**EStudy Protocol: Celiac disease and risk of microscopic colitis.**

CHANGES FROM STUDY PROTOCOL

* when making major changes after the protocol has been finalized, please note here:

**Design: Cohort Study (1990-2016)**

**First author: David Bergman**

**(sole) Senior author:** Jonas F. Ludvigsson

**Additional co-authors:** Hamed Khalili, Bjorn Roelstraete (statistician).

**Background:**

Microscopic colitis (MC)is the most recently recognized inflammatory condition of the large intestine. The most prominent symptoms are frequent, non-bloody, watery diarrhea, weight loss and abdominal pain1. Some patients may also experience nocturnal diarrhea, urgency and fecal incontinence2. The gravity of symptoms may, however, differ substantially between patients.

The term microscopiccolitis is a unifying concept for a number of subtypes, with collagenous colitis (CC) and lymphocytic colitis (LC) being the most common. Recently, a third subtype, microscopic colitis incomplete has been introduced5. In a recent cohort study examining the incidence of microscopic colitis in Sweden we identified 13 844 patients with an incident diagnosis of microscopic colitis. Lymphocytic colitis (n = 9238) constituted 67% and collagenous colitis (n = 4606) 33% of microscopic colitis. The mean age at time of diagnosis of microscopic colitis was 60.2 years (58.6 for lymphocytic colitis, 63.3 for collagenous colitis). The lifetime risk of developing microscopic colitis was 0.87% in women (95% confidence interval, CI: 0.85-0.88) and 0.35% in men (95% CI: 0.34-0.36). From 2006, the overall incidence of microscopic colitis was approximately 10.5 cases per 100 000 person-years (95% CI: 9.8-11.3) with higher rates in women (72% of cases).

In the Epidemiology Strengthened by histoPathology Reports in Sweden (ESPRESSO) cohort, Dr. Jonas F. Ludvigsson has linked histopathology data available in 2.1 million unique individuals with clinical data in the Swedish National Healthcare Registers, and has matched these individuals with up to five reference individuals from the general population and first-degree relatives resulting in a total study population of 13 million individuals.22 In this cohort, there are 14,520 histopathology reports identifying patients with a first time diagnosis of microscopic colitis. Cases were recorded from 1990 to April 2017.22 Given that the sole diagnostic criteria for microscopic colitis are certain histopathologic features, the ESPRESSO cohort offers a unique opportunity to identify and study subjects with microscopic colitis, over long-term follow-up. Simultaneously, the ESPRESSO cohort also stores data on patients with a biopsy verified celiac disease. An ongoing study on the incidence of celiac disease in Sweden has identified 44,771 patients with celiac disease.

The relationship between celiac disease (CD) and microscopic colitis (MC) has long been a matter of interest and studies have found evidence implicating a similar pathogenic mechanism in the two diseases. However, currently published studies are inadequately powered and lack in generalizability.

**Objective:**

Our overall objective is to determine the association between celiac disease and microscopic colitis using high-qualitative, population-based registers. Specifically, we will assess the risk of microscopic colitis in patients with celiac disease compared to (A) matched population-based reference individuals and (B) siblings.

**Please note** that whenever the word “before” is used in relationship to “start of follow-up/biopsy date” it means “before and including date of biopsy / matching date / date of start of follow-up”).

**Design, Cohort and Study Population**

**Setting**: Sweden

**Source population:** The source population will include all men and women living in Sweden who has a duodenal/jejunal biopsy that demonstrated a diagnosis of celiac disease, defined by SNOMED codes:

1. Topography: Duodenum T64 (all) and jejunum T65, T65000 and T651
2. SNOMED code indicating celiac disease (M58 with subgroups) or the celiac disease diagnostic code D6218. (D6218, D62180, D62188, D6218X, D6218Y (celiac diagnosis), M58, M5800, M58000, M58001, M58005, M58006, M58007)
3. Note: We will not use free text searches to define our source population. Rather, we will exclusively use duodenal/jejunal biopsy and ICD codes.
4. **Years/Dates of biopsy**: 1990-2016.

**Ages**: Include both children and adults in the full source population.

**Sex (m/f):** Include both men and women

**Start of Follow-Up**:

1. Define start of follow-up as the date of index duodenal/jejunal biopsy. The index biopsy is the first duodenal/jejunal biopsy that demonstrates celiac disease.
2. Note, if a case has >1 biopsy, the first biopsy to demonstrate celiac disease is the “index biopsy”. It does not have to be the first duodenal/jejunal biopsy ever for that individual (one individual may have a normal duodenal/jejunal biopsy first, but follow-up starts on the date he/she has a biopsy with celiac disease).

**Control Group(s)**

1. *PRIMARY*: 5 matched reference individuals from the general population.
   1. *Matching:* Reference individuals are matched to celiac disease cases by age, sex, calendar year of biopsy and county of residence by the government agency Statistics Sweden.
   2. *Note about case status:* If a control (with no celiac disease at the date of matching) later undergoes a duodenal/jejunal biopsy (this may be their first duodenal/jejunal biopsy but could also be their second biopsy or later) that demonstrates celiac disease, they should be censored at the date of the biopsy with celiac disease and on that date be re-categorized as a case.

*SECONDARY:*

* 1. *Siblings.* Since only a minority of individuals undergoing biopsy have siblings, the sibling analyses will only be based on those index individuals who have a sibling. This analysis will reduce the influence of shared early environmental risk factors and shared genetics. *Status*: These have already been identified by Bjorn Roelstraete.
  2. *Normal mucosa*: These individuals will serve as a means to identify the true importance of villous atrophy compared to individuals who undergo similar investigations. *Status*: These have already been identified by Bjorn Roelstraete.

**EXCLUSIONS from Study cohort**

Notes regarding exclusions:

1. All exclusions will be made through SNOMED-codes.
2. The same exclusion criteria will be applied to both the exposed and reference individuals.
3. **IMPORTANT NOTE**: in ICD8, Swedish ICD-system uses “comma”, while in ICD10, interpunctuation (. Dot) is used. When coding (script/syntax), the comma is used (for instance we look for 269,0 in celiac disease studies) but the interpunctuation is not (we look for K900 in the data files).

***Timing of biopsy in relation to exclusion criteria:***

1. We will only exclude persons who meet the exclusion criteria BEFORE the index duodenal/jejunal biopsy.
2. Thus, if the first recorded ICD code for an exclusion criterion is met AFTER the index duodenal/jejunal biopsy, the person is still included (i.e. if a person had a biopsy showing celiac disease in 1992, and had ICD codes for microscopic colitis in 1996, he or she should be included in the source population).

**EXCLUSION CRITERIA** (to be applied to the source population and to reference individuals):

**1. Celiac disease or microscopic colitis recorded in ESPRESSO.**

1. **Exclude:** if any of the following SNOMED codes recorded **BEFORE** the index biopsy date: SNOMED code indicating celiac disease (M58 with subgroups) or the celiac disease diagnostic code D6218. (D6218, D62180, D62188, D6218X, D6218Y (celiac diagnosis), M58, M5800, M58000, M58001, M58005, M58006, M58007) or *M406* or *M4717 indicating microscopic colitis.*



**2. Exclusion from Study cohort**

Individual has formally emigrated, despite having a biopsy in Sweden. *Status*: These individuals have already been removed by Bjorn Roelstraete.

**3. Death before start of follow-up**

**OUTCOMES (see Table 1 for codes and definitions):**

Ascertainment of Outcomes: Outcomes will be defined by the appropriate SNOMED-codes.

Primary Outcome: Microscopic colitis

**Censoring / end of follow-up for primary analysis:**

* Primary outcome: Diagnosis of microscopic colitis
* Death
* Censor if alive on Dec 31, 2016.
* Censor at date of emigration

**Covariates (Table 2, below)**

* **Covariates**
* Conditioning/matching for the following variables is done in all analyses (except Kaplan-Meier curves): *age, sex, calendar year (year of biopsy), and county*.
* *Education*, there are different ways of categorizing education in Sweden, a common way is: compulsory school (0-9 years); upper secondary (10-12 years). College or university (≥13 years). For individuals diagnosed under age 18 (or perhaps for those whose follow-up did not exceed age 18), would use the furthest educational attainment of the two parents.
* Individuals with missing values will be assigned to category “missing”. We will not impute values.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table 2. Baseline (up until and including date of biopsy/matching date) Covariates** | | | | | |
|  | ICD8 (1969-1986) | ICD9 (1987-1996) | ICD10 (1997-) | Procedure Code | Notes |
| **Sex** | n/a | n/a | n/a |  |  |
| Male |  |  |  |  |  |
| Female |  |  |  |  |  |
| **Age at index biopsy** | n/a | n/a | n/a |  |  |
| Median/Mean Age |  |  |  |  |  |
| <18y |  |  |  |  |  |
| 18y -<40y |  |  |  |  |  |
| 40y - <60y |  |  |  |  |  |
| ≥60y |  |  |  |  |  |
| **Index Biopsy Year** | n/a | n/a | n/a |  |  |
| 1990 - 2000 |  |  |  |  |  |
| 2001 - 2010 |  |  |  |  |  |
| 2011 - 2017 |  |  |  |  |  |
| **Country of birth** | n/a | n/a | n/a |  |  |
| Nordic |  |  |  |  |  |
| Other |  |  |  |  |  |
| **Level of education** | n/a | n/a | n/a |  |  |
| ≤9 years |  |  |  |  |  |
| 10-12 years |  |  |  |  |  |
| >12 years |  |  |  |  |  |
| Missing |  |  |  |  |  |
| **Comorbidities #** |  |  |  |  |  |
| Diabetes (yes/no) | 250 | 250 | E10-E14  E14.1-E14.9 |  | Also include any of the following ATC codes:  A10A  A10AB-AE  A10BA-BX |
| Inflammatory Bowel Disease (IBD) | 563,10; 563,09; 563,99; 569,02; 569,03; 569,04  : 563.00 | 556.0; 556.1; 556.2; 556.3; 556.4; 556.5; 556.6; 556.8; 556.  555.0; 555.1; 555.2; 555.9 | K51.0; K51.2; K51.3; K51.4; K51.5; K51.8; K51.9  K50.0; K50.1; K50.8; K50.9 |  |  |
| Autoimmune thyroid disease | 242.00, 242.09, 244, 245.03 | 242A, 242X, 244X, 245C, 245W | E03.5, E03.9, E05.0, E05.5, E05.9, E06.3, E06.5. |  |  |
| Rheumatoid arthritis | 712,3  714,93 | 714 | M05; M06; M08; M12.3 |  |  |

**EXPOSURES and groups:**

* 1. Celiac disease patients compared to reference individuals

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 3. SnoMed Codes for Histology Exposure Groups** | | | |
| **Characteristics** | **SnoMed code** | **Comment** | **Motivation** |
| Celiac disease | (D6218, D62180, D62188, D6218X, D6218Y (celiac diagnosis), M58, M5800, M58000, M58001, M58005, M58006, M58007) |  |  |
| Normal mucosa | M00100; M00110 |  | Used in the sensitivity analysis |

**STATISTICAL ANALYSIS**

Overall Description:

We will include celiac disease patients matched to population reference individuals by age at index biopsy (within 1 year), sex, country and calendar year of biopsy. Study outcome is microscopic colitis, defined by SNOMED-code as lymphocytic colitis or collagenous colitis.

We will calculate the cumulative incidence and absolute rate differences for the outcome according to exposure groups. We will construct unadjusted- and multivariable-adjusted Cox proportional hazards regression models to estimate adjusted hazard ratios for the primary outcome (microscopic colitis). We will first compare persons with celiac disease vs. general population controls. Age will be the underlying time scale.

* + All analyses will include four Cox PH Regression models, stratified by age at index biopsy, sex, and calendar year of biopsy:
    - 1. Unadjusted
    - 2. Adjusted for age, sex, calendar year, county where biopsy was obtained
    - 3. Adjusted for #2 + education (3 groups + missing category), baseline clinical comorbidities (defined up to and including the index biopsy date)

**OUTPUT**

**Background Data**

* Plot incidence of primary and secondary outcomes per calendar year, among:
  + 1. Celiac disease patients vs. matched reference individuals
* Check that Proportional Hazards Assumption is fulfilled.

**Descriptive data**

* Flowchart for cohort construction
  + Number of individuals eligible for inclusion (overall)
  + Number of individuals excluded due to each of the exclusion criteria (overall)
  + Number of individuals included in the cohort (overall and by exposure group).
* Mean (SD) and median (range) follow-up in years for the primary and secondary outcomes
  + Overall
  + By exposure group (1. celiac disease vs. reference individuals;
* **SHELL TABLE A1 and A2 (at the end of this document):** 
  + Baseline characteristics (**Shell Tables A1 and A2,** below). Please record the proportion that are missing for education variables in each exposure group (so that we can report missingness in the publication). Please also include a separate missing category for education in the regression models.
  + Number of emigrations after index biopsy, according to calendar year
  + Number of microscopic colitis diagnosis during follow-up
  + Number of cancer diagnosis (primary and secondary outcomes) within 365 days of the start of follow-up

**Exposure Groups:**

* + 1. Primary comparison: Celiac disease patients vs. matched reference individuals

**Statistical Analyses**

* + Sibling analyses will be run using stratified Cox regression.
  + We will construct both unadjusted and multivariable-adjusted Cox regression models, using the covariates outlined in Table 2 (also shown in footnotes to **Shell Tables**, below).
    - Clinical covariates are modeled up to and including the index biopsy date (baseline)
  + All analyses will be stratified by age at index biopsy, sex and calendar year of biopsy

**PRIMARY ANALYSIS:**

* 1. Calculate cumulative incidence rates with 95% CIs at 1, 5, 10, 15, 20 (etc) years of follow-up, comparing (A) celiac disease vs. reference individuals for the primary outcome.
  + 1A. Calculate the absolute rate differences at each 5-year time-point of follow-up
* 2. Calculate unadjusted and multivariable-adjusted hazard ratios (with 95% CIs) for the study outcome, comparing (A) celiac disease vs. reference individuals.
  + - 3A: Repeat the analysis in:
      * Men
      * Women
      * Adults ≥ age 18 years at baseline
      * Children <18 years at baseline
* 4. Stratified analyses:
  + - Clinical strata (at baseline):
      * Sex (men vs. women)
      * Country (Nordic vs. elsewhere)
      * County in Sweden where index biopsy performed (Stockholm vs. not)
      * Age (<18 vs. 18 to <30 vs. 30 to <50 vs 50 to <75 vs. 75+) at index biopsy)
      * Diabetes at baseline (yes vs. no)

**SENSITIVITY ANALYSES:**

**5.** Repeat the primary analysis (see above), for patients having undergone a gastroscopy with a finding of a normal mucosa identified by SNOMED-code M00100; M00110

**Shell FLOWCHART**.

**Inclusion / Exclusion Flowchart for biopsy-proven celiac disease and matched comparators**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Patients with celiac disease** | | | | **N patients** | **N Excluded** | **% Excluded** |
| 1. Duodenum T64 (all) and jejunum T65, T65000 and T651 2. SNOMED code indicating celiac disease (M58 with subgroups) or the celiac disease diagnostic code D6218. (D6218, D62180, D62188, D6218X, D6218Y (celiac diagnosis), M58, M5800, M58000, M58001, M58005, M58006, M58007)   **Between 1/1/1990 to 12/31/2016** | | | |  |  |  |
| Duplicate personal identity number | | | |  |  |  |
| Not found in other health care registers | | | |  |  |  |
| Death on the same date or earlier than the index biopsy date | | | |  |  |  |
| Emigration earlier than the index biopsy date | | | |  |  |  |
| Other data irregularities | | | |  |  |  |
|  | | | |  |  |  |
|  | | | |  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  | | | |  |  |  |
| No matched comparators | | | |  |  |  |
| **Total** | | | | **XX** | **XX** | **XX** |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **General population comparators** | | | | **N patients** | **N Excluded** | **% Excluded** |
| **Sampled general population comparators to the identified celiac disease patients** | | | |  |  |  |
| Duplicate personal identity number | | | |  |  |  |
| Not found in other health care registers | | | |  |  |  |
| Death on the same date or earlier than the index biopsy date | | | |  |  |  |
| Emigration earlier than the index biopsy date | | | |  |  |  |
| Other data irregularities | | | |  |  |  |
|  | | | |  |  |  |
|  | | | |  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  | | | |  |  |  |
| No matched patients with microscopic colitis | | | |  |  |  |
| **Total** | | | | **XX** | **XX** | **XX** |

**SHELL TABLE A1.**

**Baseline characteristics of patients with celiac disease and matched Population Comparators (1990-2016):**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Characteristic** | | | **Overall**  **(n=XXX)** | | **Controls (n=XXX)** | | **Celiac disease**  **(n=XXX)** | |
| Sex, n (%) | | |  | |  | |  | |  | |  |
| Women | | |  | |  | |  | |  | |  |
| Men | | |  | |  | |  | |  | |  |
| Age | | |  | |  | |  | |  | |  |
| Mean (SD) | | |  | |  | |  | |  | |  |
| Median (IQR) | | |  | |  | |  | |  | |  |
| Range, min-max | | |  | |  | |  | |  | |  |
| *Categories, no. (%)* | | |  | |  | |  | |  | |  |
| <18y | | |  | |  | |  | |  | |  |
| 18y - <40y | | |  | |  | |  | |  | |  |
| 40y - <60y | | |  | |  | |  | |  | |  |
| ≥60y | | |  | |  | |  | |  | |  |
| Country of birth, n (%) | | |  | |  | |  | |  | |  |
| Nordic country | | |  | |  | |  | |  | |  |
| Other | | |  | |  | |  | |  | |  |
| Missing | | |  | |  | |  | |  | |  |
| Level of education, n (%) | | |  | |  | |  | |  | |  |
| ≤9 years | | |  | |  | |  | |  | |  |
| 10-12 years | | |  | |  | |  | |  | |  |
| >12 years | | |  | |  | |  | |  | |  |
| Missing | | |  | |  | |  | |  | |  |
| Start year of follow-up | | |  | |  | |  | |  | |  |
| 1990-2000 | | |  | |  | |  | |  | |  |
| 2001-2010 | | |  | |  | |  | |  | |  |
| 2011-2016 | | |  | |  | |  | |  | |  |
| Comorbidities ever before start of follow-up, n (%) | | |  | |  | |  | |  | |  |
| IBD | | |  | |  | |  | |  | |  |
| Diabetes  Autoimmune thyroid disease  Rhematoid arthritis | | |  | |  | |  | |  | |  |
|  |  |  | |  | |  | |  | |
|  |  |  | |  | |  | |  | |
|  |  |  | |  | |  | |  | |

**Figures: Cumulative incidence curves of time to diagnosis of microscopic colitis (follow-up until Dec 31, 2016) in:**

1A. Patients with celiac disease colitis and matched comparators (reference group)

**SHELL TABLE A4. Subgroups: Risk of microscopic colitis in patients with celiac disease and matched comparators (1990-2016):**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Group** | **N (%)** | | **N events** | | **Incidence rate (95% CI) per 1000 PY** | | **HR\***  **(95%CI)** | **HR\*\***  **(95%CI)** |
| **CD** | **Comparators** | **CD** | **Comparators** | **CD** | **Comparators** |
| **Overall** |  |  |  |  |  |  |  |  |
| Follow-up |  |  |  |  |  |  |  |  |
| 0-<1y |  |  |  |  |  |  |  |  |
| 1-<5y |  |  |  |  |  |  |  |  |
| 5-<10y |  |  |  |  |  |  |  |  |
| ≥10y |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| *≥1y of follow-up* |  |  |  |  |  |  |  |  |
| Sex |  |  |  |  |  |  |  |  |
| Women |  |  |  |  |  |  |  |  |
| Men |  |  |  |  |  |  |  |  |
| Age |  |  |  |  |  |  |  |  |
| <18y |  |  |  |  |  |  |  |  |
| 18y - <40y |  |  |  |  |  |  |  |  |
| 40y - <60y |  |  |  |  |  |  |  |  |
| ≥60y |  |  |  |  |  |  |  |  |
| Year |  |  |  |  |  |  |  |  |
| 1969-1989 |  |  |  |  |  |  |  |  |
| 1990-2000 |  |  |  |  |  |  |  |  |
| 2001-2010 |  |  |  |  |  |  |  |  |
| 2011-2017 |  |  |  |  |  |  |  |  |
| Country of birth |  |  |  |  |  |  |  |  |
| Nordic |  |  |  |  |  |  |  |  |
| Other/Missing |  |  |  |  |  |  |  |  |
| Education |  |  |  |  |  |  |  |  |
| ≤9 years |  |  |  |  |  |  |  |  |
| 10-12 years |  |  |  |  |  |  |  |  |
| >12 years |  |  |  |  |  |  |  |  |
| Missing |  |  |  |  |  |  |  |  |
| Comorbidity |  |  |  |  |  |  |  |  |
| IBD  Autoimmune thyroid disease  Rheumatoid arthritis |  |  |  |  |  |  |  |  |
| Diabetes |  |  |  |  |  |  |  |  |

\*Conditioned on matching set (age, sex, county, and calendar period);

\*\*Conditioned on matching set and further adjusted for education and baseline comorbidities (IBD, RA, diabetes, ATD)