type 1, a normal IEL phenotype is found whereas in type 2, a clonal expansion of abnormal IELs is seen [1, 4]. RCD type 2 is associated with a poorer prognosis than type 1, with a high risk of progression to enteropathy associated T-cell lymphoma or ulcerative jejunitis and a considerable mortality [4, 147, 148]. Treatment for RCD consists of immunosuppressive drugs and sometimes autologous stem cell transplantation [148].

Dermatitis herpetiformis and gluten ataxia

Two particular gluten related conditions are the cutaneous disease dermatitis herpetiformis (DH) and the neurological disorder gluten ataxia. It is possible that CD antibodies generated in the small intestine contribute to the development of these extraintestinal manifestations [4, 149]. DH is characterized by clusters of itchy papules and vesicles on the skin. Elbows, buttocks and knees seem to be sites of preference. IgA deposits are seen in the perilesional skin, confirming diagnosis. At small intestinal biopsy, more than two thirds of the DH patients show VA and the disease responds to GFD [1, 141]. Gluten ataxia is defined as “an idiopathic sporadic ataxia and positive serum AGA even in the absence of duodenal enteropathy” [1]. Patients often show cerebellar atrophy and they all appear to have abnormalities affecting the vermis. GFD usually leads to improvement but early diagnosis seems more beneficial [142].

Infections following celiac diagnosis

In this thesis, infections as complications of CD are investigated more thoroughly. This is an association that has received increased attention during the last decade [11]. Studies have shown an excess risk of more severe infections including influenza requiring hospitalization [150], tuberculosis [151] and *Clostridium difficile* infections [152]. In addition, CD patients seem to have an increased risk of sepsis, in particular pneumococcal sepsis (HR 2.5) [12]. Although studies on pneumococcal infections are not that many [12, 153], data are sufficiently convincing for many countries to recommend pneumococcal vaccine to individuals with CD [71, 83]. The recommendation is further supported by a study showing a 28% increased risk of community-acquired pneumonia in unvaccinated individuals with...
CD compared to controls (a risk increase that not could be demonstrated in vaccinated subjects) [154].

It is possible that the increased susceptibility to infections in CD depends on an impaired spleen function [145] and/or an altered intestinal barrier [146, 155], subjects elaborated below.

Hyposplenism
The increased risk of infectious complications in CD is mainly believed to depend on the comparatively high prevalence of hyposplenism (20-40%) found in adult patients [156-158]. CD is regarded to be the disorder most frequently associated with hyposplenism, a term referring both to splenic hypofunction and atrophy [145]. However, many studies are quite old and rates found today might not be as high as estimated in earlier studies (19% in the study by Di Sabatino et al from 2006 as compared to the prevalence numbers of 33-79% seen in older studies) [157-159]. Hyposplenism seems more frequent in patients with additional autoimmune diseases (59%) as well as in those with complicated CD (80%) [158, 160]. A recent study demonstrated a reduced splenic volume in RCD whereas, surprisingly, the volume was enlarged in uncomplicated CD, indicating a large inter-individual variability [160]. The splenic function is most probably improved by a GFD although modern studies addressing this issue are lacking [156, 159].

An impaired function of the spleen results, among other things, in a reduced number of IgM-memory B-cells and defective opsonisation [145]. This predisposes to infections with encapsulated bacteria wherefore some experts argue that vaccines against meningococci and *Haemophilus influenzae* should be offered to individuals with CD in addition to pneumococcal immunization [71].

Increased intestinal permeability
The intestinal barrier and permeability seem to be altered in individuals with CD [155]. Gliadin peptides cross the mucosa through paracellular as well as transcellular routes and studies suggest that both might be affected, enabling rather intact peptides to enter the lamina propria, particularly in active CD [78, 79]. Several mechanisms are most likely responsible and in 2000, a protein resembling a cholera toxin was identified [161]. The protein, named zonulin, acts by inducing disassembly of tight junctions and the expression is increased in the acute phase of CD. Likewise, the hydrophobicity of the duodenal mucous layer might be impaired [162]. A GFD seems to improve the barrier but recent data suggest that the function not becomes entirely comparable to healthy controls [155].
It is also possible that an impaired barrier enables bacterial translocation [146] and thereby contributes to the increased prevalence of severe infections in CD. However, further research is needed.

Mortality
Whereas several studies have found a modest (about 40%) increase in overall mortality amongst CD patients [163-165] others have failed to confirm these findings [166]. Causes of death vary but include malignancies, gastrointestinal and cardiovascular diseases [163, 165, 167]. As previously mentioned, the mortality rate in RCD type 2 is significantly higher than in uncomplicated CD and is mostly attributed to the enteropathy related T-cell lymphoma and ulcerative jejunitis [4, 148].
Infections

Respiratory syncytial virus infections

Respiratory syncytial virus (RSV) is probably one of the most common causes of respiratory tract infections in infants [168, 169]. The virus can be divided into two types; RSV type A and type B. Type A is probably more prevalent and seems to generate a higher viral load which might correlate to disease severity [168]. RSV infections are seasonal and generally occur in winter months parallel to influenza [168]. By two years of age about 95% of all children have been infected with RSV, most of them managing the disease without hospital care [168]. Even so, the global burden of RSV infection on the health care system is considerable [168, 169]. The RSV Global Epidemiology Network has recently estimated that, in 2015, around 33 million lower respiratory tract RSV infections resulted in about 3.2 million hospital admissions and 59,600 in-hospital deaths in children below 5 years [169]. The main independent risk factors for severe disease are young age and prematurity. In addition, children with immunological diseases, chromosomal abnormalities and heart and lung diseases are more susceptible [168]. Hospitalization is also more prevalent in children exposed to smoking both pre- and postnatally [168, 170].

Sepsis

Sepsis is not really a specific disease but rather a syndrome triggered by an acute infection. Consequently the characteristics are formed by both pathogen factors and host factors [171, 172]. Recently new criteria for defining sepsis was launched – Sepsis-3 [171]. The paper discourages continued use of the terms systemic inflammatory response syndrome (SIRS) and severe sepsis. Instead, sepsis should be defined as “a life-threatening organ dysfunction caused by a dysregulated host response to infection” and septic shock as “a subset of sepsis in which underlying circulatory and cellular metabolism abnormalities are profound enough to substantially increase mortality” [171]. Previous definitions have involved the SIRS concept and an infection combined with two or more SIRS criteria (altered body temperature, tachycardia, tachypnea/desaturation and elevated/low white blood cell count) has been classified as sepsis, severe sepsis if organ dysfunction also was present [171, 173]. These definitions are the ones commonly used for sepsis diagnoses in the study included in this thesis. Sepsis is a very severe condition and probably the primary cause of death from infections. The mortality rates in some studies reach levels as high as 26% [172]. Early identification of the illness is crucial, as is a greater understanding of which factors influence susceptibility and outcome. Patients with an increasing number of comorbidities seem to have a higher mortality rate [173].
Pneumococcal infections

*Streptococcus pneumoniae* is a Gram-positive encapsulated bacteria. It is one of the major causes of respiratory tract infections worldwide but can also cause more invasive infections [174]. Invasive pneumococcal disease (IPD) is usually defined as growth of *Streptococcus pneumoniae* in a culture from a normally sterile site (i.e. blood, cerebrospinal fluid (CSF), pleural effusion, pericardial fluid, or synovial fluid) [174] and this is also the definition used in the study in this thesis.

Since the introduction of pneumococcal vaccination, both to children, elderly and certain risk groups, the number of IPD cases has declined but is still substantial [175]. All pneumococcal serotypes are not covered in the vaccines and in addition vaccine programmes tend to cause a shift in circulating serotypes to “non-vaccine types” [174, 175]. Since IPD is a severe condition, with a considerable mortality and global burden, research regarding this infection is still of priority [174, 176]. In Sweden pneumococcal vaccine was introduced in the childhood immunization program in 2009 [177]. It has been recommended to elderly (≥65 years) since 1994 [178] and during the years this recommendation has expanded to certain risk groups but (in contrast to other countries [83]) not yet to individuals with CD. As mentioned earlier, individuals with an impaired spleen function are at greater risk of infections with encapsulated bacteria and are therefore offered vaccine [145]. However, since splenic size and function are not routinely evaluated in CD patients, they are rarely included.

The complement system

The innate immunity is the first line of defence against pathogens and here the complement system plays an important role [179]. The complement system is an intricate network of serum and membrane proteins acting in a cascade-like manner [179]. Its main functions are the opsonisation of the surface of the pathogens, activation and recruitment of inflammatory cells and the direct killing of bacteria through formation of the membrane attack complex (MAC) [179, 180]. By facilitating phagocytosis the complement system also acts as a link between innate and adaptive immunity since it promotes antigen recognition by B-cells [179]. The complement system can be activated through three different pathways [179, 180]. The classical pathway is activated by antibodies binding to antigens and thereby exposing a binding site for complement. It can also be activated by acute phase pentraxins, for example C-reactive protein (CRP). The alternative pathway does not require antibodies but is autoactivated. There is a continuously ongoing activation that is amplified on foreign surfaces that lack the inhibiting factors present on host cells. The lectin pathway is activated by mannann-binding lectins (MBL) and ficolins...
that recognize carbohydrate structures on microbes. All pathways converge at the level of C3 and end in the fusion of C5b, C6, C7, C8, and C9 (=MAC) [179, 180].

Figure 6. Schematic overview of the complement system.

The complement system is essential in the defence against pneumococcal infections even though the capsule surrounding the bacteria possesses protective properties [179, 180]. Pneumococci mainly activate the classical pathway but also the lectin pathway seems to be of importance [179-181]. Both individuals with complement deficiencies as well as MBL deficiency are at increased risk of pneumococcal disease or adverse outcome in IPD [179]. Several studies have shown an association between variant MBL alleles and celiac disease [182] but the consequences for the protection against pneumococcal infections have not been elucidated.
The complement system in celiac disease
The role of the complement system in CD pathogenesis is poorly investigated and most studies are old. A few studies have seen an increased deposition of complement in the lamina propria of untreated CD patients compared to treated patients [183]. Research on sera instead seem to show lower levels of for example C3 and C4 [184, 185]. Data also suggest an increased complement activation in CD based on an observation of increased fractional catabolic rate of radioactively labelled C3 [186].
Rationale for this thesis
The association between CD and infections has been elaborated both in the section about environmental risk factors and in the section about complications.

Many efforts have been made to establish which pathogen or pathogens that together with gluten induce the inflammatory and autoimmune process ending in CD. Respiratory infections have been studied [44, 65, 66] but studies specifically examining the possible association with more pronounced RSV infection and CD were lacking.

CD has also been linked to severe infectious complications [11, 151]. For example, a Swedish study noticed an about 60% increased risk of sepsis [12]. Since the outcome in sepsis can be influenced by comorbidities [173] the question as to whether CD could affect the survival in sepsis was raised. The same Swedish study reported an even higher risk of pneumococcal sepsis than sepsis in general (HR 2.5) [12], confirming findings from an English study published the same year [153]. However, since the Swedish study was limited to inpatient CD data and the English study included pneumonia they might have slightly overestimated the IPD risk. Considering the severity of IPD we believed that the association between CD and IPD deserved further attention, resulting in study III.

The association with severe infections is mainly believed to be due to the hypoplasplenism that sometimes coexists with CD [145, 157]. Nevertheless, an impaired splenic function is not seen in all individuals with CD and rarely in children [187]. It is also believed to improve during GFD [156] whereas the increased risk of IPD and pneumonia seems to persist beyond one year of follow-up [12, 153, 154]. Therefore, investigating other potential immunological mechanisms that could have impact on the susceptibility to infections seemed appropriate. The role of the complement system in CD is rather unexplored and previous studies regarding possible alterations that could influence the risk of pneumococcal infections could not be found wherefore we, in study IV, decided to investigate complement activation in response to pneumococci in CD patients and controls.
Objectives

The main objective of this thesis was to increase the knowledge about the association between celiac disease and infections but also to gain further understanding about the mechanisms contributing to the increased risk of infectious complications.

Specific objectives

I. Is there an association between viral bronchiolitis, in particular RSV infection, and later CD?
II. Does CD influence survival in sepsis?
III. Is there an increased risk of invasive pneumococcal disease in CD?
IV. Does the complement activation in response to *Streptococcus pneumoniae* differ between young individuals with and without CD?
Methods

Setting and personal identity number

All studies in this thesis were performed in Sweden and were accordingly based on data and material from Swedish inhabitants. The first three studies used the same cohort of patients with biopsy-verified CD which is described further down in the methods section.

Sweden is an excellent country for register-based and epidemiological research since all citizens are assigned a personal identity number (PIN), enabling linkage between different registers and data sources [188]. Already in the 17th century the church in Sweden started to keep local registers on parish members and from the middle of the 18th century there is evidence of population statistics [188, 189]. In 1947 the PIN was introduced [188]. Initially it consisted of date of birth and a three-digit number in the end. In 1967 a fourth digit, a so called check-digit, was added which is supposed to verify that the date of birth and the three-digit number are correct. Since the early 1990s the Swedish Tax Agency has the full responsibility for the PIN, which until then had been handled by local parishes [189]. Change of PIN is not very common and is usually due to incorrect registration of birth date in immigrants. Likewise, since the PIN is sex-specific, incorrect registration at birth or change of sex subsequently result in change of PIN [188].

Registers/data sources

Statistics Sweden

In 1967 the local population registers were computerized. This enabled the government agency Statistics Sweden (through the PIN) to establish the Total Population Register (TPR) [189]. The TPR contains data on name, age, sex, place of birth, place of residence, civil status, relations (married couples, child-parent), citizenship and immigration. Also dates of death are recorded. Since 1991 these data are delivered to Statistics Sweden by the Swedish Tax Agency (an assignment that was previously carried out by local parishes) [189]. Demographical data from TPR were used in all the first three studies in the thesis.

Statistics Sweden also maintain other population-based registers, for example the Swedish Education Register and the Swedish Occupational Register. Data from these two registers were used as proxies for socioeconomic status in studies I-III.
The Swedish Patient Register
The Swedish Patient Register is administered by the National Board of Health and Welfare. The register was established in 1964 and became nationwide in 1987 and PIN-based in 1993 (reconstructions have been made so all diagnoses from 1987 and onwards can be linked to PINs) [190]. The register initially only contained inpatient diagnoses but in 2001, hospital-based outpatient diagnoses were added. Currently the coverage of discharge diagnoses and medical procedures is close to 100%. Reports on outpatient visits are lacking and the coverage was about 80% in 2011 [190], but the number of missing data seems to be declining. Diagnoses are reported with International Statistical Classification of Diseases and Related Health Problems (ICD) codes and the positive predictive values (PPV) vary between diagnoses but in general the specificity is good [190]. Data from the Patient Register were predominantly used in studies I and II but also to identify potentially confounding comorbidities in study III.

The Swedish Medical Birth Register
The Medical Birth Register was established in 1973 and contains information about all pregnancies that result in a delivery. Data on pregnancies, deliveries and the newborn babies can be found and include information on previous pregnancies, smoking, gestational length, mode of delivery, mother and child diagnoses, anthropometric data on the child etc. Inclusion in the register is not optional, resulting in a high coverage. Nevertheless, records are missing in about 1.5% of infants and a somewhat higher proportion has incomplete records where one or more variables are lacking [191]. Perinatal data from this register were used in study I which involved small children with CD.

The Swedish Cause of Death Register
As previously mentioned, Sweden has a long tradition of population statistics. As early as 1749 the Swedish parliament decided that causes of death should be recorded [192]. From 1911 all causes of death (and not only the “important” ones) have been continuously documented. Between 1911 and 1993 data were collected by Statistics Sweden but since 1994 the National Board of Health and Welfare is responsible. Annual statistics are published and data from 1952 and onwards are available electronically [192]. In the death certificates ICD codes are used. An underlying cause of death is required but it is also possible to report up to 48 contributing causes [192]. The completeness is high and an underlying cause of death is only missing in less than 1%. Determining what is an underlying cause of death and what is a contributing cause can be difficult and the declining autopsy rate has not facilitated this assignment [192]. Nonetheless, the Cause of Death Register is in
general considered a reliable source for statistics and research and was used in study II in this thesis.

The Swedish Public Health Agency
The Swedish Public Health Agency was formerly called the Swedish Institute for Infectious Disease control. It is the government agency that is responsible for the surveillance of communicable and notifiable diseases. In study III data from this agency were used to identify IPD. IPD became a notifiable disease in Sweden in 2004. The reporting system includes both active reporting from clinicians but also automated reports from all microbiological laboratories in Sweden leading to an almost 100% coverage (from 2005 and onwards).

Study populations and designs
Studies I-III in this thesis were based on a nationwide cohort of individuals with biopsy-verified CD and their matched controls. The biopsies were performed between July 1969 and February 2008 and collected between October 2006 and February 2008. They were identified through computerized searches of all 28 pathology departments in Sweden and the searches were performed by local technicians by using the SnoMed codes described in Table 3 for identification [84, 163]. During this period small intestinal biopsy was considered gold standard to diagnose all CD in Sweden [84]. Altogether 381,043 small intestinal biopsies were identified. 28,654 of these turned out to be duplicates. Due to data inconsistency another 986 biopsies were excluded. The remaining 351,403 biopsies represented 287,586 individuals. These individuals were divided into three groups: CD (VA/Marsh 3), inflammation (Marsh 1-2) and normal mucosa (Marsh 0). A few patients were later excluded due to biopsies that could potentially originate from ileum, incorrect PINs and matching problems. CD serology was retrieved for part of the material (CD: n=11,612; inflammation: n=5,302; normal mucosa: n=121,952). Potential CD (previously named latent CD) was defined as normal mucosa combined with a positive CD serology up to 180 days before and until 30 days after biopsy. The distribution of biopsy findings are presented in Table 5 [84, 163].
Table 5. Distribution of biopsy findings and some characteristics of study participants [84, 163].

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>CD/VA N=29,096</th>
<th>Inflammation N=13,306</th>
<th>Potential CD N=3,719</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at study entry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-19 years</td>
<td>11,802</td>
<td>1,226</td>
<td>941</td>
</tr>
<tr>
<td>20-39 years</td>
<td>5,312</td>
<td>3,546</td>
<td>1,150</td>
</tr>
<tr>
<td>40-59 years</td>
<td>6,477</td>
<td>4,149</td>
<td>1,064</td>
</tr>
<tr>
<td>≥60 years</td>
<td>5,505</td>
<td>4,385</td>
<td>564</td>
</tr>
<tr>
<td>Female, No. (%)</td>
<td>18,005 (61.9)</td>
<td>7,497 (56.3)</td>
<td>2,310 (62.1)</td>
</tr>
</tbody>
</table>

A subset of the material was validated in a blinded examination and pathologists correctly classified 90% of the samples with VA whereas inflammation was less accurate. Likewise, to further validate the material, patient records were manually reviewed for 160 individuals (CD/VA: n=121, inflammation: n=39) [84]. Data on all biopsies were sent to Statistics Sweden for matching with controls. For each individual up to 5 controls were identified. Matching criteria included age, sex, county and calendar period (for having biopsy). In addition, controls were not to have a record of prior small intestinal biopsy. After removal of some individuals due to data irregularities and matching problems 228,632 control individuals remained [84, 163].

Study IV did not use this CD cohort but instead used computerized patient records to identify children with CD attending Kalmar County hospital. Letters were sent out with an invitation to participate. Also control patients were identified and invited at Kalmar County Hospital, mainly through the Pediatric Outpatient Clinic where nurses informed about the study. All children were born between 1999 and 2008. Younger children were not invited due to the introduction of pneumococcal vaccine in the national immunization program in 2009 [177]. Likewise, children with an ongoing infection or other autoimmune diseases than CD were excluded. Since complement activity is independent of age and sex in this age group [193, 194], we did not require these parameters to be equal in the case and control group and therefore no matching procedure was carried out. Before study start an attempt to calculate a suitable sample size was performed. Considering p-values <0.05 significant and aiming for a study power of 80%, calculations resulted in a need for about 10-20 study participants in each group. Since the calculations were so uncertain we decided to aim at including just over 20 individuals in each group.
Study I
Respiratory syncytial virus and later celiac disease – a case-control study

In this study we wanted to examine the association between RSV (or unspecified viral bronchiolitis) and early-onset CD. Consequently we restricted our CD cohort to those who had undergone biopsy before two years of age and by doing so we identified 4,589 children. Since RSV not was introduced as an ICD code in the Patient Register until 1987 we further restricted our dataset to individuals born after 1st of January 1987. This left us with 3,835 children with CD (19,102 controls), 97 children with inflammation (487 controls) and 280 children with potential CD (1,397 controls). To obtain information on RSV and unspecified bronchiolitis diagnoses these data were linked to the Patient Register.

Demographic data were retrieved from the TPR, comorbidities from the Patient Register and perinatal information from the Medical Birth Register.

Conditional logistic regressions (see page 52) were used to obtain ORs for having attended hospital for an RSV infection or unspecified viral bronchiolitis prior to CD diagnosis. Several a priori defined subanalyses were performed including one where we looked at the risk of having had an RSV infection within a year prior to CD diagnosis compared to more than one year prior but also according to characteristics of study participants. ORs were adjusted both for perinatal (intrauterine growth retardation, caesarean section and maternal smoking) and socioeconomic (maternal education) factors.

Study II
Celiac disease and survival in sepsis – a longitudinal register-based study

CD has been linked to an excess risk of sepsis. In this study we wanted to investigate whether CD also could influence the outcome in sepsis. Through linkage of the described CD/control cohort to the Patient Register 5,470 individuals who, at some point, had received a sepsis diagnosis were identified (1,432 patients with CD (4.9%) and 4,038 controls (2.8%)). These data were subsequently linked to the

RSV: ICD9: 480B; ICD10: J20.5, J21.0, B97.4 and J12.1
Any viral bronchiolitis: ICD9: 466; ICD10: J20 and J21
Total Population Register to retrieve death dates and to the Cause of Death Register to get mortality details.

Relevant ICD codes (for sepsis) used in the study:
ICD7: 053, 057.1
ICD8: 036.0-1, 038.0-2, 038.8, 038.99
ICD9: 036C, 038
ICD10: R65.0-1, A39/A39.2, A40-41

Data on comorbidities (T1DM and autoimmune thyroiditis) were also retrieved from the Patient Register. Demographic data were obtained from the TPR.

Cox regressions (page 52) were used to estimate risk rates for overall death and death from sepsis. A model that allowed for staggered entry was chosen since the original study (data collection) enrolled participants at date of biopsies and matched controls that were alive at the corresponding date. Follow-up ended at death, emigration or end of study (31st of December 2009), whichever occurred first. Several subanalyses were performed including one where we restricted the study population to include only those who received a sepsis diagnosis after being diagnosed with CD. HRs were adjusted for sex, age at sepsis diagnosis, subtype of sepsis, calendar period at sepsis, socioeconomic status, level of education and country of birth. In a separate analyses we also adjusted for T1DM and autoimmune thyroiditis.

**Study III**

**Celiac disease and invasive pneumococcal disease – a population-based cohort study**

Study III aimed to confirm findings from previous studies that had found an increased risk of IPD in individuals with CD. The cohort of CD patients and controls was linked to data on IPD. These data were retrieved from The Swedish Public Health Agency who had access to all IPD cases since July 2004 when IPD became a notifiable disease in Sweden. Earlier data were obtained from computerized databases from the seven microbiological laboratories covering the Southern part of Sweden from 1987-1995 and onwards (different laboratories joined at different times). These local databases mainly contained data on individuals ≥18 years of age. Demographical data were collected from TPR and comorbidities were identified through the Patient Register.
To estimate risks (Hazard ratios) of IPD in CD patients compared to controls we used internally stratified Cox regression models (see page 52). Follow-up started at date of biopsy (and equal date in controls) and ended at IPD diagnosis, death, emigration or end of study (31st of December 2009) whichever took place first. Several subanalyses were performed including stratifications according to follow-up time and characteristics of study participants. Since many diseases are associated with an increased risk of IPD [195, 196] as well as with CD [7, 81, 143], adjustments for several of these potentially confounding comorbidities were performed (T1DM, chronic liver disease, end-stage renal disease, asthma and ischemic heart disease) and also for level of education, socioeconomic status and country of birth.

Study IV
Celiac disease and complement response to pneumococci
In this study we wanted to explore a possible mechanism that could contribute to the increased risk of IPD seen in CD (findings from study III). The main objective was to see if the complement activation in response to pneumococci differed between children with and without CD.

Clinical data
Questionnaires regarding diet, previous infections, hyposplenism and comorbidities were filled out by all study participants.

Blood sampling and preparation of plasma
Blood samples for analyses of complement activation products (C3a and soluble (s)C5b-9), C3, MBL, pneumococcal serology and tTG were collected from all study participants. Plasma-EDTA for complement analyses was centrifuged and frozen within 4 h from sampling.

Pneumococcal incubations
The method for removal of ethylenediaminetetraacetic acid (EDTA) (an anticoagulant and inhibitor of complement activation) had previously been developed and described by a coworker in the study [197]. In short; samples were thawed and spun through Bio-Spin P-6 gel columns (Bio-Rad Laboratories AB, Solna, Sweden) saturated with veronal-buffered saline (1.8 mM sodium barbiturate, 3.1 mM barbituric acid, 0.15 mM NaCl, 0.75 mM Ca\(^{2+}\), 2.5 mM Mg\(^{2+}\), pH 7.4) and lepirudin 50 µg/mL (Refludan®, Celgene, Windsor, UK). For stimulation we used a Streptococcus pneumoniae (serogroup 23F) that had been isolated from a patient suffering from an invasive infection. This isolate was retrieved from the Department of Clinical Microbiology at Kalmar County Hospital, Sweden. Prior to stimulation,
pilot tests were performed to determine the bacterial load and incubation time required to activate the complement cascade. Initially we evaluated stimulation with 100, 1,000 and 10,000 colony forming units (CFU)/mL aiming at imitating the pathogen load seen in a bacteremia. However, these concentrations failed to generate measurable complement activation in vitro and a bacterial concentration of $10^8$ CFU/mL was required. Also different incubation times were assessed (15, 30 and 60 minutes) and since no difference was observed between 30 and 60 minutes, an incubation time of 30 minutes was applied when stimulating the study samples. Pneumococcal stimulation was carried out by mixing 20 µL of S. pneumoniae in NaCl (0.9%) with 180 µL plasma. As control 20 µL NaCl without bacteria was added. Incubation was done in polypropylene microtubes, 30 min in 37 °C waterbath. As additional control 20 µL NaCl mixed with 180 µL plasma, without incubation was included. The complement reaction was stopped by adding 0.2 M EDTA (10 mM final concentration). Analyses were carried out as duplicates or triplicates, depending on sample volume and samples were centrifuged and frozen at -80 °C prior to complement analysis.

**Analysis of complement activation**

To estimate complement activity we chose to measure levels of C3a and sC5b-9 since they are the terminal components of the complement cascade and thereby well reflect the function of the entire system. The complement levels were assessed by ELISA. Diluted plasma was incubated in wells coated with monoclonal antibody 4SD17.3, which is specific for a neoepitope in C3a and serves as the capture antibody. C3a was detected using a biotinylated polyclonal anti-human C3a, followed by HRP-conjugated streptavidin (GE Healthcare, Little Chalfont, UK). This method was described by Nilsson Ekdahl et al already in 1992 [198] and has since then been used frequently [199]. Also detection of sC5b-9 in plasma was performed as previously described [200]. Plasma was diluted and added to wells of microtiter plates coated with the anti-neoC9 monoclonal antibody aE11 (Diatec Monoclonals AS, Oslo, Norway). Soluble C5b-9 was detected with a polyclonal biotinylated anti-C5 antibody (Acris, Herford, Germany), followed by HRP-conjugated streptavidin. Zymosan-activated serum calibrated against a known C3a and sC5b-9 concentration was used as standard. The complement response caused by pneumococci was estimated by comparing levels in stimulated and non-stimulated samples.

**C3**

To confirm the sex and age independency of complement activity we also measured the level of C3 in non-stimulated plasma. Likewise, the ratio C3a/C3 is considered a more sensitive marker than C3a alone [201]. C3 was measured by
nephelometry (Beckman Coulter Immage 800, Bromma, Sweden) using Immunochemistry Diagnostic C3 (Beckman).

**Mannan-Binding Lectin**
MBL was measured by sandwich ELISA using mouse monoclonal antibody (clone HYB 131-01) from Santa Cruz Biotechnology Inc (Santa Cruz, CA).

**Pneumococcal serology**
An in-house ELISA, meeting WHO standard, was used to quantify IgG antibody concentrations for pneumococcal serotypes 6B, 19F and 23F [199].

**IgA antibodies against tissue transglutaminase**
IgA antibodies against tissue transglutaminase (tTG) were analysed using Thermo Fisher Scientific Phadia 250 (Uppsala, Sweden). The method includes an IgA-screening (EliA™ Celikey® IgA on Phadia 250, Thermo Fisher Scientific, Uppsala, Sweden) [202].

**Statistics**
For all statistical analyses SPSS 20, 22 or 24 was used (SPSS, Inc. Chicago, IL, USA). In study II R survival package was also used. GraphPad Prism version 7.0 (GraphPad Software, San Diego, CA, USA) was used for making the graphs in study IV.

**Data characteristics**
Variables have different characteristics and can be classified into categorical (qualitative) and numerical (quantitative) data. Categorical variables can be further divided into nominal and ordinal data. In both these data types categories are mutually exclusive but where nominal data are unordered (e.g. blood group), ordinal data can be ordered (e.g. APGAR). Numerical variables can be separated into discrete data/counts where only integer values can be used (e.g. number of pregnancies) and continuous data which can take on any value in a range (e.g. weight). Outcomes in studies I-III in this thesis are all nominal (binary) (disease or not disease, death or not death) whereas some of the variables used in the regression models are ordinal (e.g. education level). Variables in study IV on the other hand are mainly continuous. Besides being aware of the type of data, statistical testing requires knowledge as to whether data are normally distributed or not. If they are not, non-parametric tests have to be used, however, if the material is large parametric tests can still be considered. In general, non-parametric tests are less powerful and thereby less likely to declare results significant. Study IV included
both normally and non-normally distributed data wherefore different statistical tests were used (see further down).

**Logistic regression**

Logistic regressions are commonly used in case-control studies. In contrast to the linear regression, where the dependent variable is continuous, the dependent variable in a logistic regression is binary (CD or not CD in study I) and cannot be fitted into a linear equation straight away but instead has to be logit transformed. The independent variable, on the other hand, does not have to be binary but can be of any type. The logistic regression can be extended to a multiple logistic regression equation where several independent variables with different characteristics can be taken into consideration. In this way potentially confounding factors can be accounted and controlled for, a possibility used in study I. Logistic regressions provide valid estimates of odds ratios associated with risk factors. Providing that the frequency of the disease (outcome) is relatively low (<10%) the odds ratios can also be interpreted as relative risks (i.e. the OR in study I can be considered equivalent to the relative risk). In study I we used a conditional logistic regression. With a conditional approach each case is individually compared with his or her specific controls, this maintains the matching and reduces the impact of the matching variables.

**Cox proportional hazards regression model**

The Cox proportional hazards regression model is a type of survival analysis. Survival analyses were originally designed to estimate risks for death but are now used to assess risks for all kind of events (like IPD in study III) and are therefore also termed time-to-event analyses. A hazard ratio (HR) is the risk of an event in one group (e.g. CD-group) divided by the risk of the event in another group (e.g. reference group). In contrast to comparing proportions in a Chi-square test, survival analyses take time and censored individuals into account. The HR gives an estimate of the overall difference between the survival (time-to-event) curves, i.e. the effect size. In the Cox proportional hazards regression model the dependent variable is the instantaneous hazard of the event. The term proportional hazards reflects the assumption that the ratio of one hazard to another remains constant over the study period. Other assumptions for survival analyses are that observations are independent and that censored participants have the same survival prospects as the non-censored participants. The proportional hazards assumption can be verified by using log-minus-log curves. Like the logistic regression the Cox regression can be extended to a multivariate model. Several covariates can be fitted into the equation and by doing so it is possible to control for potential confounders. Multivariate
Cox regression models have been used both in study II and study III. In study III we used an internally stratified model similar to the conditional logistic regression described above. Each case was compared with his or her matched controls before a summarized risk estimate (HR) was calculated.

**Independent sample t-test, Mann-Whitney U-test and Wilcoxon signed rank test**

Since some of the data in study IV were not normally distributed both parametric and non-parametric tests were used to compare differences between the study groups. The independent sample t-test is a parametric test that measures whether the mean of a continuous variable (statistically) significantly differs between two independent groups. The Mann-Whitney U test is a non-parametric test that is based on ranking of the observations and measures differences in distribution between the study groups. To compare paired samples with non-parametric characteristics (i.e. to statistically confirm that pneumococcal stimulation caused an increase in complement activity) Wilcoxon Signed rank test was used. The test ranks all changes (independent of sign) from smallest to largest and then sum the ranks for increases and decreases separately before comparing them.

**Confidence intervals and p-values**

In the studies included in this thesis odds ratios, hazard ratios and graphs are presented with 95% confidence intervals but in addition p-values are often reported. The confidence interval is a very descriptive measurement of statistical precision. If the study is unbiased there is a 95% probability that the true effect size/population value is included in the confidence interval. The confidence interval is non-committting but just describes the precision of the findings and puts the emphasis on the size of the effect. The statistical precision increases and hence the confidence interval narrows with increasing study power. P-values on the other hand describe the risk of the result being a chance finding (a false-positive conclusion). P-values can of course supplement the confidence intervals but only reflect *statistical* significance and not the clinical meaningfulness. Larger study material usually leads to smaller p-values.

In this thesis 95% CIs not including 1 (studies I-III) and P-values <0.05 were considered statistically significant.
Ethical considerations

All the studies included in this thesis were approved by Ethics Review Boards in Stockholm or Linköping. (Ref: 2006/633–31/4 and 2016/366-31).

Studies I-III are all large register-based studies and the Ethics Review Board did not request that patients should be contacted to consent. To minimize risk of identification in these type of large register-based studies, the Swedish National Board of Health and Welfare assists in the linkage of different registers and returns the datasets to the researches completely anonymised with each study participant represented by a serial number.

Study IV involved invasive procedures (blood samples) and questionnaires. To avoid unnecessary testing we aimed at including CD patients with a planned hospital appointment in near-time and control patients that were due to have a vein puncture independent of study participation. In case of abnormalities in the test results (needed to be assessed), the participant’s guardians were contacted by their attending physician or me. Before participating in the study, an informed consent was signed by the child and both its guardians. Participation was completely voluntary and did not affect or interfere with the medical care.
Results

Study I
Respiratory syncytial virus and later celiac disease

Respiratory syncytial virus
Out of the 3,835 children (<2 years) with CD, 36 (0.9%) had a prior hospital record of RSV infection. The corresponding number in controls was 117/19,102 (0.6%). This generated an OR of 1.46 (95%CI=1.03-2.07) for having attended hospital for an RSV infection prior to CD diagnosis. Adjustments for perinatal factors did not particularly influence the risk estimate (aOR 1.51; 95% CI=1.05-2.19). The highest risk of CD was seen within one year after RSV diagnosis (OR 1.63; 95%CI=1.08-2.46), beyond that no statistically significant risk increase was noted. The increased risk was predominantly seen in girls (OR 1.71) but P for interaction was not statistically significant. The same applied for age and calendar period at diagnosis. We also examined the association between RSV and biopsies showing inflammation as well as normal mucosa but positive CD serology (potential CD). Some 2.1% of the 97 children with inflammation had a prior hospital record of RSV compared to 1.2% of the matched controls. This corresponded to an OR of 1.60, however, this finding failed to attain statistical significance (95%CI=0.34-7.55). In children with potential CD the OR was even higher (4.19; 95%CI=2.00-8.77). Here a prior RSV diagnosis was found in 10/280 (3.6%) cases compared to 9/1,397 (0.6%) controls.

Unspecified viral bronchiolitis
About 3.4% (132/3,835) of the children with CD had a prior hospital record of any viral bronchiolitis or bronchitis as compared to 2.0% of the matched controls (390/19,102). This corresponded to an OR of 1.60 (95%CI=1.33-1.92). The risk of subsequent CD diagnosis was, also in this group, highest within the first year after the infection (OR 1.69; 95%CI=1.36-2.09). Subgroup analyses (by sex, age or calendar period at CD diagnosis) did not reveal any statistically significant differences.

Study II
Celiac disease and survival in sepsis
From the original cohort a total of 5,470 individuals who had received a sepsis diagnosis at any point in life were identified (4.9% of the CD patients and 2.8% of the controls). The majority were diagnosed with sepsis at older age.
All-cause mortality
During follow-up 841 individuals with CD and 2,314 controls died. This corresponded to an adjusted hazard ratio (aHR) of 1.19 (95%CI=1.09-1.29) for death from any cause in CD patients compared to controls. Having received a sepsis diagnosis early in the study period was associated with the highest risk of death (aHR 1.76; 95%CI=1.38-2.26). Likewise children had a higher risk (HR 1.67) but this finding was not statistically significant (95%CI=0.67-3.91). The risk of death was slightly lower when we restricted our data set to only include study participants who had received a sepsis diagnosis after CD diagnosis or study entry (controls). Calculations resulted in an aHR of 1.12 (95%CI=1.02-1.23).

Short-term mortality (within 28 days after sepsis diagnosis) did not seem to be influenced by CD (aHR 0.98; 95%CI=0.80-1.19).

Mortality from sepsis
A total number of 147 study participants had sepsis registered as their cause of death (CD: 33/1,432 (2.3%) and controls: 114/4,038 (2.8%)). The difference was small and not statistically significant (aHR 1.10; 95%CI=0.72-1.69). Restricting the study population to individuals who received a sepsis diagnosis after CD diagnosis (or study entry) resulted in an even lower risk estimate which remained statistically non-significant (aHR 1.04; 95%CI=0.67-1.62). Most of the study participants had “unspecified sepsis” as a diagnosis why all HRs for death according to bacterial origin were statistically insignificant with wide CIs due to few number of events.

Study III
Celiac disease and invasive pneumococcal disease
Out of the 173,269 study participants we found a record of IPD in 207. The absolute number was quite low since we only had regional data on IPD for the first part of the study period. However, the proportion of IPD was slightly higher among CD patients than controls (0.15% versus 0.11%). This corresponded to an HR of 1.46 (95%CI=1.05-2.03), with an attributable fraction of 32%. Adjustments for socioeconomic factors did not influence the risk estimate but when we added comorbidities (T1DM, asthma, chronic liver disease, end-stage renal disease and ischemic heart disease) into the regression model the HR diminished slightly and just failed to reach statistical significance (aHR 1.40; 95% CI=0.99-1.97) Restricting the study cohort to individuals diagnosed with CD after 1994, when the regional IPD databases were complete, influenced risk estimates marginally (aHR 1.58; 95%CI=1.05-2.37). The increased risk of IPD
was mainly seen in women (HR 1.90; 95%CI=1.30-2.76) whereas there were no statistically significant differences between age groups. Subanalyses also included stratification for follow-up time and the highest risk of IPD was seen between 1-4.99 years after CD diagnosis (HR 2.25; 95%CI=1.08-4.70).

**Study IV**

*Celiac disease and complement response to pneumococci*

All together 25 individuals with CD and 23 controls were included. However, during the processes 8 study participants had to be excluded from analyses (two individuals with CD and two controls had to be excluded due to bacterial contamination, one patient and one control due to inadequate bacterial concentration, one control did for unknown reason not activate complement (despite C3 level within normal range) and in addition plasma from one control was missing). The final study population therefore consisted of 22 individuals with CD and 18 controls. The median age was 15.4 years in the study group and 11.5 years in the control group. The majority of the study participants were female (CD: 16/22 and controls: 10/18). None of the participants had received pneumococcal immunization. One of the CD patients had a possible asthma diagnosis otherwise no co-morbidity or medication known to influence complement activity was reported or found when reviewing medical records. Likewise, no participant had a known affection of the spleen. One CD patient claimed to have had more than one pneumonia but this could not be verified. No recent pneumonias (last month) or previous meningitis were reported. All but one CD patient received their diagnoses after small intestinal biopsy with findings consistent with CD, the remaining patient was diagnosed without biopsy as proposed by the most recent ESPGHAN guidelines (Table 2, alternative 2) [86]. The mean duration of CD was 7.6 years (1.3-14.9). Of the control individuals, 15/18 attended the hospital for undergoing an MRI.

**Complement activity (main outcome)**

*C3a, C3 and C3a/C3*

C3a levels are presented in Fig 7. Pneumococcal stimulation did cause a statistically significant increase in C3a activity (p<0.001) but no difference in response could be demonstrated between the CD group and the control group (p=0.497). After pneumococcal stimulation C3a increased on average 4.6 times in both the CD and the
control group (median: 4.4 times and range: 1.6-11.4 in the CD group compared to a median of 4.7 times and range 2.3-6.7 in the control group). Since C3a is dependent on C3 we also analysed baseline C3. No difference between the study groups could be found (p=0.124) (Fig 7). Also, the C3a/C3 ratio was similar in both groups (Fig 7). Three study participants had a C3 concentration below the reference (0.67-1.29 g/L). Excluding these from statistical analyses did not change the results.

![Graphs of C3a, C3, C3a/C3, and sC5b-9](image)

**Fig 7.** Distribution of C3a, C3, C3a/C3 and sC5b-9 in individuals with CD and controls before and after pneumococcal stimulation.
Bars representing mean and 95%CI, (-) non-stimulated samples, (+) stimulates samples.

**sC5b-9**

The distribution of the sC5b-9 concentrations are presented in Fig 7. Pneumococcal stimulation did cause a statistically significant increase in sC5b-9 (p<0.001) but the response did not at all differ between the groups (p=0.724). In the CD group the average increase of sC5b-9 was 22 times (median: 19; range: 8-47), equal to the also 22-fold increase seen in the control group (median: 20; range: 5-39).
Mannan-binding lectin
The MBL levels did not statistically significantly differ between the study groups (p=0.508). The median concentration in the CD group was 482 kU/L (7 to >2000 kU/L) and 493 kU/L (102 to >2000 kU/L) in the control group. One CD patient had an MBL value below the reference level (7 kU/L (ref >40 kU/L)). Removing this study participant from statistical analyses did not have any impact on the results.

Pneumococcal serology
The majority of the study participants showed signs of previous pneumococcal infection (i.e. IgG antibody levels >1 mg/L). In the CD group the median IgG antibody concentration was 3.10 mg/L (0.10 to 38.00 mg/L) for pneumococcal sero-type 19F, 2.00 mg/L (0.22 to >26.00 mg/L) for serotype 23F and 5.25 mg/L (0.29 to >50 mg/L) for serotype 6B. The corresponding numbers in the control group were 4.25 mg/L (0.53 to 23 mg/L), 1.5 mg/L (0.34 to >26 mg/L) and 3.50 mg/L (0.22 to >50 mg/L) respectively. Statistical testing showed no differences between the study groups (p=0.535, p=0.431 and p=0.460 respectively).

IgA antibodies against tissue transglutaminase
None of the individuals with CD had an IgA deficiency but two of them had discretely elevated tTG (8 and 13 kIE/L respectively (ref <7 kIE/L)). Excluding these two study participants from statistical analyses did not influence the results.
Discussion

Few studies are perfect. Similar to all research, the results presented in this thesis generate further questions. It is always essential to reflect on effect sizes, clinical importance, choice of methods as well as potential flaws. The aim of this section is to review the results and their implications but also strengths, limitations and possible pitfalls.

Findings and implications

Even though the studies in this thesis are not ecological and individual data are known, it is still of importance to be aware of the logical fallacy that can occur if one deducts data at individual level from data at group level why we have refrained from doing so. However, some of the findings in this thesis can support changes in clinical guidelines. They also encourage further research.

Study I

Respiratory syncytial virus and later celiac disease

In the first study we found that children with CD had visited hospitals due to RSV infection (or any viral bronchiolitis) to a greater extent than their matched controls. The excess risk was rather moderate (OR 1.46) and while it is hard to evaluate the significance of this specific finding, the result supports previous research that indicates an association between early-life infections and pediatric CD [44, 58, 66]. Almost all children have acquired an RSV infection before two years of age and the majority does not need hospital care [168]. Most likely the children with RSV (or unspecified viral bronchiolitis) in our study suffered from a more severe disease than the average patient which generates the assumption that their infection also could have had a greater impact on intestinal barrier and microbiota as well as the general condition which potentially could have increased the risk of developing CD.

Several factors seem to predispose to a more severe RSV disease [170, 203]. A few of them are also associated with CD [51, 76] and we were able to adjust for intra-uterine growth retardation and caesarean section in our regression model and thereby avoided confounding from these variables. We also adjusted for maternal smoking which is associated with hospitalization for RSV [168, 170, 203]. Unfortunately we did not have data on breast-feeding but as described earlier in this thesis there is not enough evidence to conclude that breast-feeding protects against CD [43]. However, we cannot completely rule out that breast-feeding acted as a confounder in the study. Another possible linkage between RSV and CD is vitamin D. Low levels of vitamin D can be seen in patients with CD [204] and might also
predispose to respiratory tract infections [205, 206]. However, in Sweden all children below 2 years receive vitamin D supplement (free of charge) why vitamin D deficiency is less likely to explain the association between RSV and CD.

Interestingly we also found a 4-fold increase in hospital visits for RSV in children with potential CD. This finding might confirm previous observations that infections can generate transient elevations of celiac autoantibodies [96].

A preventive drug containing monoclonal antibodies targeting RSV (palivizumab) is available on the market. Presently this is given to premature infants and young children with severe cardiopulmonary disorders to protect them from infection [168]. The rather modest OR found in our study cannot support an extension of the risk groups receiving these expensive prophylactic injections.

Study II
Celiac disease and survival in sepsis

Since individuals with CD seem to have an increased risk of sepsis we, in this study, examined the outcome for CD patients compared to controls. We confirmed findings from previous studies that individuals with CD have an excess mortality compared to the general population [163, 164] but did not find any indication that CD influences the survival in sepsis. The 19% risk increase for overall death that we noted was due to an excess mortality beyond 28 days after sepsis and therefore seemed very unlikely to be associated with sepsis at all. The aHR for sepsis-specific mortality was only 1.10 and the estimate was not statistically significant.

The study was large and included data on almost 5,500 individuals with sepsis and 147 cases of death from sepsis. Despite that we cannot completely rule out that the study was slightly underpowered and that we have made a type II error. The number of patients with a specified bacterial origin was unfortunately too low to obtain valid risk estimates of death according to bacterial origin. Considering the increased prevalence of hyposplenism in CD it would have been interesting to have been able to specifically assess the risk of death in sepsis caused by encapsulated bacteria. However, while an increasing number of underlying diseases in some studies have been linked to an excess mortality in sepsis, research is not unanimous [173, 207]. Even though a previous study has found an increased mortality in sepsis [167] it is likely that the importance of rapid recognition and treatment overshadows the possible minimal impact of a concomitant CD diagnosis and it is doubtful whether an even larger study would contribute with clinically important
information. The impact of T1DM is probably of greater importance [173] and was adjusted for in the statistical analyses.

**Study III**

**Celiac disease and invasive pneumococcal disease**

In this third study we found a 46% risk increase of IPD in patients with CD compared to the general population. Since many diseases associated with CD [81, 133, 143] are known to increase the risk of IPD [195, 196] we adjusted for several comorbidities (T1DM, chronic liver disease, end-stage renal disease, asthma and ischemic heart disease) in our regression model and when doing so, the risk estimate only changed marginally but just failed to reach statistical significance (aHR 1.40; 95%CI=0.99-1.97). Nonetheless, our results showed a trend towards an increased risk of IPD and they support previous findings. Earlier studies by Thomas et al [153] and Ludvigsson et al [12] have found somewhat higher risk estimates (HRs 1.61-2.5). This probably reflects differences in study design. While the English study included pneumonia among their outcomes, the Swedish study looked specifically at sepsis and was limited to inpatient CD data. Also, less adjustments for comorbidities were performed. However, a recent review and meta-analysis, adding results from our study to these previous papers, supports the conclusion that individuals with CD are at-risk of IPD. Given the severity of this illness the CD population should be considered for pneumococcal vaccination [208].

The major strength of study III is the high validity. Both the CD and IPD data are very reliable (elaborated in a forthcoming section). However, for the first part of the study period we only had access to regional IPD data resulting in a low actual number of events but not influencing our risk estimates. The most commonly proposed mechanism for the increased risk of infections seen in CD, is the (previously described) comparatively high frequency of hyposplenism [145]. Splenic function is not routinely evaluated in patients with CD in Sweden so we were not able to take this into account in the study. Similarly we did not have access to vaccine status. As mentioned earlier pneumococcal vaccine has, in Sweden, been recommended to elderly since 1994 but not routinely to individuals with CD [178] and therefore vaccinations are unlikely to have influenced our HRs. In addition to elderly certain risk groups are recommended vaccination. The number of risk groups included in this recommendation has increased over the years and this study (with a 32% attributable fraction from CD) might support an extension to the celiac population.
Study IV  
Celiac disease and complement response to pneumococci

Study IV examined a possible mechanism contributing to the increased risk of IPD in CD found in study III. To our knowledge this was the first study investigating complement activation in individuals with CD compared to controls, in response to a pathogen. Pneumococcal stimulation caused an increase in complement activity in all individuals but the response was strikingly similar between CD patients and controls. Comparing non-stimulated and stimulated samples separately did not reveal any differences between groups either. Interestingly, as a secondary finding, we noticed that the majority of the study participants showed signs of immunity to the three investigated pneumococcal serotypes despite being unvaccinated. This probably confirms that most individuals are exposed to and infected by pneumococci during childhood [180]. The levels of the IgG antibodies were equally distributed between CD patients and controls so there were no indications of differences in antibody response to pneumococci. Research assessing antibody response to pneumococcal vaccine in CD is very limited but indicate appropriate postvaccination levels [209]. It is obvious that further research about mechanisms underlying the increased susceptibility to IPD is of interest.

Methodological considerations

Strengths and limitations in general

Papers I-III in this thesis were based on a very large cohort of individuals with CD and matched controls [84, 163]. The size of the cohort and the population-based setting are of course some of the major strengths of all these studies. The material is unique (with an extremely low risk of misclassification - see below) and gives opportunity to extensive epidemiological and register-based studies. Due to linkage to population-based registers the cohort also includes demographical data and information on education and occupation. Likewise, data on comorbidities are available and for individuals born 1973 or later we also have access to perinatal information. We do however lack data on smoking and alcohol consumption. While socioeconomic status can be used as a proxy for these variables, smoking does not have a clear association with CD [210] and is therefore most likely not a confounder in our studies. Research on alcohol and CD is limited. An Italian study from 2005 showed that alcohol was more frequently used in individuals with CD adhering to a GFD but did not reveal misuse [211]. Another limit is that the cohort data do not contain information on dietary adherence. However, given the large number of study participants, it is not likely that differences in dietary compliance have affected the risk estimates substantially.
Study IV included children and adolescents with and without CD. A major strength of this study was the access to patient charts and vaccine data for all participants. The pilot experiments and the study analyses were developed and performed by coworkers with long laboratory experience. Since all the analyses were performed by the same persons it most likely increased the reproducibility. One of the potential pitfalls is the difficulty that comes with finding and recruiting suitable control patients (see further down) but also that a part of the laboratory method had to be developed. However, the latter is of course a general problem when a new hypothesis is tested but is usually outweighed by the importance of trying novel ideas.

**Study designs**

Experimental studies are not always possible or suitable to perform. In general, observational studies are used to demonstrate associations and to generate hypotheses. Usually it is possible to perform large-scale studies, one of the major benefits of observational studies, enabling investigations of rare exposures and multiple outcomes. Prospective *cohort studies* are often considered to be the most prestigious epidemiological method. However, there are disadvantages, for example they are often expensive and there is a considerable risk of multiple drop-outs. The cohort used in studies I-III in this thesis is in some way both historical and prospective. Small intestinal biopsy data and diagnoses (including IPD) have been recorded prospectively but collected retrospectively. This has minimized the risk of study participants being lost to follow-up but has the disadvantage that some of the registers used are not specifically designed to answer the research question and also lack some information that might have been useful. However, linkage of the collected biopsy reports to several population-based registers has given access to an, in comparison, tremendous amount of information. In a worldwide perspective this cohort with 29,096 individuals with CD, 13,306 with inflammation, 3,719 with potential CD and 228,632 matched controls is quite unique [84, 163]. Studies II and III used this celiac cohort and linked the data to registers containing prospectively recorded information on sepsis, death and IPD. Although small data irregularities do exist in these registers it is unlikely that the proportion of missing cases and irregularities differ between cases and controls why risk estimates should not have been affected. Study I used children <2 years of age from the collected cohort. By doing so we created a *case-control study* assessing RSV as a potential risk factor for early-onset CD. Case-control studies are useful and suitable for assessing risk factors associated with an outcome (particularly if the outcome is rare). In a retrospective case-control study it is not possible to evaluate study participants continuously so in study I we only had access to hospital data on RSV. However, considering that almost all
children have undergone an RSV infection before 2 years of age [168] it was the severe cases that were of interest.

Study IV was a mechanistic study designed to examine whether complement response to pneumococci differed between young individuals with and without CD. Whereas the methods for reactivation of complement [197] and complement analyses [198, 200] had been previously used, the bacterial load and incubation time had to be established prior to analyses. To verify that our methodological concept worked we used statistical tests to confirm that the increase in complement activity after pneumococcal stimulation was significant (p<0.001). Likewise, all samples could not be analysed at the same time. To avoid this from biasing the study we analysed a few CD patients and a few controls each time. Also differences in bacterial load is inevitable. The concentrations were examined after each occasion and statistical testing showed no overall difference between the CD group and the controls (p=0.757).

Validity
Validity describes how close the interpretation of the results is in relation to the truth.

Internal validity
Internal validity describes to what degree the results of the study are correct in regard to the study sample. It depends on the study design, sample collection and analyses. Accordingly the internal validity is influenced by biases (systematic errors) but also by random variation (chance). The risks of different kind of biases and chance findings in the studies included in this thesis are elaborated below.

Selection bias
Selection bias (not mutually exclusive from confounding) can occur when comparisons are made between two groups that differ also in ways other than the outcome intended to examine. To avoid selection bias a matching procedure was performed when creating a control group from the general population to the CD cohort used in studies I-III. Matching criteria included age, sex, calendar period (at CD diagnosis/study entry) and county of residence [163]. However, since CD is associated with diabetes [7], the prevalence of T1DM was higher in the CD group compared to controls. This difference was accounted for in the papers and adjusted for in the statistical analyses to avoid diabetes bias and confounding. Likewise, in study II we identified all study participants with a record of sepsis. This “removed” the matching and therefore matching variables were included in the regression
model to avoid bias and confounding from age and sex. In study IV cases and controls were recruited from the hospital. To avoid “diabetes or other autoimmune bias” we excluded individuals, both CD patients and controls, with additional autoimmune diseases. It is possible that study participants living closer to the hospital were more prone to agree to participation. Since we were not examining an environmental exposure this should not have biased the results. Selecting controls is always a difficult process, particularly when it comes to children. Almost all of the controls were at the hospital for having an MRI (often due to headache). The mean age differed between the study groups, however, as previously described, complement activity in children (of this age category) is not associated with age or sex why this not had the potential to cause selection bias. This was also confirmed by analysing baseline levels of C3 in non-stimulated plasma. In general the study participants were healthy and we were not able to identify any significant differences of importance for this particular study between the CD group and the control group wherefore selection bias is unlikely to have influenced our results (which nor revealed any differences).

The selection of study samples (both case and control groups) of course also ultimately affects the external validity.

Misclassification

Misclassification applies to both exposures and outcomes and could most probably also be classified as a kind of construct validity (have we measured what we aimed to measure?).

Celiac disease

The risk of misclassification of CD in this thesis is very low. In studies I-III, CD was identified through biopsy records and as previously mentioned, small intestinal biopsy was considered the gold standard for diagnosing CD during the study periods. This was confirmed in a web survey where 184 gastroenterologists and 68 pediatricians participated (65% and 97% responding frequency respectively) [84]. Likewise, as described earlier, a validation study of the material was performed where a subset of the material was graded by blinded pathologists from 22 different pathology departments. About 90% of the biopsies with VA were correctly classified as defined by consensus of the Swedish National Steering Group of Small Intestinal Pathology [84]. In addition, to further validate the material and reduce the risk of misclassification, a subset of biopsy reports and a smaller number of patient records were manually reviewed [84]. A more recent Swedish study (published 2016, several years after our material was collected) did, however, show that
there presently might be both an inter- and an intra-variability when small intestinal biopsies are being classified and that agreement rates differ between pathology departments, emphasizing the need for access to clinical information and continuous education [212]. Why agreement rates were lower in this validation than in the one done on our material is hard to conclude. It is possible that the increasing demands on the health care system has given less opportunities to education?

In study IV all study participants with CD were identified through computerized patient records and all records were reviewed manually to confirm CD diagnoses before sending invitations to the study. Out of the 22 final study participants with CD, 21 received their diagnoses after small intestinal biopsy.

**Respiratory syncytial virus**

Data on RSV and unspecified viral bronchiolitis were retrieved from the Swedish Patient Register. An overall validation of the Patient Register has been performed and in general the accuracy is good and misclassification low. PPVs usually reach 85-90% [190]. A specific validation of RSV diagnoses has not been done why we cannot rule out that some records in our study were misclassified. Misclassifications should however, affect cases and controls equally and thereby ORs are unlikely to have been influenced. Also unspecified viral bronchiolitis and bronchitis were examined. Tests for RSV are not always performed why this group also can contain RSV cases.

**Sepsis**

Data on sepsis were retrieved from the Patient Register. The performance of the Patient Register is described above [190]. In addition, a validation study determining misclassification rates of sepsis diagnoses at intensive care units (ICUs) in Sweden has been performed. The specificity of community-acquired sepsis diagnosis records was 99.4% (95%CI=99.2-99.6) whereas the sensitivity was lower (51.1%; 95%CI=41.0-61.2) [213]. Most likely the majority of sepsis cases in our study are correct but the actual number might be underestimated, maybe reducing the study power somewhat.

**Invasive pneumococcal disease**

Also IPD diagnoses are very accurate. IPD events were retrieved from computerized databases from microbiological laboratories and the Swedish Public Health Agency. They are all based on positive cultures and do not rely on clinical evaluation resulting in a very low risk of misclassification.
Causes of death
The Swedish Cause of Death Register has been evaluated in studies. A 77% agreement rate between registered causes of death and discharge summaries were found. Agreement rates were higher in younger patients (e.g. 98% in individuals 15-44 years and 91% in individuals 45-64 years) and for certain diagnoses. Sepsis was not reported but the agreement rate for pneumonia reached 83% [214]. The previously mentioned study investigating diagnosis-dependent misclassification of infections in ICU patients, reported underestimated mortality rates for sepsis [213]. Most likely the accuracy of sepsis as a registered cause of death (used in study II) is quite high but the number of deaths could have been underestimated. This is unlikely to have influenced our risk estimates substantially but might have reduced the study power somewhat.

Recall bias
Basing data on the study participant’s recollection of exposures can result in recall bias. “Cases” are usually more prone than referents to report different exposures. In this thesis the designs of studies I-III remove the risk of recall bias. None of the data used are based on questionnaires or interviews. In study IV a questionnaire was used to assess previous pneumonia, meningitis and pneumococcal vaccination (none of them the primary outcome of the study). However, the information could be confirmed through review of patient charts and vaccine registers, minimizing risk of recall bias.

Surveillance bias
This type of bias can occur if one study group stand a better chance of having the outcome detected than the other. It is possible that surveillance bias could have influenced the results in study I. Children visiting hospitals for RSV infection could have been subject to more investigations and tests resulting in an earlier CD diagnosis. However, the mean duration between RSV diagnosis and CD diagnosis was 300 days, making it less likely that surveillance bias influenced our results substantially. Studies II and III assess severe infections and death as outcomes. These are not outcomes that are more likely to be detected due to the increased monitoring that comes with a CD diagnosis, ruling out surveillance bias. Study IV measures complement activity in individuals with and without CD. These are point estimates and there should be no risk of surveillance bias.
Confounding bias
A confounder is a variable that influences the risk of the outcome and also co-varies with the exposure. Confounding can be both positive and negative and can be controlled for in different ways. Still, controlling requires knowledge of which variables that potentially could act as confounders, unknown factors can always influence the result anyway. Different strategies for dealing with confounders include randomization, matching, stratification and mathematical modelling. Also in this thesis several methods for controlling for confounding have been used. The study cohort that studies I-III were based on used a matching procedure reducing the risk of confounding by age, sex, calendar period (for CD diagnosis/study entry) and county of residence. The influences of these variables were further reduced by applying the previously described conditional approach in the regression models used in studies I and III. In addition, a priori defined stratifying subanalyses were performed (including by follow-up time, sex, age and calendar period for CD diagnosis). Study II also included stratifications by age at sepsis diagnosis and bacterial origin. In all the three first studies multivariate analyses were performed in which we adjusted for potentially confounding comorbidities (see Table 4), socioeconomic status and being born in Nordic countries (in study I also perinatal factors). In study II the previously described matching variables were also included in the mathematical model since they were “removed” when we restricted the dataset to only include study participants with a sepsis diagnosis. While multivariate analysis is the only possible way to deal with many variables at one time is important to be aware that multiple adjustments might decrease the study power. This might have affected the result in study III where the lower limit of the CI decreased to just below one (0.99) after multiple adjustments. In study IV study participants with autoimmune diseases other than CD were excluded from participation to avoid potential confounding from for example T1DM and thyroiditis. We are not aware of any further confounders that could have influenced our results even though this doesn’t fully exclude that there might be some.

Chance, precision and power
The deviation of an observation from the true population value due to chance alone is termed *random variation*. As described in the statistics section chance can be approached by *estimations* (the range in which the true value is likely to be included i.e. presented with *confidence intervals*) and *hypothesis testing* (associated with *p-values*). A false-positive conclusion is referred to as a *type I error or α-error*. The risk of type I errors in studies I and III in this thesis cannot fully be ruled out. Confidence intervals did tend to approach one but those who included one and had associated p-values ≥0.05 were not claimed to be statistically significant. While
the studies, in comparison, are large it is possible that an even higher number of IPD events in study III would have resulted in a narrower confidence interval (increased statistical precision) which nor would have included one after adjustments. A false-negative conclusion is instead referred to as a type II error (β-error). The risk of a type II error depends on the statistical power of the study. Although study II was based on a large material it is not entirely excluded that the number of sepsis cases and deaths from sepsis was too low to demonstrate a difference in mortality risk (thus underpowered). However, most likely the finding reflects the “truth” and there is no clinically important difference in sepsis outcome between CD patients and controls. Calculating the sample size needed in study IV was very difficult. The lack of previous research was of course one of the rationales for performing the study and considering the extremely similar response in complement activity seen between the groups we do not believe that a larger study would be more informative. Even though it is not very likely, it is still not excluded that a type II error is responsible for the negative result.

**External validity**

The external validity reflects the generalizability. How well do the results apply to the general population or other settings? Studies I-III were all nationwide and population-based with a low risk of misclassification and therefore we consider them to have a high external validity in regard to the “general CD population”. Still, results always have to be interpreted in its context. For instance the findings do not apply to all age groups. Study I only included children <2 years whereas study III had somewhat limited access to data on children. The statistical analyses do however include stratifications facilitating inferences in regard to different groups of individuals. In study IV complement activation was measured in children and teenagers born 1999-2008. The individuals with CD most likely represented the average CD patient since no selection was performed (all pediatric patients with CD living in Southern Kalmar County attend the Pediatric Clinic at Kalmar County Hospital). The control patients on the other hand cannot be said to represent “any healthy individual” since they attended the Pediatric Clinic for a reason. However, the aim of the study was not to estimate complement levels in healthy children but to look for a difference between children with and without celiac disease and therefore inferences in regard to the general population has not been done.

**Causality**

Potential causal risk factors for CD and what complications CD might cause are discussed in the background of this thesis. In general causality is hard to determine since it is not equivalent to association. Usually, as is the case in CD, many factors
interplay to cause disease (“the web of causation”). In 1965, the British professor Sir Austin Bradford Hill proposed some aspects that, in his opinion, should be considered before establishing causation. This set of features has often been referred to as the “Hill criteria” and are still widely used [215].

Table 6. “The Hill criteria”.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Feature/comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strength of the association</td>
<td>Large risk estimates</td>
</tr>
</tbody>
</table>
| Consistency               | Repeated observations (under different circum-
|                           | stances)                                     |
| Specificity               | One factor leading to one outcome            |
| Temporality               | Cause preceding effect                       |
| Biological gradient       | Dose-response                                |
| Plausibility              | Is causation biologically plausible?         |
| Coherence                 | Causation is not in conflict with current
|                           | knowledge                                    |
| Experiment                | Does removal of the exposure prevent disease?|
| Analogy                   | Similarities to other cause-effect relation-
|                           | ships                                       |

The studies in this thesis have demonstrated some associations. The risk estimates are, however, moderate. Still, the increased risk of IPD seen in study III is consistent with previous findings and the increased frequency of hospital visits for prior RSV infection is in analogy with research supporting involvement of viral infections in CD pathogenesis. Since study I assessed more severe RSV infections there might be a biological gradient (dose-response effect) however this is nothing we have investigated. Likewise, all children are sooner or later affected by RSV and general prevention is presently not considered and accordingly its effects difficult to test. In study I we did find a temporal relationship, RSV was preceding CD and the highest risk of CD was seen within the first year after the infection. Our studies are not very specific but coherent with current knowledge and plausible. However, claiming causality would be premature and presumptuous.
Concluding remarks and future research

In summary this thesis found associations between CD and infections. We demonstrated a moderately increased risk of early-onset CD after an RSV infection needing hospital care (OR 1.46; 95%CI=1.03-2.07). The results were consistent with previous research indicating involvement of viral infections in CD pathogenesis [44, 66]. We also found an increased risk of IPD among individuals with CD (HR 1.46; 95%CI=1.05-2.03). However, after multiple adjustments the risk estimate just failed to attain statistical significance (aHR 1.40; 95%CI=0.99-1.97). Still, the finding did confirm earlier research [12, 153].

The susceptibility to infections (and in particular infections with encapsulated bacteria) has for a rather long time been attributed to the hyposplenism found to a greater extent in individuals with CD compared to the general population [145]. Since other explanations for the elevated risk of pneumococcal infections are sparsely examined we, in study IV, investigated whether there was a difference in complement activation in response to pneumococci between young CD and non-CD individuals. No differences could be found and most likely other parts of the immune system than the complement system are of importance. Although CD seems associated with more severe infections [11] the findings in study II do not suggest that it affects the prognosis. This study showed no impact of CD on the outcome in sepsis (aHR 1.10; 95% CI=0.72-1.69).

Our findings have contributed to an increased understanding about CD pathogenesis and complications. They have also supported guidelines recommending pneumococcal vaccine to CD patients [71, 83] and hopefully this will now be taken under consideration also in Sweden.

For the future it would be of interest to reassess the frequency of hyposplenism among individuals with CD, both children and adults, at diagnosis and after introduction of a GFD. Recent data on the prevalence of hyposplenism are lacking and studies in a Swedish setting have not been performed. It could also be of interest to estimate the risk of infections with other encapsulated bacteria than pneumococci. Likewise, additional mechanisms for the increased susceptibility to infections should be investigated since it could generate further possibilities for prevention. Also minimal risk reductions are of importance when it comes to severe and potentially life-threatening diseases.
Sammanfattning på svenska

Den här avhandlingen handlar om celiaki och kopplingen till infektioner.


I det första delarbetet ville vi titta närmare på infektioners roll i sjukdomsutvecklingen. Vi undersökte huruvida barn som fått celiaki före två års ålder i högre utsträckning hade besökt sjukhus på grund av ett luftvägsvirus som heter respiratory syncytial virus (RSV). För att göra detta använde vi det insamlade datamaterialet. Från detta material identifierade vi bland annat 3 835 barn med celiaki och 19 102
barn utan celiaki (kontrollpatienter). Dessa data kopplades till Patientregistret och på så sätt fick vi fram hur många av barnen i respektive grupp som hade besökt sjukhus för en RSV-infektion. Bland barnen med celiaki hade 0.9% gjort det och bland ”kontrollbarnen” hade 0.6% gjort det. Detta motsvarade en ca 46% ökad risk för att ha haft en svår RSV-infektion innan celiakidiagnos vilket skulle kunna stödja teorierna om att virus kan vara inblandande i sjukdomsprocessen.

Eftersom forskningen har visat att individer med celiaki tycks löpa en ökad risk att drabbas av sepsis var vi i studie II intresserade av att titta på om en celiakidiagnos också påverkade dödligheten i detta mycket allvarliga tillstånd. I det tidigare nämnda celiakimaterialet hittade vi totalt 5 470 individer som någon gång vårdats för sepsis. Precis som tidigare studier visat så såg vi en något ökad ”generell” dödlighet bland celiakipatienterna (ca 19% högre risk) men däremot såg vi ingen ökad dödlighet i sepsis. Sannolikt spelar andra faktorer än celiaki mycket större roll för sepsisprognosen.


Celikpatienternas ökade risk för allvarliga infektioner som t.ex invasiva pneumokocks sjukdomar anses ofta bero på en dålig mjältfunktion. Mjälten är en viktig del i immunförsvaret särskilt mot s.k. kapsrade bakterier. Just pneumokocker är en av de vanligaste kapslide bakterierna. Studier har visat att individer med celiaki, särskilt de med svårbehandlad celiaki och de som har flera autoimmuna sjukdomar kan ha en mjälte som fungerar dåligt och som ibland är förminskad. Det finns dock inga studier som säkert visar att det just är den här eventuellt dåliga mjältfunktionen som ger den ökade infektionsrisken eller om det också finns andra förklaringar. I delarbete IV tittade vi därför på en alternativ mekanism. Vi samlade

Sammanfattningsvis har denna avhandling undersökt olika kopplingar mellan celiaki och infektioner. Förhoppningsvis kan resultaten bidra till ändrade riktlinjer samt ligga till grund för framtida forskning.
Acknowledgements

So many people deserve my deepest gratitude. For guiding, encouragement and support during my research education. For giving me happiness, friendship and love.

In particular I would like to thank:

Jonas Ludvigsson, my main supervisor. It is unlike me to make car-metaphors but (besides being the cyber version of The Flash) you have been everything needed in a car (except the brake): my engine, my accelerator, steering wheel and GPS but also, not least importantly, the best driving-company one could have. I have truly enjoyed every meeting we have had, changing literature recommendations, sharing clinical problems and laughter. Thank you for this journey, I hope the trip continues!

My co-supervisor Jonas Bonnedahl. Thank you for valuable support and guiding. Besides being an excellent navigator in research and hospital administration, your social skills enriches both our planned and spontaneous meetings. You have given me time for debriefing but also taught me one or two birds.

All my coauthors. Malin Inghammar, Arne Egesten, Gunnar Kahlmeter, Pontus Nauclér and Birgitta Henriques-Normark for contribution and valuable comments in study III. A very special thank you to coworkers in study IV: Kristina Nilsson Ekdahl, Hanna Woksepp, Per Nilsson, Kerstin Sandholm, Marcus Johansson and Charlotte Dahle, without you this project could not have been performed.

Louise Emilsson and Karl Mårild. For the greatest company and introduction to the “celiac community” during my first celiac disease conference (ICDS).

My colleagues at the Pediatric Department. For making it the best possible place to work and for taking care of my patients while I was finishing my thesis. A big thank you also to the nurses at the day-care clinic for your huge and never-ending effort to include study participants.

Region Kalmar County and former and present heads of the Pediatric Department for supporting and encouraging my research.

Hampus Nobin. For being a most helpful pathologist and friend. Thank you so much for digging the archives in search for “celiac pictures”.

Pernilla Frid. For an impressive creativity and an ability to produce visual associations to celiac disease resulting in the brilliant cover illustrations.
Maria and Thomas. Friends for better and worse. For giving joy and support in everything: life, clinical work and research. What can I say, without you I hadn’t even met the love of my life!

Ana and Daniel. For sharing the most enjoyable moments of life, for being undemanding and willing to listen and talk about anything (including research) no matter what.

My university friends (in particular “the hens”) who since the beginning of medical school have given me everything a friendship can desire, including loud singing and laughter.

All not previously mentioned Kalmar friends who make this a great place to live. Pia and Lisa also for sharing our first research steps and an abnormal amount of potato terrines.

My brothers Pontus and Peter and your fantastic families. For believing in me, for making me laugh but most of all for being there.

Dad and Elisabeth. For overwhelming belief. For talks. For love.

Mum. For your unconditional love. For support, advice and respect. For always being there.

Our fantastic sons Otto and Kalle whom I love more than it is possible to describe. For fulfilling my life.

My big love and husband Ivar. It is hard to express it better than Jocke Berg: “what would I be without your breath”? I don’t know how to thank you for making me feel so precious and for your endless support in life, clinical work and research. Words are not enough....

************************
The studies in this thesis were mainly funded by the Medical Research Council of Southeast Sweden.
References


111. Celiaki hos barn och ungdomar. barnlakarforeningen.se


155. Russo F, Chimienti G, Linsalata M, Clemente C, Orlando A, Riezzo G: The obestatin/ghrelin ratio and ghrelin genetics in adult celiac patients before and after a gluten-free diet, in irritable bowel syndrome patients


205. Belderbos ME, Houben ML, Wilbrink B, Lentjes E, Bloemen EM, Kimpen JL, Rovers M, Bont L: Cord blood vitamin D deficiency is


Publications in the series
Örebro Studies in Medicine


35. Söderqvist, Fredrik (2009). Health symptoms and potential effects on the blood-brain and blood-cerebrospinal fluid barriers associated with use of wireless telephones.
41. Gustafsson, Sanna Aila (2010). The importance of being thin – Perceived expectations from self and others and the effect on self-evaluation in girls with disordered eating.
42. Johansson, Bengt (2010). Long-term outcome research on PDR brachytherapy with focus on breast, base of tongue and lip cancer.
43. Tina, Elisabet (2010). Biological markers in breast cancer and acute leukaemia with focus on drug resistance.
46. de Leon, Alex (2010). Effects of Anesthesia on Esophageal Sphincters in Obese Patients.


52. Loiske, Karin (2011). Echocardiographic measurements of the heart. With focus on the right ventricle.


64. Nordin Olsson, Inger (2012). *Rational drug treatment in the elderly: "To treat or not to treat".*


67. Thuresson, Marie (2012). *The Initial Phase of an Acute Coronary Syndrome. Symptoms, patients’ response to symptoms and opportunity to reduce time to seek care and to increase ambulance use.*


75. Gustavsson, Anders (2012): *Therapy in Inflammatory Bowel Disease.*


83. Lönn, Johanna (2013): The role of periodontitis and hepatocyte growth factor in systemic inflammation.


96. Sundh, Josefin (2013): Quality of life, mortality and exacerbations in COPD.
98. Palmetun Ekbäck, Maria (2013): Hirsutism and Quality of Life with Aspects on Social Support, Anxiety and Depression.


109. Törös, Bianca (2014): Genome-based characterization of Neisseria meningitidis with focus on the emergent serogroup Y disease


120. Pelto-Piri, Veikko (2015): Ethical considerations in psychiatric inpatient care. The ethical landscape in everyday practice as described by staff.


139. Elwin Marie (2016): *Description and measurement of sensory symptoms in autism spectrum.*

140. Östlund Lagerström, Lina (2016): *”The gut matters” - an interdisciplinary approach to health and gut function in older adults.*


157. Olsson, Emma (2017): Promoting Health in Premature Infants – with special focus on skin-to-skin contact and development of valid pain assessment.


177. Christos Karefylakis (2018): Vitamin D and its role in obesity and other associated conditions.


