國立臺灣大學公共衛生學院

流行病與預防醫學研究所 生物醫學統計組

碩士論文

Division of Biostatistics, Graduate Institute of Epidemiology

College of Public Health

National Taiwan University

Master Thesis

統計模式應用於胃幽門螺旋桿菌及胃癌前病變之家戶聚集性研究

Statistical Models for Family Aggregation of *Helicobacter Pylori* and Gastric Neoplasm

范僑芸

Chiao-Yun Fan

指導教授：陳秀熙 博士

Advisor: Hsiu-Hsi Chen, Ph.D.

中華民國 108 年 8 月

August, 2019

doi:10.6342/NTU201902662

國立臺灣大學碩士學位論文
口試委員會審定書

題目

統計模式應用於胃幽門螺旋桿菌及
胃癌前病變之家戶聚集性研究

Statistical Models for Family Aggregation of
Helicobacter Pylori and Gastric Neoplasm

本論文係 范僑芸 君（學號 R06849033）在國立臺灣大學流行病學與預防醫學研究所完成之碩士學位論文，於民國 年 月 日承下列考試委員審查通過及口試及格，特此證明。

口試委員：陳秀熙 （簽名）

（指導教授）

陳秋月 李鎮

誌謝

在追尋知識的道路上，方向最明確最充滿動力向前的一切都是在這裡發生的。謝謝陳秀熙教授給予我無比的關懷與力量，能跟在老師身邊學習的不僅僅是學術上的成長，更多是以人為本不違背己心的原則。謝謝嚴明芳老師、陳立昇老師、邱月暇老師、范靜媛老師跟著老師們一起經歷並學習的經驗都是我人生珍貴的體驗。謝謝彭思敏學姊、任小萱學姊、古玫生學姊、張維容學姊、林庭瑀學姊、王威淳學長，能跟大家一起學習成長真的很棒，衷心感謝碩班有你們的陪伴，完成了很多很酷的事。謝謝劉怡辰能在這裡遇到妳真是太好了！謝謝羅淳樺，從高中到現在都是我最棒的朋友，希望未來我們一起成長，成為一個更好的人。謝謝實驗室的大家，高中畢業時所立下的人生目標「成為一個有用的人」，謝謝你們的指引、幫助，我想我在前往我想成為的她。

謝謝我的家人對我所有決定無比的支持，堅信我的選擇並鼓勵我向前行，讓我在沒有過多壓力之下追求我想做的事，我會努力成為一個讓妳們無後顧之憂的有用的人，謝謝你們一直陪在我身邊。謝謝徐翊喬，陪我走過我人生最大的低潮，也陪我經歷碩班兩年的點點滴滴，希望你明年順利畢業。謝謝我的貓呼嚕仔，你是我人生最珍貴的寶藏，遇見你是一件太幸運的事，謝謝你的愛，我也愛你。

僑芸 2019

doi:10.6342/NTU201902662

中文摘要

背景

已有許多研究證實家戶內的胃幽門螺旋桿菌感染，但尚未有研究利用基因流行病學研究去量化家族聚集的影響。此外，盡管胃幽門螺旋桿菌的家族聚集已被推廣至胃部癌前病變如：萎縮性胃炎、胃黏膜腸上皮化生，卻忽略了家族聚集對疾病進程的動態影響。

目的

本論文的目的為利用 Correa 模式與先證者(Proband)的病例對照研究設計及其推展，配合不同統計模式去推估家族聚集對於胃幽門螺旋桿菌傳染及胃部癌前病變(萎縮性胃炎、胃黏膜腸上皮化生)進程的動態影響。

材料與方法

在此論文中，我們利用兩個資料集來研究家族聚集對胃幽門螺旋桿菌與和胃幽門螺旋桿菌相關之胃部癌前病變的影響。第一個資料集採用 35 個日本被感染胃幽門螺旋桿菌兒童的家族研究，其中包含兩種基因分型結果：多位點序列分型(MLST)及隨機擴增多態性 DNA 標記(RAPD)，用於檢測胃幽門螺旋桿菌的家族聚集與傳播途徑。第二個資料集是利用 Correa 模式與馬祖的胃部癌前病變的社區預防資料，包含三個時間段：1996-2003 為篩檢介入之前，2004-2007 為篩檢介入之後，2008 為全面性投藥之化學預防。

我們利用兩階段先證者的病例對照研究去推估在相同 MLST、RAPD 家族聚集

doi:10.6342/NTU201902662

的影響，與貝式有向無環圖(DAG)計算在給定觀察資料與未知資料之下的條件機率，利用隨機效應模式計算家庭內部成員胃幽門螺旋桿菌感染的家族聚集影響。

我們利用多元邏輯斯迴歸分析、離散狀態及時間之馬可夫鍊模式與連續時間下馬可夫過程三種模型估計在 Correa 模式之下家族聚集對胃幽門螺旋桿菌感染與胃部癌前病變動態轉移的影響。

結果

第一部份：日本基因序列與胃幽門螺旋桿菌感染的家族聚集研究

當使用第一個孩子作為指標個案時，貝式 DAG 模型的 DNA 指紋序列的估計結果顯示匹配序列的家族聚集在 MLST 為 56 倍（95%CI：3.99-1878.07），及 RAPD 為 68.10 倍（95%CI：4.85-2171.12）。

第二部分：家族聚集對胃幽門螺旋桿菌感染及胃部癌前病變動態轉移平衡的影響

應用馬祖資料與多元邏輯斯迴歸模式，在調整抽菸與飲食習慣之下，家族聚集的影響在胃幽門螺旋桿菌感染、萎縮性胃炎與胃黏膜腸上皮化生皆有統計上的顯著，其估計值自 1.39 (95% CI: 1.09-1.77) 至 2.02 (95% CI: 1.26-3.55)。

符合遍歷理論的馬可夫鍊的估計結果表明，在 Correa 模式之下於 1996 年存在強烈的家族聚集趨勢，其上三角形與下三角形的比為 2.10。然而在 2004 與 2008 的相應數值為 0.91 與 0.87。

連續時間馬可夫過程的估計結果證明家族聚集與 Correa 模式的動態轉移平衡存在高度相關。主要貢獻為從正常狀態到胃幽門螺旋桿菌感染，指標個案疾病狀態為胃幽門螺旋桿菌感染、萎縮性胃炎、及胃黏膜腸上皮化生其倍數分別為：1.30 (95% CI: 1.18-1.44), 1.21 (95% CI: 1.07-1.36), 和 1.35 (95% CI: 1.14-1.59)。而在第二階段自胃幽門螺旋桿菌感染轉移至萎縮性胃炎，其倍數根據不同指標個案狀態為

doi:10.6342/NTU201902662

0.74 (95% CI: 0.62-0.88), 1.33 (95% CI: 1.10-1.60), 和 1.06 (95% CI: 0.82-1.37)。在第三階段萎縮性胃炎轉移至胃黏膜腸上皮化生結果未達統計上顯著，倍數分別為 1.28 (95% CI: 0.94-1.75), 0.95 (95% CI: 0.68-1.34), 和 1.14 (95% CI: 0.72-1.81)。

結論

本論文利用統計模式去評估胃幽門螺旋桿菌感染與 Correa 模式下胃部癌前病變動態轉移的影響。所提出的模式應用於一個具有基因分型結果的資料集，與一基於社區的胃部癌前病變資料在預防措施介入前與介入後。

doi:10.6342/NTU201902662

ABSTRACT

Introduction

Although intra-familial aggregation of *helicobacter pylori infection* has been well studied before, the effect size of family aggregation has been scarcely elucidated by using a well-designed genetic epidemiological study. While the familial aggregation of HP has been extended to include pre-cancerous lesions like atrophy gastritis (AG) and intestine metaplasia (IM), the effect size of familial aggregation accounting for the dynamic transition study has been even neglected.

Aims

The objective of this thesis is to employ a case-control proband study design and its variants in conjunction with different statistical models to estimate the effect size of familial aggregation of genetic typing associated with HP infection and the dynamic transitions between HP, AG, and IM under the context of the Correa model.

Material and Methods

Two data sources were used for studying familial aggregation of HP and HP-related pre-cancerous lesions. The first dataset was derived from a 35 index Japanese pediatric patients familial study with the available information on two genetic typing procedures, Multilocus Sequence Typing (MLST) and random amplified polymorphic DNA (RAPD) fingerprinting used for detecting transmission route of familial aggregation. The second dataset was descended from the Matsu community-based prevention of gastric neoplasm under the context of Correa model with three periods, 1996-2003 before intervention, 2004-2007 after screening, and 2008 after chemoprevention. The case-control proband study design with two-state was designed to estimate the effect size of familial aggregation associated with the same sequence of MLST and also RAPD.

doi:10.6342/NTU201902662

Bayesian directed acyclic graphic (DAG) model was built up to develop fully conditional distribution given the observed data and unknown quantity to estimate the effect size of family aggregation of HP infection making allowance for the correlation of HP infection across the same family members with the random effect model.

To model how family aggregation affects the dynamic transition of HP and gastric pre-cancerous neoplasm under the context of the Correa multistate model, three statistical approaches were used, including the multi-nominal logistic regression model, the discrete-state and discrete-time Markov chain model, and continuous-time Markov process.

Results

Part I Japanese familial aggregation study on genotyping associated with HP infection

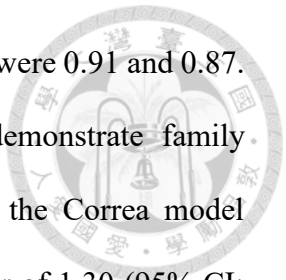
When using first child as index case the estimated results on the DNA fingerprint sequence with Bayesian DAG model show the tendency of familial aggregation for the matched sequence of RAPD was 56-fold (95% CI: 3.99-1878.07) for RAPD and 68.10-fold (95% CI: 4.85-2171.12) for MLST compared with the unmatched sequence.

Part II Family aggregation for the dynamic transition of HP and gastric neoplasm

With the application of multi-nominal logistic regression model to the Matsu data, the effect sizes of familial aggregation of HP infection, AG, and IM were statistically significant with the range from 1.39 (95% CI: 1.09-1.77) to 2.02 (95% CI: 1.26-3.55) after adjustment for smoking and dietary factors.

The estimated results of Markov chain with the ergodicity theory show there was a strong tendency of familial aggregation for reaching equilibrium with the property of the Correa model in 1996 but not in the two periods after interventions as the ratio of the summation of transition probability in the upper triangle to that of in the lower triangle

doi:10.6342/NTU201902662



was around 2.24 whether the corresponding values in 2004 and 2008 were 0.91 and 0.87.

The estimated results of continuous-time Markov process demonstrate family aggregation was highly associated with the dynamic transition of the Correa model mainly in the transition from normal to HP infection with the order of 1.30 (95% CI: 1.18-1.44), 1.21 (95% CI: 1.07-1.36), and 1.35 (95% CI: 1.14-1.59) for the index case of HP infection , AG, and IM. The influence of family aggregation dwindled from HP infection to AG with the order of 0.74 (95% CI: 0.62-0.88), 1.33 (95% CI: 1.10-1.60), and 1.06 (95% CI: 0.82-1.37) to AG to IM with insignificant findings, 1.28 (95% CI: 0.94-1.75), 0.95 (95% CI: 0.68-1.34), and 1.14 (95% CI: 0.72-1.81)

Conclusions

Statistical models are proposed here to model familial aggregation of HP infection and the dynamic transition of HP-related gastric precancerous lesion under the context of the Correa model. The proposed models were applied to one data with genotyping HP infection and the community-based data before and after prevention of gastric neoplasm.

doi:10.6342/NTU201902662

CONTENTS

doi:10.6342/NTU201902662

doi:10.6342/NTU201902662

doi:10.6342/NTU201902662

List of Figures

doi:10.6342/NTU201902662

List of Tables

doi:10.6342/NTU201902662

doi:10.6342/NTU201902662

Chapter 1 Introduction

Several causes responsible for the occurrence of gastric cancer have been well studied, including life style factors, dietary factors, and Helicobacter pylori infection in accordance with the Correa model [Suerbaum and Michetti, 2002]. Helicobacter pylori is the single most common cause of the development of gastric ulcer and stomach cancer [Blaser, 1996; Fukase et al., 2008; IARC Helicobacter pylori Working Group, 2014]. The results from previous studies have shown that more than 90% of gastric cancers is associated with H. pylori infection [Lee et al., 2016], suggesting that gastric cancer development possibly depends on individual's genetic susceptibility. Genetic susceptibility of individuals has been proposed as an additional important cause which may contribute to gastric cancer and its precursors [Bamford et al., 1993; Goodman and Correa, 1995]. The genetic susceptibility of individuals to environmental exposure factors considering the familial clustering phenomenon of gastric cancer also plays a role in the development of gastric cancer [Yokota et al., 2015; Liang and Pulver, 1996].

The interaction between transition mode of Helicobacter pylori infection and genetic susceptibility within family has increasingly gained attention. Intrafamililal infection of H. pylori has become preferential, especially in developed countries. The possibility of intraspousal infection of H.pylori was evaluated in Japan [Yokota et al., 2015]. However, the effect size of family aggregation has been scarcely elucidated by using a well-designed genetic epidemiological study. It might not be very efficient to conduct genomic studies without understanding the role of family aggregation on H.pylori infection and gastric cancer development. To gain a better understanding of how family aggregation accounts for different staging of gastric carcinogenesis is very important for designing genomic epidemiological study in association with gastric cancer.

doi:10.6342/NTU201902662

Although the case–control/family sampling design has been proposed by Liang et al. [Liang et al., 1996] and applied to assess the effect size of familial aggregation of metabolic syndrome in Chiu et al.'s study [Chiu et al., 2007], the statistical method underpinning Bayesian approach for evaluating the effect size of familial aggregation with small sample sizes has been rarely addressed. Moreover, the effect size of familial aggregation accounting for the dynamic transition study has been even neglected while the familial aggregation of HP has been extended to include pre-cancerous lesions like atrophy gastritis (AG) and intestine metaplasia (IM). This requires the development of a mathematical model to estimate the effect size of familial aggregation on the dynamic transitions between HP, AG, and IM under the context of the Correa model

We therefore aimed to employ a case-control proband study design and its variants in conjunction with different statistical models to estimate the effect size of familial aggregation of genetic typing associated with HP infection and the dynamic transitions between HP, AG, and IM using data derived from two data sources, one from a 35 index Japanese pediatric patients familial study and another from a population-based Taiwanese Matsu screening cohort.

.

doi:10.6342/NTU201902662

Chapter 2 Literature Review

2.1 Familial aggregation of Helicobacter pylori Infection

2.1.1 Introduction of Helicobacter pylori

Helicobacter pylori infection has been proved to be a high-risk factor for some gastrointestinal disorders, including gastritis, gastric and duodenal ulcers, gastric adenocarcinoma, and gastric lymphoma. [Suerbaum et al., 2002; Blaser 1996; Fukase et al., 2008; Chiu et al., 2012; Watari et al., 2014] Furthermore, there are a lot of reports showing Helicobacter pylori is the most important etiologic factor for gastric cancer [Lee et al., 2013], it had been taken more than 720,000 human's lives away each year. This makes the control for infection of Helicobacter pylori an important issue worldwide.

Helicobacter pylori is one kind of bacterium, which could be treated by antibiotics, so, the new era of treatment for gastric diseases has been discovered; on the other hand, Helicobacter pylori would infect others by various transmission modes, therefore, the control for Helicobacter pylori infection not only take individual consideration, but also the probability of contact from aggregation (ex. family), cluster, or environment [Parsonnet et al., 1991; Nomura et al., 1991].

2.1.2 Infection and transmission of *Helicobacter pylori*-Familial aggregation

Oral-oral, gastro-oral, and fecal-oral are the main transmission modes for Infection and transmission of *Helicobacter pylori*, therefore socioeconomic status and hygiene status play important roles in reducing the prevalence of *Helicobacter pylori* infected. However, both gastro-oral and fecal-oral routes might be through contaminated water of

doi:10.6342/NTU201902662

food, such as raining water collection by wells, underground water, etc. There were many studies demonstrated the cluster infection of *Helicobacter pylori* through the same contaminated water use, especially for those area in developing countries. The procedure of boiling and clean pipe-water supply is strongly recommended to prevent *Helicobacter pylori* infection [Aziz et al., 2015]. Besides the same source sharing of *Helicobacter pylori* infection, the transmission modes may be through vertical mode (from parents to children) or horizontal mode (across environment and individual contact). In the rural or low economic area, the horizontal mode is the important pathway for transmission because the environment with high *Helicobacter pylori* prevalence rate, therefore, the contaminated water and food are common sharing the infectious source. But, in low prevalence rate of *Helicobacter pylori* areas, the vertical transmission route is the key for *Helicobacter pylori* infection, like the eating pattern between parents and children and high contact rate [Vale and Vitor, 2010].

From the viewpoint of water supply and use as Helicobacter pylori transmission route, the family cluster would be the most probable way due to the common exposure source sharing. Therefore, family plays an important role for *Helicobacter pylori* transmission pathway investigation, such as dietary pattern, family size, and residential density type, behavioral relationship, and the prevalence rate of *Helicobacter pylori* in outside surrounding as background prevalence, etc. The intrafamilial clustering of *Helicobacter pylori* infection would be represented on numbers of *Helicobacter pylori* infection and the contact rate within family members.

2.1.3 DNA-fingerprinting of H pylori. infection

The intrafamilial clustering not only presents the environmental sharing, but also indicates the host genetic susceptible for *Helicobacter pylori* infection [Kivi et al., 2006]. Those who got *Helicobacter pylori* are believed been infected when they are under

doi:10.6342/NTU201902662

5 years old, as secretion of gastric acid is immature [Suerbaum et al., 2002; Rajindrajith et al., 2009]. In third world countries, children are sent to care workers who take care of a lot of children from different families, with the highly contact the *Helicobacter pylori* can be spread through contaminated water and food. Nevertheless, higher socioeconomic status and hygiene status in developed country make the *Helicobacter pylori* cannot be spread by horizontal transmission route. That is the reason why familial infected become much important in developed country [Dominici et al., 1999; Pryczynicz et al., 2002; Miline et al., 2009].

The technology for genome scan has been developed decades ago, therefore, the specific origin virus or bacteria can be very efficiently using polymerase chain reaction (PCR) amplifying tech determined which sub-branch or subtype they are. The examination can be performed from blood/serum or fecal DNA also can be extracted to determine *Helicobacter pylori* [McMillan et al., 2011; Mamishi et al., 2016; Puz et al., 2008]. In 2003, Roma-Giannikou et al. conducted the PCR-based random amplified polymorphic DNA (RAPD) fingerprinting method to identify the source of *Helicobacter pylori* infection for the intrafamilial spread cases, including 32members of 11 families. For the family members, they recruited child, mother, and father simultaneously to perform the genomic analysis. The result showed the DNA fingerprinting pattern of members of same family were very close. This evidence strongly supported the transmission route of *Helicobacter pylori* in family is person-to-person or from the same infectious source [Roma-Giannikou et al., 2003].

Besides the cross-sectional association in family spread, in 2005 in Japan, Konno et al. employed the family cohort study design with 5 years follow-up to revealed *Helicobacter pylori* through mother-to-child transmission. Recruiting 350 pregnant women, there were 69 women with *Helicobacter pylori* positive results, using serology

doi:10.6342/NTU201902662

or stool test. Those babies were followed-up by RAPD surveillance from birth to 5 years old. Those babies' gastric juice was examined by RAPD method. Finally, there were 5 babies acquired *Helicobacter pylori* infection. The infected time is around 1 to 2 years since birth. The most important information is the strains of *Helicobacter pylori* identified by RAPD were identical to mother, proving the intrafamilial mother-to-child transmission (Konno et al., 2005). Regarding those children aged 4-19 years, Konno et al. also invited 66 familial members who serum or stool with positive results of *H pylori*. Among those patients, there were 76% of those who had identical type with any one fof amilial member and 69% were same as their mother's type. The identical relation or mother-to-child was significantly higher than father-to-child (Konno et al., 2008). The evidence strongly suggested that *Helicobacter pylori* infection in intrafamilial transmission is tend to dominate by mother-to-child route. The similar result also revealed by study from Bangladesh [Nahar et al., 2009].

To achieve the examination convenience and success, especially for children without endoscopy examination using biopsy, the multilocus sequence typing (MLST) analysis has become popularly and the *H pylori*. housekeeping genes has been published in website for analysis, therefore, this method would help intrafamilial infection studies for exploration [Osaki et al, 2013; Osaki et al, 2015]. In a Japanese study, Yokota et al used MLST to prove 25 of 35 infected children in Japan can find out their mother who had also been infected by the same gene *Helicobacter pylori* [Yokota et al., 2015] These results provide strong evidence that there are clearly familial aggregation in *Helicobacter pylori* infection.

2.1.4 Critiques

To our knowledge, the most efficient way of treating *Helicobacter pylori* is the use of drug therapy. However, *Helicobacter pylori* has been reported to acquire resistance to

doi:10.6342/NTU201902662

various antimicrobial drugs, and the proportion of those with drug-resistant *Helicobacter pylori* has been increasing in recent years [Kobayashi et al., 2007; Kobayashi et al., 2013; Hashinaga et al., 2016]. These may result in declining success in eradicating *Helicobacter pylori.* Thus, identifying the most probable transmission route is the most critical issue in eradicating *Helicobacter pylori*.

According to the guidelines, the following are most frequently adopted diagnosed methods: a rapid urease test, direct microscopic count, microbial culture, urea breath test, antibody measurement, or stool antigen test [JSHR, 2016]. However, patients are determined according to qualitative result, yet these methods could not efficiently identify those being positive that were infected by the same transmission route or shared the same origin, especially for the clustering infection or intrafamilial spreading events. As a result, some studies have started to use state-of-the-art molecular typing or DNA-fingerprinting methods to evaluate the possibility of intra-familial transmission.

Although both studies in Japan and Bangladesh have suggested that mother-to-children is the dominant transmission route in *Helicobacter pylori* [Nahar et al., 2009; Yokota et al., 2015], they cannot provide definite evidence for a transmission route. In addition to the source identification of same trait of *Helicobacter pylori* by (RAPD) fingerprinting method and the multilocus sequence typing (MLST) analysis, these studies did not evaluate or discuss the quantitative estimation of familial aggregation of genetic typing associated with *Helicobacter pylori* infection. Specifically, they lack the data and statistical models to identify the effect of familial aggregation in consideration of lifestyle patterns and shared environment in a family, such as the close contact of father and child or mother and child. In addition, the dynamic transitions between *Helicobacter pylori*, atrophy gastritis (AG) and intestine metaplasia (IM) under the context of the Correa model have rarely been taken into account.

doi:10.6342/NTU201902662

Since *Helicobacter pylori* is the most important etiologic factor for gastric cancer [Lee et al., 2013], we should take into account all possible scenarios of *Helicobacter pylori* infection. According to the previous studies, nearly 75% have identical strain of HP with any one of familial member and 69% are the same as their mother's [Konno et al., 2005; Konno et al., 2008]. Evidence have led to a strong evidence that there is clearly familial aggregation in Helicobacter pylori infection. As a result, quantitative analysis is relatively important in this stage to evaluate the effect of familial aggregation. Based on the viewpoints and evidence above, the construction of a population-based dataset which consists of the information of family structures as well as the characteristics and variables of interest among all family members is required. Family-based pedigree or population-based proband-oriented pedigree information system is also needed to completely capture all possible scenarios of HP infection. Application of different statistical models to evaluate the effect size of familial aggregation on the dynamic transition of HP and gastric pre-cancerous neoplasm should also be adopted.

2.2 Model for familial aggregation

2.2.1 Case-control proband study on Familial aggregation data

Based on the case–control/family sampling design proposed by Liang and Pulver Liang eta l., 1996], Chiu et al study familial aggregation of metabolic syndrome with separated proband in to two groups: healthy or Met-s, to compare the difference of prevalence on Met-s in relatives of probands [Chiu et al., 2007]. This study random sample 1500 proband with Met-s, otherwise in order to adjust the difference baseline of proband this study use 1:2 matching process, shows in Figure1. The relationship between relatives of same proband might not be independent, thus the assumption of independent in the conventional statistical model might not be appropriate. Therefore, using the logistic

doi:10.6342/NTU201902662

model provided by Liang and Pulver [Liang eta l., 1996] which allows the correlation between familial aggregation, environmental risk have be defined as

$$\log\left(\frac{P_{ij}}{1-P_{ij}}\right) = \alpha + \beta_1 X_1 + \cdots + \beta_p X_p + \delta Z_i$$

With p environmental risk factor ($X_1 - X_p$, $X_1 = \text{coffee}$, $X_2 = \text{meat}, \ldots$) and Z_i show the status of disease for i_{th} proband. The odds ratio (95% C.I) of familial aggregation based on case-proband and control-proband is 1.70 (1.41-2.04), this result shows there are strong familial aggregation in association of Metabolic syndrome.

2.2.2 Population-based proband-oriented pedigree information system

In order to investigate further into the familial aggregation of different disease, Chiu et al based-on case-control proband study (Chiu et al., 2012), build a proband-oriented pedigree information system, calculating the relative score based on relationship with proband to define different level of relatives. However, the screening data set might not full cover all the family member in registration system, Chiu et al adjusted the difference of Real-life family structure and Theoretical familial relationships. Take first generation for example, the theoretical family structure among the identified pedigree of families is denoted by K, was been calculated by

$$Z_1 + Z_2 + Z_3 + Z_6 + Z_7 = k - 1$$

Which Z_j (j in 1,2,3,6,7) represents the number of family members in relative status of j. But the numbers of relatives are subject to the social norms of relative relationships, thus some theoretical combinations (eg, spouse≥2) are unreasonable. Therefore let r stand for the upper bond to number of probable relative relationships, and the number of theoretical combinations subject to the constraints r and total k people in this family can be defined as

doi:10.6342/NTU201902662

$$H_{k-1}^{r} - H_{(k-1)-2}^{r}, \text{ if } Z_1 \leq 1 \; and \; 0 \leq Z_1 \leq k-1$$

After adjusting the difference of real capture family structure and theoretical family structure, the study shows the higher score of proband-oriented the higher the chance you get hypertension. The adjusted hazard ratio of relative score ≥15 verse <6 is 1.72(95 CI: 1.55-1.90), showing the highly evidence of familial aggregation about hypertension.

doi:10.6342/NTU201902662

Chapter 3 Material and Method

3.1 Data

3.1.1 Japan HP infection Data

In 2015, Yokota et al. collected family aggregation data based on 35 index pediatric patients with their family members to analyze the infrafamilial infection route based on both approaches of multilocus sequence typing (MLST) and random amplified polymorphic DNA (RAPD). The PCCR-based RAPD were carried out by 4 primers to stabilize the DNA amplification and analysis. For the MLST method, they used eight housekeeping (atpA, efp, mutY, ppa, trpC, ureI, yphC, and vacA) to determine the genome types from website (http://pubmlst.org/helicobacter/). All information from 35 families including family member (children or mother or father), H pylori. positive/negative using stool antigen test, RAPD pattern, and MLST types were provided in this article. This article with the details of family information is very useful to analyze the cluster of H pylori. interfamilial transmission and infection among families (Yokota et al., 2015). Therefore, the data were reorganized as the following tabulate data for our further analysis. It should be noted that the analysis was based on 34 families because we excluded one family with more complicate HP transmission.

doi:10.6342/NTU201902662

Table 3.1 Data used for studying family aggregation of carried gene on HP

Family No.	Family member	Index Child	HP infection	RAPD patterns	Allele sequence type								
					atpA	efp	mutY	ppa	trpC	urel	yphC	vacA	mlst
1	F		+	○	2322	2203	2362	56	2434	2409	2408	521	2766
	M		+	○	2322	2204	2362	56	2434	2409	2408	521	2783
	C1	✓	+	○	2322	2203	2362	56	2434	2409	2408	521	2766
	C2		+	○	2322	2203	2362	56	2434	2409	2408	521	2766
	C3		+	○	2322	2203	2362	56	2434	2409	2408	521	2766
2	F		+	○	1232	2206	2368	52	2436	2411	2410	522	2784
	M		+	○	1232	2206	2368	52	2436	2411	2410	522	2784
	C1		+	○	1232	2206	2368	52	2436	2411	2410	522	2784
	C2	✓	+	○	1232	2206	2368	52	2436	2411	2410	522	2784
3	F		+	○	2328	2219	2370	2248	2441	2425	2419	526	2787

doi:10.6342/NTU201902662

Family No.	Family member	Index Child	HP infection	RAPD patterns	Allele sequence type								
					atpA	efp	mutY	ppa	trpC	urel	yphC	vacA	mlst
	M		+	○	2328	2219	2376	2248	2441	2425	2419	526	2788
	C1	✓	+	○	2328	2219	2370	2248	2441	2425	2419	526	2787
	C2		+	○	2328	2219	2370	2248	2441	2425	2419	526	2787
4	F		+	○	2319	2191	2390	945	954	36	957	529	2790
	M		+	○	2319	2191	2355	945	954	36	957	529	2757
	C1	✓	+	○	2319	2191	2355	945	954	36	957	529	2757
	C2		+	○	2319	2191	2355	945	954	36	957	529	2757
5	F		+	○	2331	2220	2384	1125	2442	2426	2422	527	2789
	M		+	○	2331	2220	2384	1125	2442	2426	2422	527	2789
	C1	✓	+	○	2331	2220	2384	1125	2442	2426	2422	527	2789
	C2		+	n.d.	.	.	.	.	.	.	.	.	.

doi:10.6342/NTU201902662

Family No.	Family member	Index Child	HP infection	RAPD patterns	Allele sequence type								
					atpA	efp	mutY	ppa	trpC	urel	yphC	vacA	mlst
6	F		+	○	2327	2213	1865	2246	2440	2419	2417	524	2786
	M		+	○	2327	2213	1865	2246	2440	2419	2417	524	2786
	C1		+	n.d.	.	.	.	.	.	.	.	.	.
	C2	✓	+	○	2327	2213	1865	2246	2440	2419	2417	524	2786
7	F		+	○	2360	2246	2425	502	2499	2471	2453	542	2802
	M		+	○	2360	2246	2425	502	2499	2471	2453	543	2802
	C1	✓	+	○	2360	2246	2425	502	2499	2471	2453	542	2802
	C2		-	n.d.	.	.	.	.	.	.	.	.	.
8	F		+	○	1332	2212	2369	2244	2438	2418	2414	523	2785
	M		+	○	1332	2212	2369	2244	2438	2418	2414	523	2785
	C1	✓	+	○	1332	2212	2369	2244	2438	2418	2414	523	2785

doi:10.6342/NTU201902662

Family No.	Family member	Index Child	HP infection	RAPD patterns	Allele sequence type								
					atpA	efp	mutY	ppa	trpC	urel	yphC	vacA	mlst
9	F		+	○	1760	2185	2354	2232	457	2393	457	528	2754
	M		+	○	1760	2185	2354	2232	457	2393	457	528	2754
	C1	✓	+	○	1760	2185	2354	2232	457	2400	457	528	2755
10	F		+	●	2305	2266	2342	2153	2502	2390	2307	544	2803
	M		+	○	2366	2269	2347	1125	2419	2391	2345	460	2804
	C1	✓	+	○	2309	2269	2347	1125	2419	2391	2345	546	2805
	C2		+	n.d.	.	.	.	.	.	.	.	.	.
	C3		+	n.d.	.	.	.	.	.	.	.	.	.
11	F		+	●	2339	2234	2395	2256	2447	2445	2434	532	2793
	M		+	○	2340	2236	2397	2257	2448	2553	2435	533	2794
	C1	✓	+	○	2340	2236	2397	2257	2448	2553	2435	533	2794

doi:10.6342/NTU201902662

Family No.	Family member	Index Child	HP infection	RAPD patterns	Allele sequence type								
					atpA	efp	mutY	ppa	trpC	urel	yphC	vacA	mlst
	C2		-	n.d.	.	.	.	.	.	.	.	.	.
	C3		-	n.d.	.	.	.	.	.	.	.	.	.
12	F		+	●	2356	2220	2414	2271	2475	2462	2448	538	2798
	M		+	○	2357	2242	2419	445	2482	2463	2451	539	2799
	C1	✓	+	○	2357	2242	2419	445	2482	2463	2451	539	2799
	C2		-	n.d.	.	.	.	.	.	.	.	.	.
	C3		-	n.d.	.	.	.	.	.	.	.	.	.
13	F		+	●	2313	2276	2359	942	2508	1968	2401	547	2806
	M		+	○	2374	2202	2360	2235	2433	2406	2404	548	2807
	C1		+	○	2374	2202	2360	2235	2433	2406	2404	548	2807
	C2	✓	+	○	2374	2202	2360	2235	2433	2406	2404	548	2807

doi:10.6342/NTU201902662

Family No.	Family member	Index Child	HP infection	RAPD patterns	Allele sequence type								
					atpA	efp	mutY	ppa	trpC	urel	yphC	vacA	mlst
14	F		+	●	2377	2285	459	2282	458	1968	2463	549	2808
	M		+	○	2382	2290	2368	2290	2513	2480	2469	550	2809
	C1	✓	+	○	2382	2290	2368	2290	457	2480	2469	550	2810
	C2		-	n.d.	.	.	.	.	.	.	.	.	.
15	F		+	●	255	2230	2393	226	2443	2442	2426	530	2791
	M		+	○	2338	2232	2394	179	2445	2443	2431	531	2792
	C1	✓	+	○	2338	2232	2394	179	2445	2443	2431	531	2792
16	F		+	●	949	2244	946	2272	2491	969	460	540	2800
	M		+	○	2313	462	2369	2274	2498	2470	2452	540	2801
	C1	✓	+	○	2313	462	2369	2274	2498	2470	2452	540	2801
17	F		+	●	2345	455	2399	2261	2419	966	1247	535	2795

doi:10.6342/NTU201902662

Family No.	Family member	Index Child	HP infection	RAPD patterns	Allele sequence type								
					atpA	efp	mutY	ppa	trpC	urel	yphC	vacA	mlst
	M		+	○	2350	2237	2400	2266	1225	2456	2440	536	2796
	C1	✓	+	○	2350	2237	2400	2266	1225	2456	2440	536	2796
	C2		+	■	2353	2239	2402	2268	2472	2460	2414	537	2797
18	F		-	n.d.	.	.	.	.	.	.	.	.	.
	M		+	○	1760	455	2463	2295	2516	2506	2484	556	2817
	C1		-	n.d.	.	.	.	.	.	.	.	.	.
	C2	✓	+	○	1760	455	2463	2295	2516	2506	2484	556	2817
	C3		-	n.d.	.	.	.	.	.	.	.	.	.
19	F		-	n.d.	.	.	.	.	.	.	.	.	.
	M		+	○	2388	2308	2464	2296	2519	2507	2485	557	2818
	C1	✓	+	○	2308	2309	2464	2296	2519	2507	2485	557	2819

doi:10.6342/NTU201902662

Family No.	Family member	Index Child	HP infection	RAPD patterns	Allele sequence type								
					atpA	efp	mutY	ppa	trpC	urel	yphC	vacA	mlst
	C2		-	n.d.	.	.	.	.	.	.	.	.	.
	C3		-	n.d.	.	.	.	.	.	.	.	.	.
20	F		-	n.d.	.	.	.	.	.	.	.	.	.
	M		+	○	2364	2244	454	2276	458	2437	2447	536	2843
	C1	✓	+	○	2364	2244	454	2276	458	2437	2447	536	2843
	C2		-	n.d.	.	.	.	.	.	.	.	.	.
21	F		-	n.d.	.	.	.	.	.	.	.	.	.
	M		+	○	940	1772	2451	2292	2516	2493	2481	553	2813
	C1	✓	+	○	940	1772	2451	2292	2516	2495	2481	553	2814
22	F		+	○	2393	2313	2419	2298	252	2524	2493	561	2823
	M		+	●	2391	2312	2467	420	423	2516	2492	560	2822

doi:10.6342/NTU201902662

Family No.	Family member	Index Child	HP infection	RAPD patterns	Allele sequence type								
					atpA	efp	mutY	ppa	trpC	urel	yphC	vacA	mlst
	C1		+	n.d.	.	.	.	.	.	.	.	.	.
	C2		-	n.d.	.	.	.	.	.	.	.	.	.
	C3	✓	+	○	2393	2313	2419	2298	252	2524	2493	561	2823
23	F		+	●	2340	2236	2397	2330	2448	2553	2435	577	2837
	M		-	n.d.	.	.	.	.	.	.	.	.	.
	C1		-	n.d.	.	.	.	.	.	.	.	.	.
	C2	✓	+	○	2463	2327	2484	26	2538	1362	2551	578	2838
	C3		-	n.d.	.	.	.	.	.	.	.	.	.
24	F		+	●	2421	2320	2475	2306	2530	2536	2538	569	2829
	M		-	n.d.	.	.	.	.	.	.	.	.	.
	C1		+	○	2430	453	2476	445	2531	2545	460	536	2830

doi:10.6342/NTU201902662

Family No.	Family member	Index Child	HP infection	RAPD patterns	Allele sequence type								
					atpA	efp	mutY	ppa	trpC	urel	yphC	vacA	mlst
	C2	✓	+	○	2430	453	2476	445	2531	2545	460	536	2830
25	F		+	●	1760	2330	449	445	2541	44	1937	581	2841
	M		-	n.d.	.	.	.	.	.	.	.	.	.
	C1	✓	+	○	2466	2331	2463	2337	2542	50	2557	582	2842
	C2		+	○	2466	2331	2463	2337	2542	50	2557	582	2842
26	F		+	●	2439	2321	2477	2307	2532	2547	2539	571	2831
	M		-	n.d.	.	.	.	.	.	.	.	.	.
	C1	✓	+	○	2447	2322	2478	942	458	2548	2541	572	2832
27	F		+	●	2458	2324	2481	420	2535	2551	1962	571	2835
	M		-	n.d.	.	.	.	.	.	.	.	.	.
	C1	✓	+	○	2461	2184	2482	45	2532	2552	2547	576	2836

doi:10.6342/NTU201902662

Family No.	Family member	Index Child	HP infection	RAPD patterns	Allele sequence type								
					atpA	efp	mutY	ppa	trpC	urel	yphC	vacA	mlst
28	F		+	●	2449	455	2479	2308	2533	2549	2543	573	2833
	M		-		.	.	.	.	.	.	.	.	.
	C1	✓	+	○	2452	2293	2480	2328	2534	2550	2546	574	2834
29	F		+	●	2464	2328	2485	2331	2539	2554	2555	579	2839
	M		-	n.d.	.	.	.	.	.	.	.	.	.
	C1	✓	+	○	2464	2329	2486	2332	2540	2555	2556	580	2840
30	F		-	n.d.	.	.	.	.	.	.	.	.	.
	M		+	●	2386	909	2452	2294	2517	2504	2307	554	2815
	C1	✓	+	○	1760	2307	2454	445	457	1247	457	555	2816
31	F		-	n.d.	.	.	.	.	.	.	.	.	.
	M		+	●	2389	2310	2465	2297	2520	2510	2486	558	2820

doi:10.6342/NTU201902662

Family No.	Family member	Index Child	HP infection	RAPD patterns	Allele sequence type								
					atpA	efp	mutY	ppa	trpC	urel	yphC	vacA	mlst
	C1	✓	+	○	2390	2311	2466	502	2521	2514	2488	559	2821
32	F		-	n.d.	.	.	.	.	.	.	.	.	.
	M		+	●	2398	2316	2471	2301	2472	2529	2525	565	2825
	C1	✓	+	○	2400	2317	2472	2301	2527	2533	2533	566	2826
33	F		-	n.d.	.	.	.	.	.	.	.	.	.
	M		+	●	940	2246	2450	502	2418	2482	2477	551	2811
	C1	✓	+	○	2384	2298	2450	502	2515	2485	2477	552	2812
34	F		-	n.d.	.	.	.	.	.	.	.	.	.
	M		+	●	2357	2202	2360	2235	2433	2406	2535	548	2827
	C1	✓	+	○	2420	2319	2474	1254	2529	459	2536	568	2828

doi:10.6342/NTU201902662

3.1.2 Matsu gastric cancer screening program

The study samples were derived from community-based gastric cancer screening program in the Matsu islands in Taiwan. Between 1995 and 1998, a total of 2,201 residents participated in a two-stage screening project for gastric neoplasia. Of non-attenders (n=1,370), some precursors were identified through opportunistic screening.

We can rebuild the familial pedigree of 1,637 families, which at least one person has complete information on Helicobacter pylori infection test and gastroscopy test, and other subjects that form the basis of the following analysis based on screening and house registry profiles. In order to study the role of familial aggregation in Pre-Malignance of gastric, the study design is based on the case–control family proband sampling design proposed by Chiu (Chiu et al., 2007), to compare the prevalence of pre-cancerous lesions with status of proband between normal and gastric pre-malignance.

doi:10.6342/NTU201902662

3.2 Case-control proband method

3.2.1 Generalized estimated equation model (GEE)

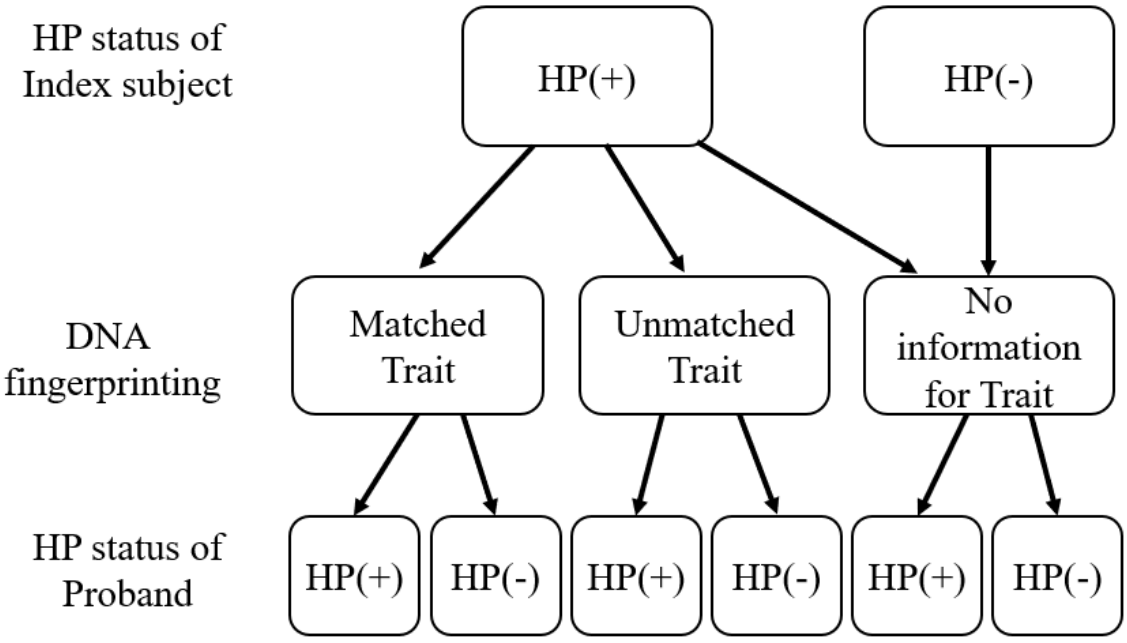


Figure 3.1 Flow chart of case-control proband study

We used the case-control proband study design to estimate the effect size of family aggregation associated with the same sequence of MLST only, RAPD only and MLST in conjunction with RADP. In order to capture the correlation of HP infection across the same family, the generalized estimating equations (GEE) was developed to estimate the effect size of family aggregation making allowance for the correlation of HP infection across family members. The advantage of using GEE is that the estimates of the effect size based on GEE model which focus on the marginal effect are robust when the covariance structure are mis-specified.

In our model, the dependent variable (Y_{ij}) is the binary outcome of positive HP (y = 1) and negative (y = 0) HP in the j_{th} proband derived from the i_{th} index subject, and the independent variable (Z_{ij}) which is set to 1 for HP(+) index and 0 for HP(-) index.

doi:10.6342/NTU201902662

For the advanced model, we also consider the sequence of MLST and RAPD as the independent variable (Z_{1ij},Z_{2ij}), which is set to (1,0) for the matched traits of DNA sequence, (0,1) for the unmatched traits and (0,0) for HP(-) index without information on DNA sequence because of negative result of HP. The model can be expressed as

$$logit\ P_{ij}(Y=1) = \alpha_i + \gamma_1 Z_{1ij} + \gamma_2 Z_{2ij},$$

$$\begin{cases} Z_{1ij} = 1, matched \text{ DNA sequence in HP}(+) \\ Z_{2ij} = 1, unmatched \text{ DNA sequence in HP}(+) \end{cases}$$

ij^{th}: j^{th} proband derived from the i^{th} index subject,

The covariance structure of $\boldsymbol{Y}$ in the GEE model can be expressed as

$$Cov(Y) = \begin{bmatrix} \boldsymbol{\sigma}_{G_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\sigma}_{G_m} \end{bmatrix}_{m\times m} \quad m = i \times j$$

$$\sigma_G = \begin{bmatrix} \rho+\sigma & \rho & \cdots & \rho \\ \rho & & & \vdots \\ \vdots & & & \rho \\ \rho & \cdots & \rho & \rho+\sigma \end{bmatrix}_{j\times j} \quad \text{(compound symmetry)}$$

In this model, we assume that the covariance structure for the GEE model is compound symmetry which means the correlation with any other family members are the same because the numbers of parameters needed to be estimate are only two rather than

doi:10.6342/NTU201902662

15 needed for the unstructured covariance, for which the sample size are not enough to estimate such large number of parameters.

The GEE model was used to estimate the effect size of family aggregation, and after taking the exponent of regression coefficient, we can get the odds ratios and corresponding 95% confidence intervals, which indicates the risk of HP(+) among family members (proband) for HP(+) index compared with family members (proband) for HP(-) index.

3.2.2 Bayesian DAG (directed acyclic graphic model)

Bayesian directed acyclic graphic (DAG) model was built up to develop fully conditional distribution given the observed data and unknown quantity to estimate the effect size of family aggregation of HP infection making allowance for the heterogeneity of HP infection across different households with the random effect model (Figure 3.2).

doi:10.6342/NTU201902662

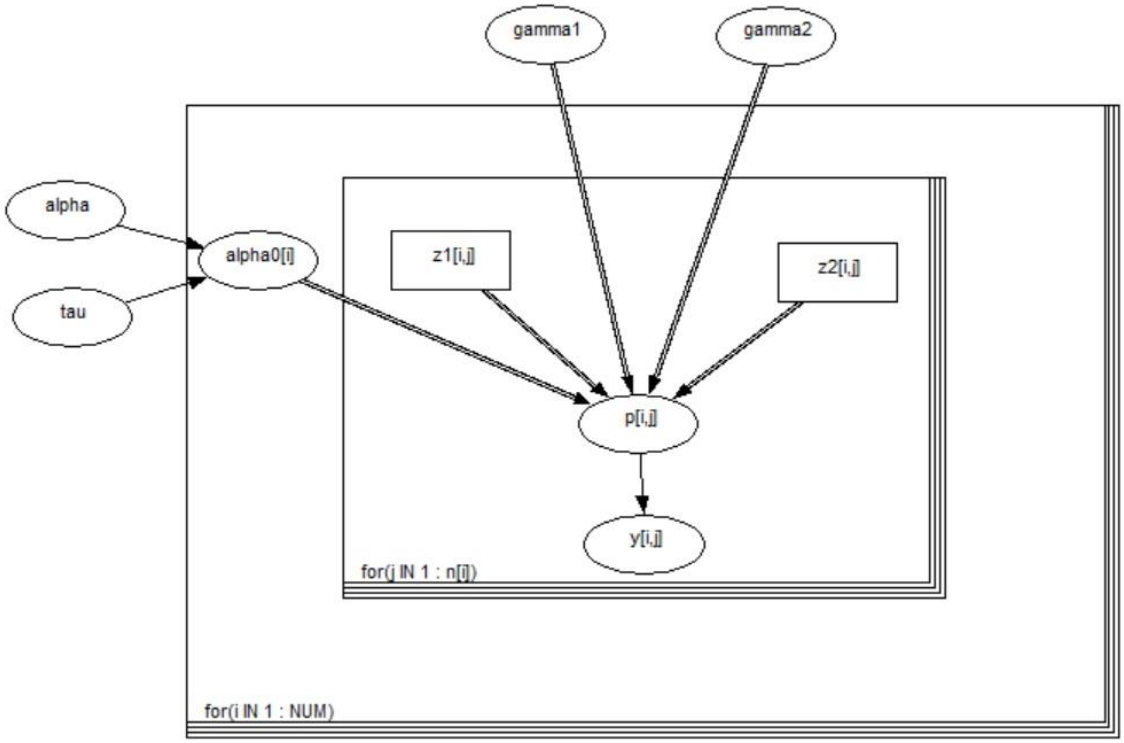


Figure 3.2 Bayesian directed acyclic graphical (DAG) model

Annotation:

a. NUM:34

b. n[i]: Number of probands in i_{th} family

c. i, j: j_{th} proband in i_{th} family

d. Y: Outcome of HP

e. P: probability of HP infection ~ Bernoulli distribution

f. Z1=1, represent matched trait of DNA finger printing; else z1=0.

g. Z2=1, represent heterogeneity trait of DNA finger printing; else z2=0.

h. Gamma1: coefficients of z1 ~ Normal distribution

i. Gamma2: coefficients of z2~ Normal distribution

j. Alpha0[i]: random intercept of i_{th} family ~ Normal distribution

k. Alpha: Mean of alpha0

l. Tau: variance of alpha0 ~ Gamma distribution

Let Y[i,j] be the binary outcome of HP for the j_{th} proband (j=1,…,n[i], where n[i] is denoted by total number of probands for subject i) derived from the i_{th} index subject (i=1,…,34), which follows Bernoulli distribution with probability of HP infection P[i,j]. To take into account the information of DNA fingerprinting (MLST/RAPD) and also the heterogeneity of HP infection across households, we used the logistic regression model with random effect to estimate the effect of family aggregation of HP infection, which can be expressed as

$$\text{logit}\,(P[i,j]) = \text{logit}\left(\text{P(Y[i,j]} = 1)\right) = \alpha_0[i] + \gamma_1 Z_1[i,j] + \gamma_2 Z_2[i,j],$$

where $Z_1[i,j] = 1$ representing the HP (+) group with the matched trait of DNA fingerprinting, $Z_2[i,j] = 1$ representing the HP (+) group with unmatched trait of DNA fingerprinting, otherwise $Z_1[i,j] = 0$ and $Z_2[i,j] = 0$ representing the reference group of HP (-) without information of DNA fingerprinting, and γ_1 and γ_2 representing respective regression coefficients. The random intercept $\alpha_0[i]$ describes the baseline risk of HP for index subject i, which follows normal distribution with mean α and variance τ.

These parameters were estimated by using the Markov Chain Monte Carlo simulation with Bayesian approach. We assigned non-informative priors with normal distribution to regression coefficients $\boldsymbol{\Gamma} = (\gamma_1, \gamma_2)$ denoted by

$$\boldsymbol{\Gamma}\sim\text{normal}(0,0.0001).$$

For random intercept, we also assigned different non-informative priors denoted by

$$\alpha_0[i]\sim normal(\alpha,\tau),$$

$$\alpha\sim\text{normal}(0,0.0001)$$

$$\tau\sim\text{gamma}(0.01,0.01),$$

therefore, the heterogeneity among index subjects can be derived as $\sigma = 1/\sqrt{\tau}$. Then, the

doi:10.6342/NTU201902662

full joint probability distribution can be given as

$$\mathrm{P}(\boldsymbol{Y}[\boldsymbol{i},\boldsymbol{j}],\boldsymbol{\Gamma},\alpha,\tau) \propto \mathrm{P}(\boldsymbol{Y}[\boldsymbol{i},\boldsymbol{j}]|\boldsymbol{\Gamma},\alpha,\tau) \times \mathrm{P}(\boldsymbol{\Gamma}) \times \mathrm{P}(\alpha) \times \mathrm{P}(\tau).$$

The priors would be updated by data (likelihood), then we can obtain the posterior samples, and also estimate the effect of family aggregation of HP infection with the corresponding 95% CI based on the marginal posterior distribution.

3.3 Matsu gastric cancer screening program

The first gastric cancer intervention program in Matsu was aimed to prevent deaths from gastric cancer as its morbidity and mortality were very high. To reach this goal, a secondary preventive program using pepsinogen as first line tool for early detection of gastric cancer was launched in 1996. During 1996-1998, among the 3,571 target residents, 2,201 subjects attended this gastric cancer screening program. Information on dietary was collect via structured questionnaire. After positive finding, the atrophic gastritis, intestinal metaplasia, and dysplasia was diagnosed by endoscopic examination. Although single screening programs have been demonstrated to facilitate pepsinogen test as first line screening and endoscopy as confirmatory methods, the program was extended to more preventive services not only H. pylori screening but also considering eradication of H. pylori in 2004. The 13C-urea breath test was used as first line to detect H. pylori. A total of 4,121 participants participated in the mass eradication program in 2004 and 1,334 participated in the second round of the program in 2008.

We can rebuild the familial pedigree of 1,637 families, which at least one person has complete information on Helicobacter pylori infection test and gastroscopy test, and other subjects that form the basis of the following analysis based on screening and house registry profiles. In order to study the role of familial aggregation in Pre-Malignance of

doi:10.6342/NTU201902662

gastric, the study design is based on the case–control family proband sampling design proposed by Chiu (Chiu et al., 2007), to compare the prevalence of pre-cancerous lesions with status of proband between normal and gastric pre-malignance.

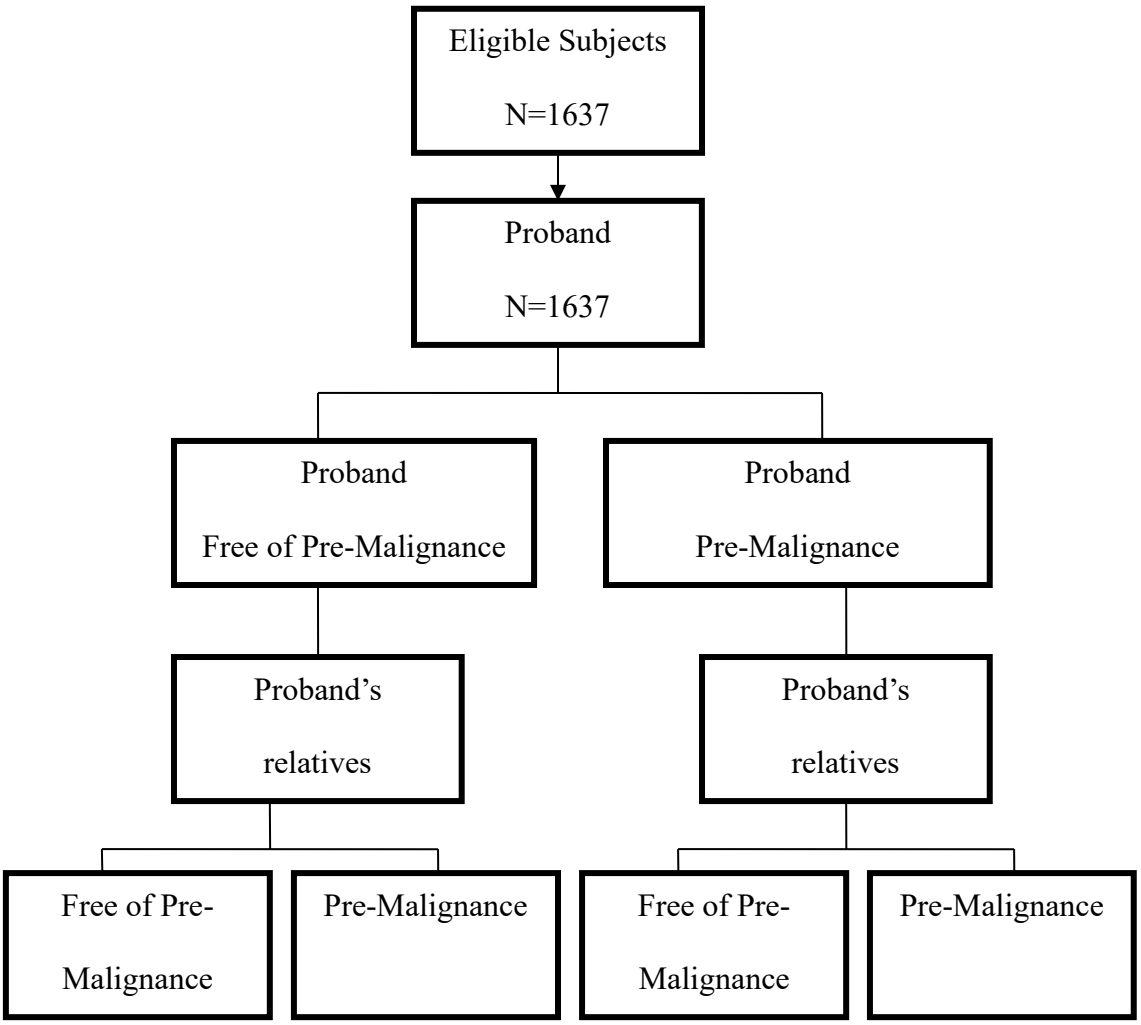


Figure 3.3 Flow chart of Matsu data

doi:10.6342/NTU201902662

3.3.1 Multi-nominal logistic regression model

In order to model how family aggregation affects the dynamic transition of HP and gastric pre-cancerous neoplasm under the context of the Correa multistate model, the multi-nominal logistic regression model can be applied.

In addition to using the binary outcome considering the infection of *Helicobacter pylori* (HP), the second dataset used for analysis were derived from the Matsu community with other correlated outcomes, including Atrophic gastritis (AG) and Intestinal metaplasia (IM). We can simultaneously assesse the effect size of familial aggregation of index status and some factors associated with infection of Helicobacter pylori (HP) and occurrence of AG and IM.

Multi-nominal logistic regression model is used when there are more than three type response categories. We used the model to assess the effect sizes of familial aggregation of HP infection, AG, and IM and effect of relevant covariate including demographic characteristics such as age and sex, smoking behavior, and dietary factors. The model is specified as follows

$$\text{logit}(\pi_j) = \log(\frac{\pi_j}{\pi_1}) = \mathrm{x}_j^T \beta_j \quad , j=2, 3, \ldots\ldots, J$$

where π_j denotes whether the Jth event occurred during the observation period.

This model firstly needs to specify the reference group, and then the other groups are compared to this reference group in order. For example, in our dataset, we can specify normal state as reference group, so we can get the three models expressed by

$$\log(\frac{\pi_{HP}}{\pi_{normal}}) = \alpha_2 + \beta_2 x$$

$$\log(\frac{\pi_{AG}}{\pi_{normal}}) = \alpha_3 + \beta_3 x$$

$$\log(\frac{\pi_{IM}}{\pi_{normal}}) = \alpha_4 + \beta_4 x$$

doi:10.6342/NTU201902662

The above three models can be regarded as three binary logistic regression models. Then we can use these models to calculate the effect size of familial aggregation of index status and other factors associated with infection of Helicobacter pylori (HP) and occurrence of AG and IM.

3.3.2 Bayesian DAG (directed acyclic graphic model)

We also build up the Bayesian directed acyclic graphic (DAG) model to develop fully conditional distribution to estimate the effect size of family aggregation of HP infection and gastric pre-cancerous neoplasm making allowance for the heterogeneity of HP infection across different households. (Figure 3.4).

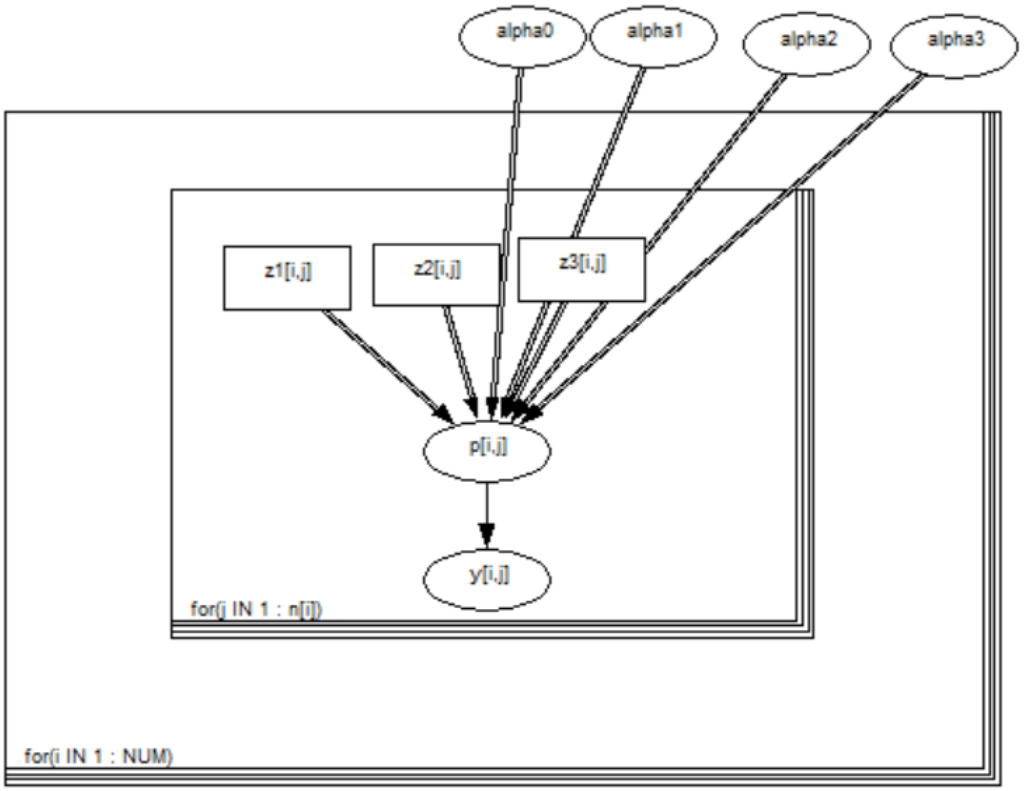


Figure 3.4

a.NUM: Number of families

a. n[i]: Number of probands in i_{th} family

doi:10.6342/NTU201902662

b. i, j: j_{th} proband in i_{th} family

c. Y: Outcome of HP

d. Z1=1, represent disease stage HP infection; else z1=0.

e. Z2=1, represent disease stage Atrophic gastritis; else z2=0.

f. Z3=1, represent disease stage Intestinal metaplasia or gastric cancer; else z3=0.

g. alpha0: intercept ~ Normal distribution

h. alpha1: coefficients of z1 ~ Normal distribution

i. alpha 2: coefficients of z2~ Normal distribution

j. alpha 3: coefficients of z3~ Normal distribution

Let Y[i,j] be the multinomial outcome of HP infection and gastric pre-cancerous neoplasm for the j_{th} proband (j=1,…,n[i], where n[i] is denoted by total number of probands for subject i) derived from the i_{th} index subject with probability of disease state P[i,j]. To take into account the information of disease state of index case, we used the multinomial logistic regression model to estimate the effect of family aggregation, which can be expressed as

$$\text{logit}\,(P[i,j]) = \text{logit}\left(\text{P(Y[i,j]} = 1)\right) = \alpha_0 + \gamma_1 Z_1[i,j] + \gamma_2 Z_2[i,j] + \gamma_3 Z_3[i,j],$$

where $Z_1[i,j] = 1$ representing the HP (+) group, $Z_2[i,j] = 1$ representing the Atrophic gastritis, $Z_3[i,j] = 1$ representing the Intestinal metaplasia or gastric cancer otherwise $Z_1[i,j] = 0, Z_2[i,j] = 0$ and $Z_3[i,j] = 0$ representing the reference group of Normal, and γ_1, γ_2 and γ_3 representing respective regression coefficients.

These parameters were estimated by using the Markov Chain Monte Carlo simulation with Bayesian approach. We assigned non-informative priors with normal distribution to regression coefficients $\boldsymbol{\Gamma} = (\gamma_1, \gamma_2, \gamma_3)$ denoted by

$$\boldsymbol{\Gamma} \sim \text{normal}(0, 0.0001).$$

For intercept, we also assigned different non-informative priors denoted by

doi:10.6342/NTU201902662

$$\alpha \sim \text{normal}(0, 0.0001)$$

$$\tau \sim \text{gamma}(0.01, 0.01),$$

Then, the full joint probability distribution can be given as

$$\mathrm{P}(\boldsymbol{Y}[\boldsymbol{i},\boldsymbol{j}], \boldsymbol{\Gamma}, \alpha, \tau) \propto \mathrm{P}(\boldsymbol{Y}[\boldsymbol{i},\boldsymbol{j}] | \boldsymbol{\Gamma}, \alpha, \tau) \times \mathrm{P}(\boldsymbol{\Gamma}) \times \mathrm{P}(\alpha).$$

The priors would be updated by data (likelihood), then we can obtain the posterior samples, and also estimate the effect of family aggregation of HP infection and gastric pre-cancerous neoplasm with the corresponding 95% CI based on the marginal posterior distribution.

3.3.3 Family aggregation with Markov Chain Model

In this chapter, we use two different methods with Markov chain model to demonstrate the situation of family aggregation. One is to derive the equilibrium distribution of Markov Chain, and the other is to generate an indicator by constructing the transition matrix defined by index case and probands. First, suppose the disease in question with the tendency of family aggregation has K states that include non-disease state, say $\mathrm{K} = 1$, characterized by integer from 1 to m and $2, \dots, \mathrm{m}$ representing a spectrum of disease.

$$\Omega = \{1, 2, \dots, \mathrm{m}\}$$

When studying family aggregation of the disease of interest, we select one index case which can be one of possible states as defined above. We are interested in whether and how these index cases are more likely to identify more probands with the states that are identical or severe compared with the index case.

To test this postulate, the conditional probability governed by a Markov chain is defined as follow

doi:10.6342/NTU201902662

$$\mathrm{P}(Index\ case = j | Proband = i) - ①\ \text{where}\ \ i, j \in \Omega$$

The short hand P_{ij} is used to represent this conditional probability throughout the text. This kind of sequence forms an example of Markov chain.

Given the equation ①, it is very interesting to assess whether and how distribution of disease status can reach the equilibrium in association with family aggregation defined through the equation ①.

Suppose we have the matrix of transition probabilities ($\boldsymbol{P}$) defined by the index case and probands given $\mathrm{K} = 4$

$$\begin{array}{c} \\ proband_i \end{array} \begin{array}{c} \begin{array}{cccc} Index & & Case_j & \end{array} \\ \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} \end{array}$$

From the limiting theorem and properties of Markov chain, if all states are communicable, which means in the same class, this Markov chain is called an irreducible chain. Irreducibility is an important property when we want to find the equilibrium distribution of a Markov chain. Here, another property of Markov Chain is required such as aperiodic, which means that the return of state doesn't follow a specific cycle of time. For example, the return of state (P_{ii}) can be any times t, $2t$, $5t$, $11t$, …etc. By contrast, a periodic chain only return at times t, $2t$, $3t$,…etc.

With these two properties (irreducibility and aperiodicity), this chain is an ergodic chain and it has a limiting distribution.

$$\lim_{n\to\infty} \boldsymbol{P}^n = \begin{pmatrix} \boldsymbol{\pi} \\ \boldsymbol{\pi} \\ \vdots \\ \boldsymbol{\pi} \end{pmatrix} = 1 \cdot \boldsymbol{\pi}$$

doi:10.6342/NTU201902662

This limiting distribution ($\boldsymbol{\pi}$) is also an equilibrium distribution independent of the initial conditions. That is to say, for any initial distribution ($\boldsymbol{p}^0$), we have

$$\lim_{n\to\infty} \boldsymbol{p}^0 \boldsymbol{P}^n = \boldsymbol{p}^0 1\boldsymbol{\pi} = \boldsymbol{\pi}$$

Besides, it is also a stationary distribution so we can easily derive the distribution by solving the following equation.

$$\boldsymbol{\pi P} = \boldsymbol{\pi}$$

After getting the distribution, we can know the long-term distribution of disease status. The distribution will demonstrate the effect of our intervention like HP eradication program. For example, if the intervention is effective, then the long-term probability of HP infection should decrease which also implies the reduction of familial aggregation.

Besides, the matrix of transition probabilities ($\boldsymbol{P}$) defined by the index case and probands can also be used to see the temporal sequence of family aggregation. Let $\sum_{c\in R} {P_{ij}}^{U^c}$ and $\sum_{l\in R} {P_{ij}}^{L^c}$ represent the summation of the transition probabilities of the upper triangle and the lower triangle. Set up the null hypothesis without family aggregation and the alternative hypothesis with family aggregation lower triangle off the diagonal matrix.

$$H_0: \sum_{c\in R} {P_{ij}}^{U^c} = \sum_{l\in R} {P_{ij}}^{L^c}$$

$$H_1: \sum_{c\in R} {P_{ij}}^{U^c} \neq \sum_{l\in R} {P_{ij}}^{L^c}$$

As family aggregation in related to the disease status, which in above governed by temporal sequence whether $\sum_{c\in R} {P_{ij}}^{U^c} > \sum_{l\in R} {P_{ij}}^{L^c}$ is also dependent on the temporal sequence of the selected index case and probands. Thus, we can use the ratio of $\sum_{c\in R} {P_{ij}}^{U^c}$ and $\sum_{l\in R} {P_{ij}}^{L^c}$ to be an indicator of familial aggregation.

doi:10.6342/NTU201902662

3.3.4 Continuous-time Discrete-state Markov Model

In this section, we have applied a continuous-time discrete-state Markov model to build up the natural history of gastric neoplasm over time. To apply a four-state Markov process, the state space is defined as follows: Ω={Stage1 (Free of HP infection and gastric neoplasm (Normal)), Stage2 (HP-infected), Stage3 (Atrophic gastritis), Stage4 (Intestinal metaplasia or gastric cancer) }. When applying the Markov model, we assumed a progressive assumption in a phase-type manner. That means that there no possibility of reversible transition in natural course of these gastric-related infection and disease. Our multistate Markov model is illustrated in Figure 1, using arrows to indicate the direction in which instantaneous transitions are possible, with the transition rates between four states denoted as $\lambda_i s$, and i=1, 2, and 3.

Thus, the transition rate matrix, $\boldsymbol{Q}$, dominating this four-state Markov process, can be written as follows

$$\mathrm{Q} = \begin{array}{c} \\ Normal \\ HP(+) \\ AG \\ IM+GC \end{array} \begin{array}{c} \begin{array}{cccc} Normal & HP(+) & AG & IM+GC \end{array} \\ \begin{bmatrix} -\lambda_1 & \lambda_1 & 0 & 0 \\ 0 & -\lambda_2 & \lambda_2 & 0 \\ 0 & 0 & -\lambda_3 & \lambda_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

By using the forward Kolmogorov differential equation, we can further derive the elements in transition probability matrix $\boldsymbol{P}$ derived from the intensity matrix $\boldsymbol{Q}$.

$$\mathbf{P}(t) = \begin{array}{c} \\ 1: Normal \\ 2: H.P(+) \\ 3: AG \\ 4: IM \end{array} \begin{array}{c} \begin{array}{cccc} Normal & H.P(+) & \mathrm{AG} & \mathrm{IM} \end{array} \\ \begin{bmatrix} P_{11}(t) & P_{12}(t) & P_{13}(t) & P_{14}(t) \\ 0 & P_{22}(t) & P_{23}(t) & P_{24}(t) \\ 0 & 0 & P_{33}(t) & P_{34}(t) \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

The residencies of Matsu can be identified as index case with the four status of gastric lesion mentioned above. The two-state (normal or with diseased status) in the conventional case-control proband study design is now extended to four status of gastric

doi:10.6342/NTU201902662

lesion mentioned above. Following the selected index case, their family members were identified as the probands. The disease status is ascertained as one of the four possibilities of the state space **Ω**.

The transition probability can be written as follows

$$P_{11} = e^{-\lambda_1 t}$$

$$P_{12} = -\left({\lambda_1}\Big/{(\lambda_1-\lambda_2)e^{\lambda_1 t}}\right) + \left({\lambda_1}\Big/{(\lambda_1-\lambda_2)e^{\lambda_2 t}}\right)$$

$$P_{13} = \left(\frac{(\lambda_1\lambda_2)}{(\lambda_1-\lambda_2)(\lambda_2-\lambda_3)} - {\frac{(\lambda_1\lambda_2)}{(\lambda_1-\lambda_3)(\lambda_2-\lambda_3)}}\Big/{e^{\lambda_1 t}}\right)$$

$$-\left({(\lambda_1\lambda_2)}\Big/{(\lambda_1-\lambda_2)(\lambda_2-\lambda_3)e^{\lambda_2 t}}\right)$$

$$+\left({(\lambda_1\lambda_2)}\Big/{(\lambda_1-\lambda_3)(\lambda_2-\lambda_3)e^{\lambda_3 t}}\right)$$

$$P_{14} = 1 + \left(-1 + \frac{\lambda_1}{(\lambda_1-\lambda_2)} - \frac{(\lambda_1\lambda_2)}{(\lambda_1-\lambda_2)(\lambda_2-\lambda_3)} + {\frac{(\lambda_1\lambda_2)}{(\lambda_1-\lambda_3)(\lambda_2-\lambda_3)}}\Big/{e^{\lambda_1 t}}\right)$$

$$+\left({\lambda_1\left(-1+\frac{\lambda_2}{(\lambda_2-\lambda_3)}\right)}\Big/{(\lambda_1-\lambda_2)e^{\lambda_2 t}}\right)$$

$$-\left({(\lambda_1\lambda_2)}\Big/{(\lambda_1-\lambda_3)(\lambda_2-\lambda_3)e^{\lambda_3 t}}\right)$$

Let the random variable X(t) denote the state of the proband at age t, X(t) ∈**Ω**. With the Markov assumption, we can construct our likelihood for each person according to their status as follows,

$$\text{Normal: } \left(\frac{P_{11}(t)}{P_{11}(t)+P_{12}(t)+P_{13}(t)+P_{14}(t)}\right),$$

$$\text{HP(+): } \left(\frac{P_{12}(t)}{P_{11}(t)+P_{12}(t)+P_{13}(t)+P_{14}(t)}\right),$$

doi:10.6342/NTU201902662

AG: $\left(\frac{P_{13}(t)}{P_{11}(t)+P_{12}(t)+P_{13}(t)+P_{14}(t)}\right)$

IM or GC: $\left(\frac{P_{14}(t)}{P_{11}(t)+P_{12}(t)+P_{13}(t)+P_{14}(t)}\right)$.

With above equations, we can define the total likelihood function by applying the multinomial distribution which is written as

$$L(\boldsymbol{\lambda}) = \prod_{i=1}^{n}\left[\frac{P_{11}(t)}{P_{11}(t)+P_{12}(t)+P_{13}(t)+P_{14}(t)}\right]^{1-y_{1i}-y_{2i}-y_{3i}}\left[\frac{P_{12}(t)}{P_{11}(t)+P_{12}(t)+P_{13}(t)+P_{14}(t)}\right]^{y_{1i}}$$
$$\left[\frac{P_{12}(t)}{P_{11}(t)+P_{12}(t)+P_{13}(t)+P_{14}(t)}\right]^{y_{2i}}\left[\frac{P_{12}(t)}{P_{11}(t)+P_{12}(t)+P_{13}(t)+P_{14}(t)}\right]^{y_{3i}}$$

where *n* represnts the number of family and y_{1i} ,y_{2i} ,and y_{3i} are the dummy variables that are reagrded as the indicators for the subject with HP infection (y_{1i}=1), atrophic gastritis (y_{2i}=1), intestinal metaplasia or gastric cancer (y_{3i}=1), or normal (y_{1i}= y_{2i}= y_{3i}=0), and $\boldsymbol{\lambda}$=(λ_1,λ_2,λ_3) corresponding to the transition rates in four-state Markov model.

Furthermore, in order to assess the contribution of familial aggregation in natural history of gastric neoplasm, we separated status of the index case in each family into 4 groups (Normal, H.pylori infected, Atrophic gastritis, Intestinal metaplasia) and incorporated it into a Markov regression model as independent variable.

By applying the Cox proportional hazards regression form, the transition rate specific for state_i is defined by

$$\lambda_i = \lambda_{i0}\exp\left(\gamma_1 Z_1 + \gamma_2 Z_2 + \gamma_3 Z_3 + \beta X\right),$$

where λ_{i0} is the baseline hazard for transition, Z_1, Z_2, and Z_3 are three dummy variables represent the disease status of index case, γ_1, γ_2, and γ_3 are the corresponding regression coefficients for Z_1, Z_2, and Z_3. Here other adjusted covariates, ***X****s* , are gender and the habit of cigarette smoking of proband, and the $\boldsymbol{\beta}$s are the corresponding

doi:10.6342/NTU201902662

regression coefficients.

doi:10.6342/NTU201902662

Chapter 4 Result

4.1 Case-control proband analysis for family aggregation of HP infection considering DNA fingerprinting matching

The data used for case-control proband study were derived from 34 index Japan pediatric patients. Among 34 families, 14 families with 3 family members, 12 families with four family members, and 8 family with five family members (Table 4.1.1).

We selected 51 fathers with positive HP and 11 fathers with negative HP as index subjects to find the positive rate of HP among proband's relatives. The positive HP rate (76.5%) was lower in proband with HP infection than 81.8% in proband without HP infection in children. Similar result was found for spouse. For mother as index, the positive HP rate (81.3%) was higher in proband with HP infection than 64.3% in proband without HP infection. For sibling as index, the positive rate (58.3%) was higher in proband with HP infection than 50% in proband without HP infection in sibling.

Table 4.1.3 shows the percentage of HP within different number of family member according to the status of proband's relateves. Compared with proband with negative HP, the percentages of HP are higher in proband with positive HP while family number over 3 in household.

The odds ratios of family aggregation for HP were estimated as 2.67 (95% CI: 0.61-11.74) and 2.67 (95% CI: 0.60-11.86) for fix effect model and random effect model,

doi:10.6342/NTU201902662

respectively (Table 4.1.4). The index subjects with matched MLST type in proband subjects are more likely to have family aggregation of HP infection. The 26 (95%CI: 2.2-310.6) of odds ratio was found in random effect model. The matched RAPD type in either index or proband subjects are also more likely to have family aggregation of HP infection. The odds ratio was 35 (95%CI: 2.9-414.7) in random effect model. The matching with MLST or RAPD type has the significant effect on HP's family aggregation (OR=35 (95%CI: 2.9-414.7)). For those unmatched types for MLST or RAPD in index or proband subjects, there is no significant family aggregation for HP infection.

Using first child as index subjects, the both of matched RAPD and MLST types have the significant effect on HP's family aggregation (Table 4.1.5). The odds ratios were 55.9 (95%CI: 4.0-1878.1) and 68.1 (95%CI: 4.9-2171.1) for RAPD and MLST type, respectively.

Table 4.1.1 Number of family members in 34 families

Number of family members	N	%
3	14	41.18
4	12	35.29
5	8	23.53
Total	34	100

doi:10.6342/NTU201902662

Table 4.1.2 Number of HP infected by different index and relative

Index status		Proband status: First child HP(+)	First child HP(-)	Father HP(+)	Father HP(-)	Mother HP(+)	Mother HP(-)	Other children HP(+)	Other children HP(-)
First child	HP(+)	-	-	24 (75.00%)	8 (25.00%)	26 (81.25%)	6 (18.75%)	14 (58.33%)	10 (41.67%)
	HP(-)	-	-	1 (50.00%)	1 (50.00%)	1 (50.00%)	1 (50.00%)	2 (50.00%)	2 (50.00%)
Father	HP(+)	24 (96.00%)	1 (4.00%)	-	-	18 (72.00%)	7 (28.00%)	15 (65.22%)	8 (34.78%)
	HP(-)	8 (88.89%)	1 (11.11%)	-	-	9 (100.00%)	0 (0.00%)	1 (20.00%)	4 (80.00%)
Mother	HP(+)	26 (96.30%)	1 (3.70%)	18 (66.67%)	9 (33.33%)	-	-	13 (54.17%)	11 (45.83%)
	HP(-)	6 (85.71%)	1 (14.29%)	7 (100.00%)	0 (0.00%)	-	-	3 (75.00%)	1 (25.00%)

doi:10.6342/NTU201902662

Table 4.1.3 Number of relatives with HP (+) by household number

		Number of relatives with HP(+)				
Index	**Status**	**0**	**1**	**2**	**3**	**4**
			Household number=3			
Mother	HP(-)	0 (0%)	0 (0%)	4 (100%)	-	-
	HP(+)	0 (0%)	6 (60%)	4 (40%)	-	-
Father	HP(-)	0 (0%)	0 (0%)	6 (100%)	-	-
	HP(+)	0 (0%)	4 (50%)	4 (50%)	-	-
First child	HP(-)	0	0	0	-	-
	HP(+)	0 (0%)	14 (71%)	4 (29%)	-	-
Other children	HP(-)	-	-	-	-	-
	HP(+)	-	-	-	-	-
			Household number=4			
Mother	HP(-)	0 (0%)	0 (0%)	0 (0%)	2 (100%)	-
	HP(+)	0 (0%)	1 (10%)	2 (20%)	8 (30%)	-
Father	HP(-)	0 (0%)	0 (0%)	1 (100%)	0 (0%)	-
	HP(+)	0 (0%)	0 (0%)	4 (36%)	8 (64%)	-
First child	HP(-)	0	0	0	0	-
	HP(+)	0 (0%)	1 (8%)	4 (33%)	8 (58%)	-
Other children	HP(-)	0 (0%)	0 (0%)	1 (33%)	2 (67%)	-
	HP(+)	0 (0%)	0 (0%)	2 (22%)	7 (78%)	-
			Household number=5			
Mother	HP(-)	0 (0%)	0 (0%)	1 (100%)	0 (0%)	0 (0%)
	HP(+)	0 (0%)	2 (29%)	2 (29%)	1 (13%)	2 (29%)
Father	HP(-)	0 (0%)	0 (0%)	2 (100%)	0 (0%)	0 (0%)
	HP(+)	0 (0%)	1 (17%)	2 (33%)	1 (17%)	2 (33%)
First child	HP(-)	0 (0%)	0 (0%)	2 (100%)	0 (0%)	0 (0%)
	HP(+)	0 (0%)	1 (17%)	2 (33%)	1 (17%)	2 (33%)
Other children	HP(-)	0 (0%)	2 (22%)	2 (22%)	4 (45%)	1 (11%)
	HP(+)	0 (0%)	2 (29%)	0 (0%)	0 (0%)	5 (71%)

doi:10.6342/NTU201902662

Table 4.1.4 Estimates of relative risk from HP family aggregation with and without random effect model

Variable	OR (95% C.I.)			
	Model 0 (Crude)	***Model 1 (MLST)***	***Model 2 (RAPD)***	***Model 3 (MLST or RAPD)***
			Logistic regression model	
Index status (vs. normal)				
HP (+)	2.67 (0.61-11.74)	-	-	-
HP (+), same trait	-	26.00 (2.21-305.26)	35.00 (3.00-407.68)	35.00 (3.00-407.68)
HP (+), different trait	-	1.65 (0.37-7.40)	1.26 (0.28-5.71)	1.26 (0.28-5.71)
AIC	118.22	106.48	97.62	97.62
			Generalized estimating equation model	
Index status (vs. normal)				
HP (+)	2.67 (1.64-4.29)	-	-	-
HP (+), same trait	-	38.16 (3.43-424.09)	47.90 (6.89-332.95)	47.90 (6.89-332.95)
HP (+), different trait	-	1.44 (0.83-2.52)	1.02 (0.57-1.85)	1.02 (0.57-1.85)
AIC	116.25	105.29	96.06	96.06

doi:10.6342/NTU201902662

Table 4.1.5 Estimates of Bayesian directed acyclic graphical (DAG) model

Variable	OR (95% CI)		
	Model 1 (MLST)	***Model 2 (RAPD)***	***Model 3 (MLST or RAPD)***
Index status (vs. normal)			
HP (+), same trait	55.92 (3.99-1878.7)	68.10 (4.85-2171.12)	68.10 (4.85-2171.12)
HP (+), different trait	2.00 (0.37-14.06)	1.32 (0.22-7.49)	1.32 (0.22-7.49)
DIC	108.73	98.66	98.66

doi:10.6342/NTU201902662

4.2 Case-control proband analysis for family aggregation of HP infection and gastric pre-malignance

4.2.1 Family aggregation analysis based on full family data

The data used for family aggregation analysis were derived from 1637 index Japan pediatric patients. Among 1637 families, 689 families with 2 family members, 424 families with 3 family members, 271 families with 4 family members, and other 253 families with five or more family members (Table 4.2.1.1). Table 4.2.1.2-6 shows the percentage of HP, AG, or IM/GC according to the disease status of proband across different family members in household by the disease status of index subjects.

The odds ratios for having HP infection among family members for cases vs. control probands were 1.63 (95%CI: 1.36-1.96), 1.08 (95%CI: 0.87-1.35), and 1.51 (95%CI: 1.10-2.06) for index subjects with HP infection, AG, and IM, respectively (Table 4.2.1.7). The family aggregation of HP infection with adjusting other covariates was noted with statistically significance for both of index subjects with HP infection (aOR: 1.61 (95%CI:1.33-1.95) and IM (aOR: 1.43 (95%CI:1.04-1.97). There is negative association between smoking and proband with HP infection (aOR: 0.65 (95%CI:0.53-0.79).

In terms of other pre-gastric cancer lesion, the adjusted odds ratios of family aggregation for AG were 1.14 (95% CI: 0.9-1.43), 1.39 (95% CI: 1.09-1.77), and 1.59 (95% CI: 1.11-2.28) for index subjects with HP infection, AG, and IM, respectively. The positive association between smoking and proband with AG was found.(aOR: 1.38 (95%CI:1.08-1.77). The odds ratios for having IM among family members for cases vs. control probands were 1.51 (95%CI: 1.11-2.07), 1.51 (95%CI: 1.06-2.14), and 2.04 (95%CI: 1.28-3.27) for index subjects with HP infection, AG, and IM, respectively. After controlling for other confounding factors of probands, the tendency to family aggregation

doi:10.6342/NTU201902662

still remained statistically significant.

4.2.2 Family aggregation analysis based on sampling family data

Table 4.2.1.8 shows the family aggregation of HP infection and gastric pre-malignancy based on sampling data. The tendency to family aggregation of HP infection are similar as the results in full data. The results for family aggregation of AG are also similar but there is no statistical significance due to the smaller sample sizes. The odds ratio for having IM among family members for cases vs. control probands were 2.17 (95%CI: 1.12-4.18) for index subjects with IM.

4.2.3 Family aggregation analysis based on sampling family data in 1996

Using data from early screening period in 1996, the adjusted odds ratios for having HP infection among family members for cases vs. control probands were 1.41 (95%CI: 1.05-1.91), 1.44 (95%CI: 0.90-2.31), and 1.71 (95%CI: 0.96-3.14) for index subjects with HP infection, AG, and IM, respectively (Table 4.2.1.9). The adjusted odds ratios of family aggregation for AG were 1.72 (95% CI: 1.08-2.73), 3.82 (95% CI: 2.12-6.88), and 3.76 (95% CI: 1.81-7.78) for index subjects with HP infection, AG, and IM, respectively. The adjusted odds ratios of family aggregation for IM were 1.75 (95% CI: 1.02-2.99), 3.05 (95% CI: 1.50-6.21), and 2.21 (95% CI: 0.86-5.69) for index subjects with HP infection, AG, and IM, respectively.

4.2.4 Family aggregation analysis based on sampling family data in 2004

Using data from second screening period in 2004, the adjusted odds ratios for having HP infection among family members for cases vs. control probands were 4.43 (95%CI:

doi:10.6342/NTU201902662

2.28-8.59), 1.37 (95%CI: 0.80-2.35), and 1.45 (95%CI: 0.79-2.66) for index subjects with HP infection, AG, and IM, respectively (Table 4.2.1.10). Female probands are more likely to have family aggregation of HP (aOR=2.17 (95%CI: 1.22-3.87). The adjusted odds ratios of family aggregation for AG were 1.40 (95% CI: 0.82-2.39), 1.34 (95% CI: 0.96-1.87), and 1.29 (95% CI: 0.88-1.91) for index subjects with HP infection, AG, and IM, respectively. The adjusted odds ratios of family aggregation for IM were 1.89 (95% CI: 1.03-3.47), 1.47 (95% CI: 0.99-2.20), and 1.96 (95% CI: 1.26-3.05) for index subjects with HP infection, AG, and IM, respectively.

4.2.5 Family aggregation analysis based on sampling family data in 2008

Using data from HP eradicated period in 2008, the adjusted odds ratios for having HP infection among family members for cases vs. control probands were 3.80 (95%CI: 1.96-7.37), 1.44 (95%CI: 0.84-2.48), and 1.18 (95%CI: 0.66-2.11) for index subjects with HP infection, AG, and IM, respectively (Table 4.2.1.11). Female probands are more likely to have family aggregation of HP (aOR=2.12 (95%CI: 1.19-3.78). Probands with diet factor 1 are more likely to have family aggregation of HP (aOR=1.15 (95%CI: 1.02-1.29). But diet factor 2 is less likely to have family aggregation of HP (aOR=0.71(95%CI: 0.57-0.89)). The adjusted odds ratios of family aggregation for AG were 1.41 (95% CI: 0.82-2.43), 1.38 (95% CI: 0.96-1.98), and 1.27 (95% CI: 0.87-1.84) for index subjects with HP infection, AG, and IM, respectively. The adjusted odds ratios of family aggregation for IM were 1.46 (95% CI: 0.81-2.64), 1.35 (95% CI: 0.91-1.99), and 1.72 (95% CI: 1.16-2.55) for index subjects with HP infection, AG, and IM, respectively.

4.2.6 Family aggregation analysis based on sampling family data with

doi:10.6342/NTU201902662

three periods

Table 4.2.1.12 shows the period effect on family aggregation of HP infection and gastric pre-malignancy. The later periods are less likely to have family aggregation of HP infection. The odds ratio was 0.36 (95% CI: 0.27-0.47) for year of 2004 against year of 1996. The similar finding for year of 2008 (OR=0.35 (95%CI: 0.27-0.46). For the gastric pre-malignancy, the later periods are more likely to have family aggregation for both of AG and IM. The odds ratios of AG were 3.45 (95% CI: 2.67-4.45) and 3.03 (95% CI: 2.34-3.93) for year of 2004 and year of 2008. The odds ratios of IM were 4.41 (95% CI: 3.27-5.95) and 5.28 (95% CI: 3.92-7.11) for year of 2004 and year of 2008. 4.2.1 Case-control proband study

doi:10.6342/NTU201902662

Table 4.2.1.1 Number of family members in 1637 families

Number of family members	N	%
2	689	42.09
3	424	25.9
4	271	16.55
5	125	7.64
6	75	4.58
7	35	2.14
8	14	0.86
9	3	0.18
10	1	0.06
Total	1637	100

doi:10.6342/NTU201902662

Table 4.2.1.2 Number of relatives with different statuses of disease by household number (full data set)

(a) Number of relatives with HP infection

Number of relatives with HP infection										
Index	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Household number=2										
Free of HP infection, AG, and IM/GC	197 (73.23%)	72 (26.77%)	-	-	-	-	-	-	-	-
HP infection	138 (58.97%)	96 (41.03%)	-	-	-	-	-	-	-	-
AG	108 (78.83%)	29 (21.17%)	-	-	-	-	-	-	-	-
IM/GC	28 (57.14%)	21 (42.86%)	-	-	-	-	-	-	-	-
Household number=3										
Free of HP infection, AG, and IM/GC	119 (74.84%)	0 (0%)	40 (25.16%)	-	-	-	-	-	-	-
HP infection	82 (55.78%)	0 (0%)	65 (44.22%)	-	-	-	-	-	-	-
AG	61 (70.11%)	0 (0%)	26 (29.89%)	-	-	-	-	-	-	-
IM/GC	20 (64.52%)	0 (0%)	11 (35.48%)	-	-	-	-	-	-	-
Household number=4										

doi:10.6342/NTU201902662

Free of HP infection, AG, and IM/GC/GC	54 (65.85%)	0 (0%)	0 (0%)	28 (34.15%)	-	-	-	-	-	-
HP infection	57 (58.76%)	0 (0%)	0 (0%)	40 (41.24%)	-	-	-	-	-	-
AG	41 (63.08%)	0 (0%)	0 (0%)	24 (36.92%)	-	-	-	-	-	-
IM/GC	16 (59.26%)	0 (0%)	0 (0%)	11 (40.74%)	-	-	-	-	-	-
Household number=5										
Free of HP infection, AG, and IM/GC	32 (62.75%)	0 (0%)	0 (0%)	0 (0%)	19 (37.25%)	-	-	-	-	-
HP infection	29 (58%)	0 (0%)	0 (0%)	0 (0%)	21 (42%)	-	-	-	-	-
AG	11 (68.75%)	0 (0%)	0 (0%)	0 (0%)	5 (31.25%)	-	-	-	-	-
IM/GC	8 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	-	-	-	-	-
Household number≥6										
Free of HP infection, AG, and IM/GC	32 (66.67%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	9 (18.75%)	6 (12.5%)	1 (2.08%)	0 (0%)	0 (0%)
HP infection	21 (65.63%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	8 (25%)	1 (3.13%)	2 (6.25%)	0 (0%)	0 (0%)
AG	28 (68.29%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	8 (19.51%)	2 (4.88%)	1 (2.44%)	1 (2.44%)	1 (2.44%)
IM/GC	6 (85.71%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (14.29%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

doi:10.6342/NTU201902662

(b) Number of relatives with AG

Index	Number of relatives with AG							
	0	1	2	3	4	5	6	7
Household number=2								
Free of HP infection, AG, and IM/GC	222 (82.53%)	47 (17.47%)	-	-	-	-	-	-
HP infection	207 (88.46%)	27 (11.54%)	-	-	-	-	-	-
AG	92 (67.15%)	45 (32.85%)	-	-	-	-	-	-
IM/GC	36 (73.47%)	13 (26.53%)	-	-	-	-	-	-
Household number=3								
Free of HP infection, AG, and IM/GC	124 (77.99%)	0 (0%)	35 (22.01%)	-	-	-	-	-
HP infection	119 (80.95%)	0 (0%)	28 (19.05%)	-	-	-	-	-
AG	60 (68.97%)	0 (0%)	27 (31.03%)	-	-	-	-	-
IM/GC	28 (90.32%)	0 (0%)	3 (9.68%)	-	-	-	-	-
Household number=4								
Free of HP infection, AG, and IM/GC	63 (76.83%)	0 (0%)	0 (0%)	19 (23.17%)	-	-	-	-
HP infection	73 (75.26%)	0 (0%)	0 (0%)	24 (24.74%)	-	-	-	-

doi:10.6342/NTU201902662

AG	50 (76.92%)	0 (0%)	0 (0%)	15 (23.08%)	-	-	-	-
IM/GC	20 (74.07%)	0 (0%)	0 (0%)	7 (25.93%)	-	-	-	-
Household number=5								
Free of HP infection, AG, and IM/GC	43 (84.31%)	0 (0%)	0 (0%)	0 (0%)	8 (15.69%)	-	-	-
HP infection	41 (82%)	0 (0%)	0 (0%)	0 (0%)	9 (18%)	-	-	-
AG	14 (87.5%)	0 (0%)	0 (0%)	0 (0%)	2 (12.5%)	-	-	-
IM/GC	6 (75%)	0 (0%)	0 (0%)	0 (0%)	2 (25%)	-	-	-
Household number≥6								
Free of HP infection, AG, and IM/GC	35 (72.92%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	8 (16.67%)	4 (8.33%)	1 (2.08%)
HP infection	26 (81.25%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	5 (15.63%)	0 (0%)	1 (3.13%)
AG	33 (80.49%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (7.32%)	2 (4.88%)	3 (7.32%)
IM/GC	4 (57.14%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (14.29%)	2 (28.57%)	0 (0%)

(c) Number of relatives with IM/GC

Number of relatives with IM/GC

doi:10.6342/NTU201902662

Index	0	1	2	3	4	5	6	7	8
Household number=2									
Free of HP infection, AG, and IM/GC	248 (92.19%)	21 (7.81%)	-	-	-	-	-	-	-
HP infection	206 (88.03%)	28 (11.97%)	-	-	-	-	-	-	-
AG	123 (89.78%)	14 (10.22%)	-	-	-	-	-	-	-
IM/GC	47 (95.92%)	2 (4.08%)	-	-	-	-	-	-	-
Household number=3									
Free of HP infection, AG, and IM/GC	148 (93.08%)	0 (0%)	11 (6.92%)	-	-	-	-	-	-
HP infection	136 (92.52%)	0 (0%)	11 (7.48%)	-	-	-	-	-	-
AG	80 (91.95%)	0 (0%)	7 (8.05%)	-	-	-	-	-	-
IM/GC	29 (93.55%)	0 (0%)	2 (6.45%)	-	-	-	-	-	-
Household number=4									
Free of HP infection, AG, and IM/GC	77 (93.9%)	0 (0%)	0 (0%)	5 (6.1%)	-	-	-	-	-
HP infection	90 (92.78%)	0 (0%)	0 (0%)	7 (7.22%)	-	-	-	-	-
AG	58 (89.23%)	0 (0%)	0 (0%)	7 (10.77%)	-	-	-	-	-

doi:10.6342/NTU201902662

IM/GC	24 (88.89%)	0 (0%)	0 (0%)	3 (11.11%)	-	-	-	-	-
Household number=5									
Free of HP infection, AG, and IM/GC/GC	45 (88.24%)	0 (0%)	0 (0%)	0 (0%)	6 (11.76%)	-	-	-	-
HP infection	48 (96%)	0 (0%)	0 (0%)	0 (0%)	2 (4%)	-	-	-	-
AG	15 (93.75%)	0 (0%)	0 (0%)	0 (0%)	1 (6.25%)	-	-	-	-
IM/GC	8 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	-	-	-	-
Household number≥6									
Free of HP infection, AG, and IM/GC	45 (93.75%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (4.17%)	0 (0%)	0 (0%)	1 (2.08%)
HP infection	30 (93.75%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (6.25%)	0 (0%)	0 (0%)	0 (0%)
AG	37 (90.24%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (2.44%)	3 (7.32%)	0 (0%)	0 (0%)
IM/GC	6 (85.71%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (14.29%)	0 (0%)	0 (0%)

doi:10.6342/NTU201902662

Table 4.2.1.3 Number of relatives with different statuses of disease by household number (index case from random sample in full cohort)

(a) Number of relatives with HP infection

Number of relatives with HP infection					
Index	**0**	**1**	**2**	**3**	**4**
Household number=2					
Free of HP infection, AG, and IM/GC	112 (67.47%)	54 (32.53%)	-	-	-
HP infection	66 (52.8%)	59 (47.2%)	-	-	-
AG	66 (75.86%)	21 (24.14%)	-	-	-
IM/GC	24 (57.14%)	18 (42.86%)	-	-	-
Household number=3					
Free of HP infection, AG, and IM/GC	42 (52.5%)	24 (30%)	14 (17.5%)	-	-
HP infection	27 (37.5%)	33 (45.83%)	12 (16.67%)	-	-
AG	30 (57.69%)	16 (30.77%)	6 (11.54%)	-	-
IM/GC	2 (14.29%)	10 (71.43%)	2 (14.29%)	-	-
Household number=4					
Free of HP infection, AG, and IM/GC	11 (26.19%)	19 (45.24%)	7 (16.67%)	5 (11.9%)	-
HP infection	13 (27.66%)	21 (44.68%)	11 (23.4%)	2 (4.26%)	-
AG	13 (39.39%)	8 (24.24%)	12 (36.36%)	0 (0%)	-
IM/GC	4 (33.33%)	3 (25%)	4 (33.33%)	1 (8.33%)	-
Household number=5					
Free of HP infection, AG, and IM/GC	5 (22.73%)	8 (36.36%)	6 (27.27%)	3 (13.64%)	0 (0%)
HP infection	4 (20%)	8 (40%)	3 (15%)	5 (25%)	0 (0%)
AG	4 (44.44%)	4 (44.44%)	1 (11.11%)	0 (0%)	0 (0%)
IM/GC	1 (25%)	1 (25%)	2 (50%)	0 (0%)	0 (0%)
Household number≥6					
Free of HP infection, AG, and IM/GC	5 (20.83%)	8 (33.33%)	5 (20.83%)	4 (16.67%)	2 (8.33%)
HP infection	2 (9.52%)	4 (19.05%)	9 (42.86%)	4 (19.05%)	2 (9.52%)
AG	3 (30%)	3 (30%)	2 (20%)	2 (20%)	0 (0%)
IM/GC	0 (0%)	1 (33.33%)	1 (33.33%)	0 (0%)	1 (33.33%)

doi:10.6342/NTU201902662

(b) Number of relatives with AG

Number of relatives with AG					
Index	**0**	**1**	**2**	**3**	**4**
Household number=2					
Free of HP infection, AG, and IM/GC	135 (81.33%)	31 (18.67%)	-	-	-
HP infection	111 (88.8%)	14 (11.2%)	-	-	-
AG	60 (68.97%)	27 (31.03%)	-	-	-
IM/GC	34 (80.95%)	8 (19.05%)	-	-	-
Household number=3					
Free of HP infection, AG, and IM/GC	51 (63.75%)	25 (31.25%)	4 (5%)	-	-
HP infection	49 (68.06%)	21 (29.17%)	2 (2.78%)	-	-
AG	30 (57.69%)	18 (34.62%)	4 (7.69%)	-	-
IM/GC	11 (78.57%)	2 (14.29%)	1 (7.14%)	-	-
Household number=4					
Free of HP infection, AG, and IM/GC	22 (52.38%)	15 (35.71%)	4 (9.52%)	1 (2.38%)	-
HP infection	16 (34.04%)	18 (38.3%)	13 (27.66%)	0 (0%)	-
AG	11 (33.33%)	21 (63.64%)	1 (3.03%)	0 (0%)	-
IM/GC	7 (58.33%)	3 (25%)	2 (16.67%)	0 (0%)	-
Household number=5					
Free of HP infection, AG, and IM/GC	10 (45.45%)	6 (27.27%)	5 (22.73%)	1 (4.55%)	0 (0%)
HP infection	10 (50%)	8 (40%)	2 (10%)	0 (0%)	0 (0%)
AG	0 (0%)	6 (66.67%)	2 (22.22%)	1 (11.11%)	0 (0%)
IM/GC	3 (75%)	1 (25%)	0 (0%)	0 (0%)	0 (0%)
Household number≥6					
Free of HP infection, AG, and IM/GC	4 (16.67%)	6 (25%)	9 (37.5%)	4 (16.67%)	1 (4.17%)
HP infection	7 (33.33%)	5 (23.81%)	6 (28.57%)	3 (14.29%)	0 (0%)
AG	0 (0%)	5 (50%)	4 (40%)	1 (10%)	0 (0%)
IM/GC	1 (33.33%)	0 (0%)	1 (33.33%)	1 (33.33%)	0 (0%)

doi:10.6342/NTU201902662

(c) Number of relatives with IM/GC

Number of relatives with IM/GC				
Index	**0**	**1**	**2**	**3**
Household number=2				
Free of HP infection, AG, and IM/GC	160 (96.39%)	6 (3.61%)	-	-
HP infection	116 (92.8%)	9 (7.2%)	-	-
AG	78 (89.66%)	9 (10.34%)	-	-
IM/GC	40 (95.24%)	2 (4.76%)	-	-
Household number=3				
Free of HP infection, AG, and IM/GC	70 (87.5%)	10 (12.5%)	0 (0%)	-
HP infection	60 (83.33%)	12 (16.67%)	0 (0%)	-
AG	45 (86.54%)	7 (13.46%)	0 (0%)	-
IM/GC	12 (85.71%)	2 (14.29%)	0 (0%)	-
Household number=4				
Free of HP infection, AG, and IM/GC	37 (88.1%)	5 (11.9%)	0 (0%)	0 (0%)
HP infection	38 (80.85%)	8 (17.02%)	1 (2.13%)	0 (0%)
AG	25 (75.76%)	6 (18.18%)	2 (6.06%)	0 (0%)
IM/GC	6 (50%)	4 (33.33%)	2 (16.67%)	0 (0%)
Household number=5				
Free of HP infection, AG, and IM/GC	12 (54.55%)	8 (36.36%)	2 (9.09%)	0 (0%)
HP infection	11 (55%)	7 (35%)	2 (10%)	0 (0%)
AG	4 (44.44%)	4 (44.44%)	1 (11.11%)	0 (0%)
IM/GC	3 (75%)	1 (25%)	0 (0%)	0 (0%)
Household number≥6				
Free of HP infection, AG, and IM/GC	16 (66.67%)	5 (20.83%)	2 (8.33%)	1 (4.17%)
HP infection	16 (76.19%)	2 (9.52%)	3 (14.29%)	0 (0%)
AG	8 (80%)	0 (0%)	2 (20%)	0 (0%)
IM/GC	0 (0%)	3 (100%)	0 (0%)	0 (0%)

doi:10.6342/NTU201902662

Table 4.2.1.4 Number of relatives with different statuses of disease by household number (index case from random sample in 1996)

(a) Number of relatives with HP infection

Number of relatives with HP infection					
Index	**0**	**1**	**2**	**3**	**4**
Household number=2					
Free of HP infection, AG, and IM/GC	94 (61.44%)	59 (38.56%)	-	-	-
HP infection	60 (46.51%)	69 (53.49%)	-	-	-
AG	25 (65.79%)	13 (34.21%)	-	-	-
IM/GC	14 (53.85%)	12 (46.15%)	-	-	-
Household number=3					
Free of HP infection, AG, and IM/GC	28 (46.67%)	24 (40%)	8 (13.33%)	-	-
HP infection	21 (29.58%)	35 (49.3%)	15 (21.13%)	-	-
AG	9 (36%)	11 (44%)	5 (20%)	-	-
IM/GC	5 (41.67%)	5 (41.67%)	2 (16.67%)	-	-
Household number=4					
Free of HP infection, AG, and IM/GC	5 (17.86%)	10 (35.71%)	7 (25%)	6 (21.43%)	-
HP infection	6 (20%)	9 (30%)	14 (46.67%)	1 (3.33%)	-
AG	4 (50%)	1 (12.5%)	2 (25%)	1 (12.5%)	-
IM/GC	1 (50%)	0 (0%)	1 (50%)	0 (0%)	-
Household number=5					
Free of HP infection, AG, and IM/GC	3 (14.29%)	6 (28.57%)	10 (47.62%)	2 (9.52%)	0 (0%)
HP infection	2 (13.33%)	9 (60%)	3 (20%)	1 (6.67%)	0 (0%)
AG	1 (33.33%)	0 (0%)	1 (33.33%)	1 (33.33%)	0 (0%)
IM/GC	0 (0%)	2 (66.67%)	1 (33.33%)	0 (0%)	0 (0%)
Household number≥6					
Free of HP infection, AG, and IM/GC	0 (0%)	1 (25%)	1 (25%)	1 (25%)	1 (25%)
HP infection	0 (0%)	1 (16.67%)	3 (50%)	2 (33.33%)	0 (0%)
AG	0 (0%)	1 (33.33%)	1 (33.33%)	1 (33.33%)	0 (0%)
IM/GC	0 (0%)	1 (50%)	0 (0%)	0 (0%)	1 (50%)

doi:10.6342/NTU201902662

(b) Number of relatives with AG

Number of relatives with AG					
Index	**0**	**1**	**2**	**3**	**4**
Household number=2					
Free of HP infection, AG, and IM/GC	138 (90.2%)	15 (9.8%)	-	-	-
HP infection	121 (93.8%)	8 (6.2%)	-	-	-
AG	27 (71.05%)	11 (28.95%)	-	-	-
IM/GC	23 (88.46%)	3 (11.54%)	-	-	-
Household number=3					
Free of HP infection, AG, and IM/GC	52 (86.67%)	8 (13.33%)	0 (0%)	-	-
HP infection	50 (70.42%)	21 (29.58%)	0 (0%)	-	-
AG	15 (60%)	10 (40%)	0 (0%)	-	-
IM/GC	6 (50%)	4 (33.33%)	2 (16.67%)	-	-
Household number=4					
Free of HP infection, AG, and IM/GC	26 (92.86%)	2 (7.14%)	0 (0%)	0 (0%)	-
HP infection	18 (60%)	11 (36.67%)	1 (3.33%)	0 (0%)	-
AG	5 (62.5%)	2 (25%)	1 (12.5%)	0 (0%)	-
IM/GC	1 (50%)	1 (50%)	0 (0%)	0 (0%)	-
Household number=5					
Free of HP infection, AG, and IM/GC	12 (57.14%)	6 (28.57%)	3 (14.29%)	0 (0%)	0 (0%)
HP infection	10 (66.67%)	5 (33.33%)	0 (0%)	0 (0%)	0 (0%)
AG	2 (66.67%)	0 (0%)	1 (33.33%)	0 (0%)	0 (0%)
IM/GC	1 (33.33%)	2 (66.67%)	0 (0%)	0 (0%)	0 (0%)
Household number≥6					
Free of HP infection, AG, and IM/GC	0 (0%)	4 (100%)	0 (0%)	0 (0%)	0 (0%)
HP infection	0 (0%)	5 (83.33%)	1 (16.67%)	0 (0%)	0 (0%)
AG	0 (0%)	3 (100%)	0 (0%)	0 (0%)	0 (0%)
IM/GC	1 (50%)	0 (0%)	1 (50%)	0 (0%)	0 (0%)

doi:10.6342/NTU201902662

(c) Number of relatives with IM/GC

Number of relatives with IM/GC					
Index	**0**	**1**	**2**	**3**	**4**
Household number=2					
Free of HP infection, AG, and IM/GC	146 (95.42%)	7 (4.58%)	-	-	-
HP infection	117 (90.7%)	12 (9.3%)	-	-	-
AG	36 (94.74%)	2 (5.26%)	-	-	-
IM/GC	23 (88.46%)	3 (11.54%)	-	-	-
Household number=3					
Free of HP infection, AG, and IM/GC	55 (91.67%)	4 (6.67%)	1 (1.67%)	-	-
HP infection	64 (90.14%)	7 (9.86%)	0 (0%)	-	-
AG	19 (76%)	6 (24%)	0 (0%)	-	-
IM/GC	12 (100%)	0 (0%)	0 (0%)	-	-
Household number=4					
Free of HP infection, AG, and IM/GC	25 (89.29%)	3 (10.71%)	0 (0%)	0 (0%)	-
HP infection	24 (80%)	5 (16.67%)	1 (3.33%)	0 (0%)	-
AG	3 (37.5%)	4 (50%)	1 (12.5%)	0 (0%)	-
IM/GC	2 (100%)	0 (0%)	0 (0%)	0 (0%)	-
Household number=5					
Free of HP infection, AG, and IM/GC	12 (57.14%)	6 (28.57%)	3 (14.29%)	0 (0%)	0 (0%)
HP infection	8 (53.33%)	3 (20%)	4 (26.67%)	0 (0%)	0 (0%)
AG	3 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
IM/GC	1 (33.33%)	2 (66.67%)	0 (0%)	0 (0%)	0 (0%)
Household number≥6					
Free of HP infection, AG, and IM/GC	3 (75%)	1 (25%)	0 (0%)	0 (0%)	0 (0%)
HP infection	6 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
AG	2 (66.67%)	0 (0%)	1 (33.33%)	0 (0%)	0 (0%)
IM/GC	0 (0%)	2 (100%)	0 (0%)	0 (0%)	0 (0%)

doi:10.6342/NTU201902662

Table 4.2.1.5 Number of relatives with different statuses of disease by household number (index case from random sample in 2004)

(a) Number of relatives with HP infection

Number of relatives with HP infection				
Index	**0**	**1**	**2**	**3**
Household number=2				
Free of HP infection, AG, and IM/GC	84 (87.5%)	12 (12.5%)	-	-
HP infection	28 (75.68%)	9 (24.32%)	-	-
AG	127 (89.44%)	15 (10.56%)	-	-
IM/GC	68 (91.89%)	6 (8.11%)	-	-
Household number=3				
Free of HP infection, AG, and IM/GC	44 (95.65%)	2 (4.35%)	0 (0%)	-
HP infection	10 (76.92%)	3 (23.08%)	0 (0%)	-
AG	64 (87.67%)	9 (12.33%)	0 (0%)	-
IM/GC	31 (81.58%)	7 (18.42%)	0 (0%)	-
Household number=4				
Free of HP infection, AG, and IM/GC	13 (54.17%)	10 (41.67%)	1 (4.17%)	0 (0%)
HP infection	6 (42.86%)	5 (35.71%)	1 (7.14%)	2 (14.29%)
AG	28 (62.22%)	15 (33.33%)	2 (4.44%)	0 (0%)
IM/GC	14 (73.68%)	4 (21.05%)	1 (5.26%)	0 (0%)
Household number=5				
Free of HP infection, AG, and IM/GC	8 (88.89%)	1 (11.11%)	0 (0%)	0 (0%)
HP infection	2 (33.33%)	1 (16.67%)	3 (50%)	0 (0%)
AG	7 (77.78%)	1 (11.11%)	0 (0%)	1 (11.11%)
IM/GC	4 (40%)	6 (60%)	0 (0%)	0 (0%)
Household number≥6				
Free of HP infection, AG, and IM/GC	7 (87.5%)	1 (12.5%)	0 (0%)	0 (0%)
HP infection	1 (100%)	0 (0%)	0 (0%)	0 (0%)
AG	7 (87.5%)	1 (12.5%)	0 (0%)	0 (0%)
IM/GC	7 (77.78%)	2 (22.22%)	0 (0%)	0 (0%)

(b) Number of relatives with AG

doi:10.6342/NTU201902662

Index	Number of relatives with AG				
	0	1	2	3	4
Household number=2					
Free of HP infection, AG, and IM/GC	55 (57.29%)	41 (42.71%)	-	-	-
HP infection	26 (70.27%)	11 (29.73%)	-	-	-
AG	80 (56.34%)	62 (43.66%)	-	-	-
IM/GC	47 (63.51%)	27 (36.49%)	-	-	-
Household number=3					
Free of HP infection, AG, and IM/GC	15 (32.61%)	24 (52.17%)	7 (15.22%)	-	-
HP infection	5 (38.46%)	5 (38.46%)	3 (23.08%)	-	-
AG	20 (27.4%)	46 (63.01%)	7 (9.59%)	-	-
IM/GC	19 (50%)	10 (26.32%)	9 (23.68%)	-	-
Household number=4					
Free of HP infection, AG, and IM/GC	9 (37.5%)	7 (29.17%)	6 (25%)	2 (8.33%)	-
HP infection	5 (35.71%)	6 (42.86%)	0 (0%)	3 (21.43%)	-
AG	10 (22.22%)	10 (22.22%)	25 (55.56%)	0 (0%)	-
IM/GC	3 (15.79%)	11 (57.89%)	3 (15.79%)	2 (10.53%)	-
Household number=5					
Free of HP infection, AG, and IM/GC	3 (33.33%)	1 (11.11%)	3 (33.33%)	2 (22.22%)	0 (0%)
HP infection	2 (33.33%)	3 (50%)	1 (16.67%)	0 (0%)	0 (0%)
AG	1 (11.11%)	2 (22.22%)	3 (33.33%)	1 (11.11%)	2 (22.22%)
IM/GC	2 (20%)	4 (40%)	3 (30%)	1 (10%)	0 (0%)
Household number≥6					
Free of HP infection, AG, and IM/GC	0 (0%)	4 (50%)	3 (37.5%)	1 (12.5%)	0 (0%)
HP infection	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (100%)
AG	0 (0%)	6 (75%)	1 (12.5%)	1 (12.5%)	0 (0%)
IM/GC	0 (0%)	0 (0%)	5 (55.56%)	2 (22.22%)	2 (22.22%)

doi:10.6342/NTU201902662

(c) Number of relatives with IM/GC

Number of relatives with IM/GC					
Index	**0**	**1**	**2**	**3**	**4**
Household number=2					
Free of HP infection, AG, and IM/GC	77 (80.21%)	19 (19.79%)	-	-	-
HP infection	26 (70.27%)	11 (29.73%)	-	-	-
AG	113 (79.58%)	29 (20.42%)	-	-	-
IM/GC	56 (75.68%)	18 (24.32%)	-	-	-
Household number=3					
Free of HP infection, AG, and IM/GC	32 (69.57%)	12 (26.09%)	2 (4.35%)	-	-
HP infection	8 (61.54%)	4 (30.77%)	1 (7.69%)	-	-
AG	39 (53.42%)	32 (43.84%)	2 (2.74%)	-	-
IM/GC	24 (63.16%)	11 (28.95%)	3 (7.89%)	-	-
Household number=4					
Free of HP infection, AG, and IM/GC	16 (66.67%)	6 (25%)	2 (8.33%)	0 (0%)	-
HP infection	9 (64.29%)	5 (35.71%)	0 (0%)	0 (0%)	-
AG	29 (64.44%)	11 (24.44%)	4 (8.89%)	1 (2.22%)	-
IM/GC	7 (36.84%)	8 (42.11%)	4 (21.05%)	0 (0%)	-
Household number=5					
Free of HP infection, AG, and IM/GC	2 (22.22%)	2 (22.22%)	3 (33.33%)	2 (22.22%)	0 (0%)
HP infection	2 (33.33%)	2 (33.33%)	2 (33.33%)	0 (0%)	0 (0%)
AG	5 (55.56%)	3 (33.33%)	1 (11.11%)	0 (0%)	0 (0%)
IM/GC	2 (20%)	2 (20%)	5 (50%)	1 (10%)	0 (0%)
Household number≥6					
Free of HP infection, AG, and IM/GC	4 (50%)	0 (0%)	3 (37.5%)	1 (12.5%)	0 (0%)
HP infection	0 (0%)	1 (100%)	0 (0%)	0 (0%)	0 (0%)
AG	2 (25%)	1 (12.5%)	2 (25%)	1 (12.5%)	2 (25%)
IM/GC	3 (33.33%)	2 (22.22%)	0 (0%)	3 (33.33%)	1 (11.11%)

doi:10.6342/NTU201902662

Table 4.2.1.6 Number of relatives with different statuses of disease by household number (index case from random sample in 2008)

(a) Number of relatives with HP infection

Number of relatives with HP infection				
Index	**0**	**1**	**2**	**3**
Household number=2				
Free of HP infection, AG, and IM/GC	84 (87.5%)	12 (12.5%)	-	-
HP infection	28 (75.68%)	9 (24.32%)	-	-
AG	127 (89.44%)	15 (10.56%)	-	-
IM/GC	68 (91.89%)	6 (8.11%)	-	-
Household number=3				
Free of HP infection, AG, and IM/GC	44 (95.65%)	2 (4.35%)	0 (0%)	-
HP infection	10 (76.92%)	3 (23.08%)	0 (0%)	-
AG	64 (87.67%)	9 (12.33%)	0 (0%)	-
IM/GC	31 (81.58%)	7 (18.42%)	0 (0%)	-
Household number=4				
Free of HP infection, AG, and IM/GC	13 (54.17%)	10 (41.67%)	1 (4.17%)	0 (0%)
HP infection	6 (42.86%)	5 (35.71%)	1 (7.14%)	2 (14.29%)
AG	28 (62.22%)	15 (33.33%)	2 (4.44%)	0 (0%)
IM/GC	14 (73.68%)	4 (21.05%)	1 (5.26%)	0 (0%)
Household number=5				
Free of HP infection, AG, and IM/GC	8 (88.89%)	1 (11.11%)	0 (0%)	0 (0%)
HP infection	2 (33.33%)	1 (16.67%)	3 (50%)	0 (0%)
AG	7 (77.78%)	1 (11.11%)	0 (0%)	1 (11.11%)
IM/GC	4 (40%)	6 (60%)	0 (0%)	0 (0%)
Household number≥6				
Free of HP infection, AG, and IM/GC	7 (87.5%)	1 (12.5%)	0 (0%)	0 (0%)
HP infection	1 (100%)	0 (0%)	0 (0%)	0 (0%)
AG	7 (87.5%)	1 (12.5%)	0 (0%)	0 (0%)
IM/GC	7 (77.78%)	2 (22.22%)	0 (0%)	0 (0%)

doi:10.6342/NTU201902662

(b) Number of relatives with AG

Number of relatives with AG					
Index	**0**	**1**	**2**	**3**	**4**
Household number=2					
Free of HP infection, AG, and IM/GC	55 (57.29%)	41 (42.71%)	-	-	-
HP infection	26 (70.27%)	11 (29.73%)	-	-	-
AG	80 (56.34%)	62 (43.66%)	-	-	-
IM/GC	47 (63.51%)	27 (36.49%)	-	-	-
Household number=3					
Free of HP infection, AG, and IM/GC	15 (32.61%)	24 (52.17%)	7 (15.22%)	-	-
HP infection	5 (38.46%)	5 (38.46%)	3 (23.08%)	-	-
AG	20 (27.4%)	46 (63.01%)	7 (9.59%)	-	-
IM/GC	19 (50%)	10 (26.32%)	9 (23.68%)	-	-
Household number=4					
Free of HP infection, AG, and IM/GC	9 (37.5%)	7 (29.17%)	6 (25%)	2 (8.33%)	-
HP infection	5 (35.71%)	6 (42.86%)	0 (0%)	3 (21.43%)	-
AG	10 (22.22%)	10 (22.22%)	25 (55.56%)	0 (0%)	-
IM/GC	3 (15.79%)	11 (57.89%)	3 (15.79%)	2 (10.53%)	-
Household number=5					
Free of HP infection, AG, and IM/GC	3 (33.33%)	1 (11.11%)	3 (33.33%)	2 (22.22%)	0 (0%)
HP infection	2 (33.33%)	3 (50%)	1 (16.67%)	0 (0%)	0 (0%)
AG	1 (11.11%)	2 (22.22%)	3 (33.33%)	1 (11.11%)	2 (22.22%)
IM/GC	2 (20%)	4 (40%)	3 (30%)	1 (10%)	0 (0%)
Household number≥6					
Free of HP infection, AG, and IM/GC	0 (0%)	4 (50%)	3 (37.5%)	1 (12.5%)	0 (0%)
HP infection	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (100%)
AG	0 (0%)	6 (75%)	1 (12.5%)	1 (12.5%)	0 (0%)
IM/GC	0 (0%)	0 (0%)	5 (55.56%)	2 (22.22%)	2 (22.22%)

doi:10.6342/NTU201902662

(c) Number of relatives with IM/GC

Number of relatives with IM/GC					
Index	**0**	**1**	**2**	**3**	**4**
Household number=2					
Free of HP infection, AG, and IM/GC	77 (80.21%)	19 (19.79%)	-	-	-
HP infection	26 (70.27%)	11 (29.73%)	-	-	-
AG	113 (79.58%)	29 (20.42%)	-	-	-
IM/GC	56 (75.68%)	18 (24.32%)	-	-	-
Household number=3					
Free of HP infection, AG, and IM/GC	32 (69.57%)	12 (26.09%)	2 (4.35%)	-	-
HP infection	8 (61.54%)	4 (30.77%)	1 (7.69%)	-	-
AG	39 (53.42%)	32 (43.84%)	2 (2.74%)	-	-
IM/GC	24 (63.16%)	11 (28.95%)	3 (7.89%)	-	-
Household number=4					
Free of HP infection, AG, and IM/GC	16 (66.67%)	6 (25%)	2 (8.33%)	0 (0%)	-
HP infection	9 (64.29%)	5 (35.71%)	0 (0%)	0 (0%)	-
AG	29 (64.44%)	11 (24.44%)	4 (8.89%)	1 (2.22%)	-
IM/GC	7 (36.84%)	8 (42.11%)	4 (21.05%)	0 (0%)	-
Household number=5					
Free of HP infection, AG, and IM/GC	2 (22.22%)	2 (22.22%)	3 (33.33%)	2 (22.22%)	0 (0%)
HP infection	2 (33.33%)	2 (33.33%)	2 (33.33%)	0 (0%)	0 (0%)
AG	5 (55.56%)	3 (33.33%)	1 (11.11%)	0 (0%)	0 (0%)
IM/GC	2 (20%)	2 (20%)	5 (50%)	1 (10%)	0 (0%)
Household number≥6					
Free of HP infection, AG, and IM/GC	1 (11.11%)	0 (0%)	4 (44.44%)	0 (0%)	4 (44.44%)
HP infection	0 (0%)	1 (100%)	0 (0%)	0 (0%)	0 (0%)
AG	1 (10%)	2 (20%)	2 (20%)	0 (0%)	5 (50%)
IM/GC	3 (30%)	1 (10%)	0 (0%)	3 (30%)	3 (30%)

doi:10.6342/NTU201902662

Table 4.2.1.7 Effect of family aggregation on gastric pre-cancer lesion (full dataset)

Variable	OR	95% CI		aOR	95% CI	
Proband status - HP(+)						
Index status (vs normal)						
HP(+)	1.63	1.36	1.96	1.61	1.33	1.95
AG	1.08	0.87	1.35	1.08	0.86	1.36
IM/GC	1.51	1.10	2.06	1.43	1.04	1.97
Age	-	-	-	0.99	0.98	0.99
Female (vs male)	-	-	-	1.06	0.89	1.25
Smoking (vs no)	-	-	-	0.65	0.53	0.79
Salty food	-	-	-	1.00	0.97	1.04
Seafood	-	-	-	0.98	0.92	1.05
Proband status - AG						
Index status (vs normal)						
HP(+)	1.08	0.87	1.35	1.14	0.90	1.43
AG	1.52	1.20	1.92	1.39	1.09	1.77
IM/GC	1.49	1.05	2.11	1.59	1.11	2.28
Age	-	-	-	0.97	0.96	0.97
Female (vs male)	-	-	-	0.99	0.82	1.20
Smoking (vs no)	-	-	-	1.38	1.08	1.77
Salty food	-	-	-	0.99	0.95	1.04
Seafood	-	-	-	0.99	0.91	1.07
Proband status – IM/GC						
Index status (vs normal)						
HP(+)	1.51	1.11	2.07	1.53	1.12	2.11
AG	1.51	1.06	2.14	1.47	1.03	2.09
IM/GC	2.04	1.28	3.27	2.02	1.26	3.25
Age	-	-	-	0.99	0.99	1.00
Female (vs male)	-	-	-	0.82	0.63	1.07
Smoking (vs no)	-	-	-	0.75	0.53	1.05
Salty food	-	-	-	1.00	0.94	1.06
Seafood	-	-	-	1.00	0.90	1.12

Abbreviation: HP: Helicobacter pylori; AG: Atrophic gastritis; IM: Intestinal metaplasia

doi:10.6342/NTU201902662

Table 4.2.1.8 Effect of family aggregation on gastric pre-cancer lesion (random sample)

Variable	OR	95% CI		aOR	95% CI	
Proband status - HP(+)						
Index status (vs normal)						
HP(+)	1.39	1.07	1.80	1.35	1.03	1.77
AG	0.83	0.61	1.14	0.80	0.58	1.10
IM/GC	1.60	1.03	2.47	1.65	1.06	2.59
Age	-	-	-	0.99	0.98	0.99
Female (vs male)	-	-	-	0.99	0.75	1.32
Smoking (vs no)	-	-	-	0.86	0.61	1.21
Salty food	-	-	-	1.07	1.01	1.13
Seafood	-	-	-	0.95	0.86	1.05
Proband status - AG						
Index status (vs normal)						
HP(+)	1.06	0.78	1.43	1.00	0.73	1.38
AG	1.30	0.94	1.79	1.18	0.83	1.66
IM/GC	1.02	0.60	1.75	1.03	0.59	1.80
Age	-	-	-	0.96	0.95	0.97
Female (vs male)	-	-	-	0.71	0.51	0.98
Smoking (vs no)	-	-	-	0.70	0.48	1.03
Salty food	-	-	-	0.96	0.89	1.04
Seafood	-	-	-	0.91	0.79	1.05
Proband status – IM/GC						
Index status (vs normal)						
HP(+)	1.44	0.93	2.25	1.44	0.93	2.26
AG	1.47	0.91	2.39	1.49	0.91	2.42
IM/GC	2.17	1.12	4.18	2.26	1.17	4.39
Age	-	-	-	1.00	0.99	1.01
Female (vs male)	-	-	-	0.71	0.46	1.10
Smoking (vs no)	-	-	-	0.53	0.31	0.92
Salty food	-	-	-	1.01	0.92	1.11
Seafood	-	-	-	1.03	0.87	1.21

Abbreviation: HP: Helicobacter pylori; AG: Atrophic gastritis; IM: Intestinal metaplasia

doi:10.6342/NTU201902662

Table 4.2.1.9 Effect of family aggregation on gastric pre-cancer lesion (random sample in 1996 (Period 1))

Variable	OR	95% CI		aOR	95% CI	
Proband status - HP(+)						
Index status (vs normal)						
HP(+)	1.48	1.11	1.97	1.41	1.05	1.91
AG	1.48	0.94	2.33	1.44	0.90	2.31
IM/GC	1.60	0.90	2.83	1.74	0.96	3.14
Age	-	-	-	0.98	0.97	0.99
Female (vs male)	-	-	-	0.86	0.60	1.21
Smoking (vs no)	-	-	-	0.65	0.43	0.98
Salty food	-	-	-	1.06	0.99	1.13
Seafood	-	-	-	0.94	0.84	1.05
Proband status - AG						
Index status (vs normal)						
HP(+)	1.74	1.10	2.74	1.72	1.08	2.73
AG	3.75	2.10	6.68	3.82	2.12	6.88
IM/GC	3.56	1.73	7.32	3.76	1.81	7.78
Age	-	-	-	0.99	0.97	1.00
Female (vs male)	-	-	-	0.67	0.42	1.08
Smoking (vs no)	-	-	-	0.49	0.27	0.90
Salty food	-	-	-	1.05	0.95	1.16
Seafood	-	-	-	0.90	0.76	1.06
Proband status – IM/GC						
Index status (vs normal)						
HP(+)	1.69	0.99	2.86	1.75	1.02	2.99
AG	2.83	1.41	5.67	3.05	1.50	6.21
IM/GC	2.20	0.87	5.59	2.21	0.86	5.69
Age	-	-	-	1.01	0.99	1.03
Female (vs male)	-	-	-	0.57	0.33	0.99
Smoking (vs no)	-	-	-	0.39	0.19	0.80
Salty food	-	-	-	1.04	0.93	1.17
Seafood	-	-	-	1.01	0.84	1.22

Abbreviation: HP: Helicobacter pylori; AG: Atrophic gastritis; IM: Intestinal metaplasia

doi:10.6342/NTU201902662

Table 4.2.1.10 Effect of family aggregation on gastric pre-cancer lesion (random sample in 2004 (Period 2))

Variable	OR	95% CI		aOR	95% CI	
Proband status - HP(+)						
Index status (vs normal)						
HP(+)	4.65	2.42	8.95	4.43	2.28	8.59
AG	1.44	0.85	2.44	1.37	0.80	2.35
IM/GC	1.43	0.78	2.61	1.45	0.79	2.66
Age	-	-	-	1.00	0.99	1.01
Female (vs male)	-	-	-	2.17	1.22	3.87
Smoking (vs no)	-	-	-	1.34	0.69	2.63
Salty food	-	-	-	1.12	0.99	1.26
Seafood	-	-	-	0.75	0.60	0.94
Proband status - AG						
Index status (vs normal)						
HP(+)	1.44	0.85	2.46	1.40	0.82	2.39
AG	1.37	0.98	1.90	1.34	0.96	1.87
IM/GC	1.31	0.89	1.92	1.29	0.88	1.91
Age	-	-	-	0.99	0.99	1.00
Female (vs male)	-	-	-	0.69	0.49	0.98
Smoking (vs no)	-	-	-	0.62	0.41	0.95
Salty food	-	-	-	1.03	0.95	1.11
Seafood	-	-	-	0.91	0.79	1.04
Proband status – IM/GC						
Index status (vs normal)						
HP(+)	1.74	0.95	3.16	1.89	1.03	3.47
AG	1.33	0.90	1.97	1.47	0.99	2.20
IM/GC	1.76	1.14	2.72	1.96	1.26	3.05
Age	-	-	-	1.02	1.01	1.03
Female (vs male)	-	-	-	0.68	0.46	1.01
Smoking (vs no)	-	-	-	0.80	0.50	1.30
Salty food	-	-	-	1.05	0.97	1.14
Seafood	-	-	-	0.98	0.85	1.13

Abbreviation: HP: Helicobacter pylori; AG: Atrophic gastritis; IM: Intestinal metaplasia

doi:10.6342/NTU201902662

Table 4.2.1.11 Effect of family aggregation on gastric pre-cancer lesion (random sample in 2008 (Period 3))

Variable	OR	95% CI		aOR	95% CI	
Proband status - HP(+)						
Index status (vs normal)						
HP(+)	4.07	2.12	7.82	3.80	1.96	7.37
AG	1.52	0.89	2.59	1.44	0.84	2.48
IM/GC	1.18	0.66	2.10	1.18	0.66	2.11
Age	-	-	-	1.00	0.99	1.01
Female (vs male)	-	-	-	2.12	1.19	3.78
Smoking (vs no)	-	-	-	1.33	0.68	2.61
Salty food	-	-	-	1.15	1.02	1.29
Seafood	-	-	-	0.71	0.57	0.89
Proband status - AG						
Index status (vs normal)						
HP(+)	1.46	0.85	2.51	1.41	0.82	2.43
AG	1.43	1.00	2.03	1.38	0.96	1.98
IM/GC	1.30	0.90	1.88	1.27	0.87	1.84
Age	-	-	-	0.99	0.98	1.00
Female (vs male)	-	-	-	0.75	0.53	1.06
Smoking (vs no)	-	-	-	0.63	0.41	0.97
Salty food	-	-	-	1.02	0.94	1.11
Seafood	-	-	-	0.89	0.77	1.03
Proband status – IM/GC						
Index status (vs normal)						
HP(+)	1.34	0.75	2.40	1.46	0.81	2.64
AG	1.24	0.85	1.82	1.35	0.91	1.99
IM/GC	1.52	1.03	2.23	1.72	1.16	2.55
Age	-	-	-	1.02	1.01	1.03
Female (vs male)	-	-	-	0.61	0.42	0.89
Smoking (vs no)	-	-	-	0.76	0.48	1.19
Salty food	-	-	-	1.07	0.99	1.15
Seafood	-	-	-	0.97	0.85	1.11

Abbreviation: HP: Helicobacter pylori; AG: Atrophic gastritis; IM: Intestinal metaplasia

doi:10.6342/NTU201902662

Table 4.2.1.12 Effect of family aggregation on gastric pre-cancer lesion (random sample combing three periods (1996, 2004, 2008))

Variable	OR	95% CI		aOR	95% CI	
Proband status - HP(+)						
Index status (vs normal)						
HP(+)	1.39	1.07	1.80	1.83	1.43	2.34
AG	0.83	0.61	1.14	1.16	0.89	1.52
IM/GC	1.60	1.03	2.47	1.14	0.83	1.55
Age	-	-	-	0.99	0.98	0.99
Female (vs male)	-	-	-	1.22	0.95	1.57
Smoking (vs no)	-	-	-	0.91	0.67	1.22
Salty food	-	-	-	1.08	1.02	1.13
Seafood	-	-	-	0.87	0.80	0.94
Period (vs 1996)						
2004	-	-	-	0.36	0.27	0.47
2008	-	-	-	0.35	0.27	0.46
Proband status - AG						
Index status (vs normal)						
HP(+)	1.06	0.78	1.43	1.27	0.97	1.67
AG	1.30	0.94	1.79	1.57	1.25	1.96
IM/GC	1.02	0.60	1.75	1.46	1.14	1.87
Age	-	-	-	0.99	0.98	1.00
Female (vs male)	-	-	-	0.67	0.54	0.84
Smoking (vs no)	-	-	-	0.57	0.44	0.75
Salty food	-	-	-	1.02	0.98	1.07
Seafood	-	-	-	0.91	0.84	0.99
Period (vs 1996)						
2004	-	-	-	3.45	2.67	4.45
2008	-	-	-	3.03	2.34	3.93
Proband status – IM/GC						
Index status (vs normal)						
HP(+)	1.44	0.93	2.25	1.51	1.10	2.06
AG	1.47	0.91	2.39	1.56	1.20	2.01
IM/GC	2.17	1.12	4.18	1.95	1.48	2.56
Age	-	-	-	1.02	1.01	1.02
Female (vs male)	-	-	-	0.60	0.48	0.77
Smoking (vs no)	-	-	-	0.68	0.50	0.90
Salty food	-	-	-	1.06	1.00	1.11
Seafood	-	-	-	0.99	0.91	1.08
Period (vs 1996)						

doi:10.6342/NTU201902662

Variable	OR	95% CI		aOR	95% CI	
2004	-	-	-	4.41	3.27	5.95
2008	-	-	-	5.28	3.92	7.11

Abbreviation: HP: Helicobacter pylori; AG: Atrophic gastritis; IM/GC: Intestinal metaplasia or gastric cancer

doi:10.6342/NTU201902662

Table 4.2.1.13 Effect of family aggregation on gastric pre-cancer lesion with DAG model (full dataset)

Index status (vs normal)	OR	95% CI	
Proband status - HP(+)			
HP(+)	1.63	1.36	1.96
AG	1.08	0.87	1.35
IM/GC	1.51	1.1	2.06
Proband status - AG			
HP(+)	1.10	0.88	1.38
AG	1.55	1.22	1.94
IM/GC	1.53	1.08	2.11
Proband status – IM/GC			
HP(+)	1.54	1.09	2.16
AG	1.53	1.05	2.22
IM/GC	2.08	1.24	3.31

Table 4.2.1.14 Effect of family aggregation on gastric pre-cancer lesion with DAG model (random sample)

Index status (vs normal)	OR	95% CI	
Proband status - HP(+)			
HP(+)	1.38	1.07	1.80
AG	0.83	0.61	1.13
IM/GC	1.60	1.04	2.48
Proband status - AG			
HP(+)	1.05	0.78	1.42
AG	1.30	0.94	1.79
IM/GC	1.01	0.58	1.72
Proband status – IM/GC			
HP(+)	1.70	1.01	2.92
AG	2.82	1.38	5.62
IM/GC	2.12	0.76	5.31

doi:10.6342/NTU201902662

Table 4.2.1.15 Effect of family aggregation on gastric pre-cancer lesion with DAG model (random sample in 1996 (Period 1))

Index status (vs normal)	OR	95% CI	
Proband status - HP(+)			
HP(+)	1.48	1.12	1.97
AG	1.49	0.95	2.35
IM/GC	1.61	0.91	2.88
Proband status - AG			
HP(+)	1.75	1.11	2.79
AG	3.77	2.13	6.81
IM/GC	3.57	1.74	7.29
Proband status – IM/GC			
HP(+)	1.44	0.84	2.49
AG	1.36	0.98	1.89
IM/GC	1.31	0.89	1.93

Table 4.2.1.16 Effect of family aggregation on gastric pre-cancer lesion with DAG model (random sample in 2004 (Period 2))

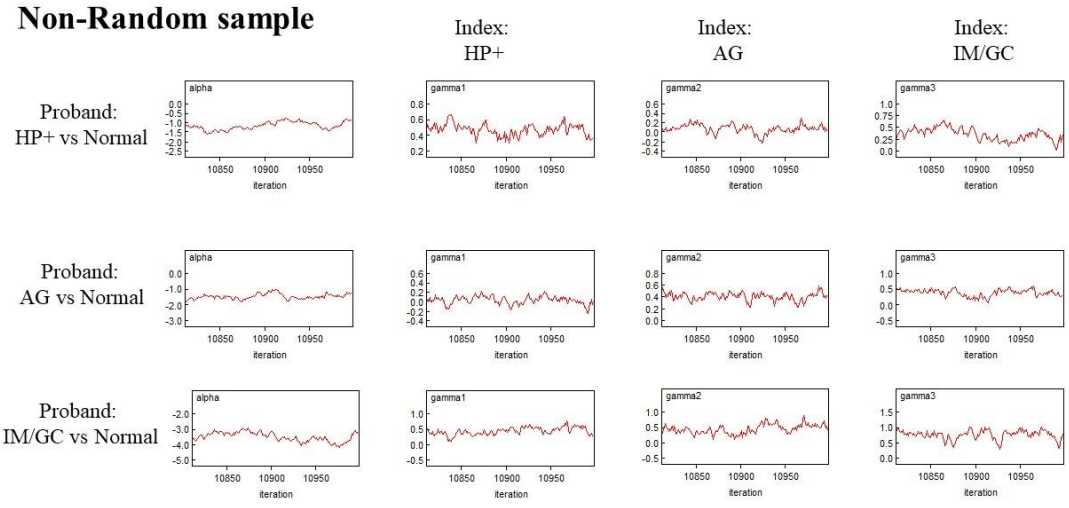
Index status (vs normal)	OR	95% CI	
Proband status - HP(+)			
HP(+)	4.70	2.42	9.27
AG	1.45	0.85	2.53
IM/GC	1.43	0.78	2.64
Proband status - AG			
HP(+)	1.44	0.84	2.49
AG	1.36	0.98	1.89
IM/GC	1.31	0.89	1.93
Proband status – IM/GC			
HP(+)	1.35	0.76	2.41
AG	1.24	0.85	1.82
IM/GC	1.53	1.04	2.24

doi:10.6342/NTU201902662

Table 4.2.1.17 Effect of family aggregation on gastric pre-cancer lesion with DAG model (random sample in 2008 (Period 3))

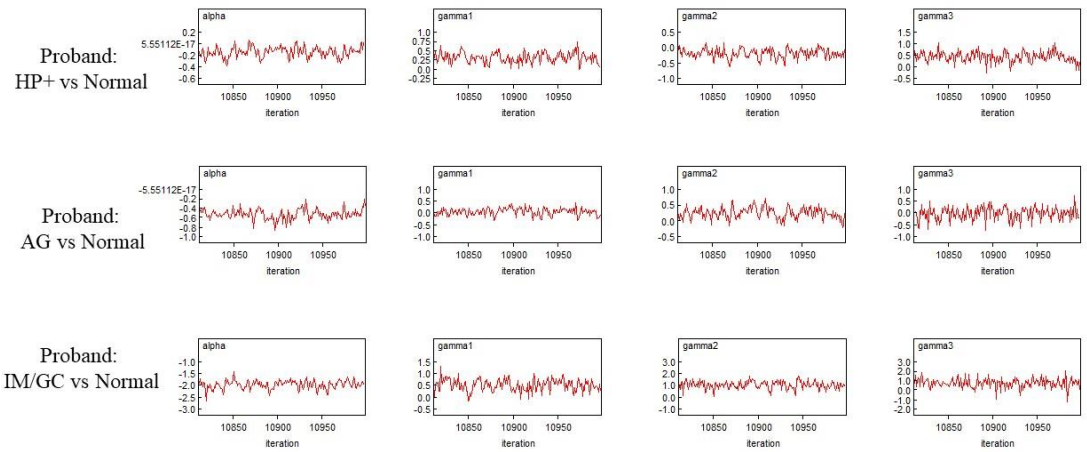
Index status (vs normal)	OR	95% CI	
Proband status - HP(+)			
HP(+)	4.08	2.10	8.04
AG	1.52	0.89	2.66
IM/GC	1.17	0.65	2.11
Proband status - AG			
HP(+)	1.47	0.86	2.54
AG	1.43	1.01	2.03
IM/GC	1.30	0.91	1.89
Proband status – IM/GC			
HP(+)	1.35	0.76	2.41
AG	1.24	0.85	1.82
IM/GC	1.53	1.04	2.24

4.2.7 Trace Plots of DAG Model

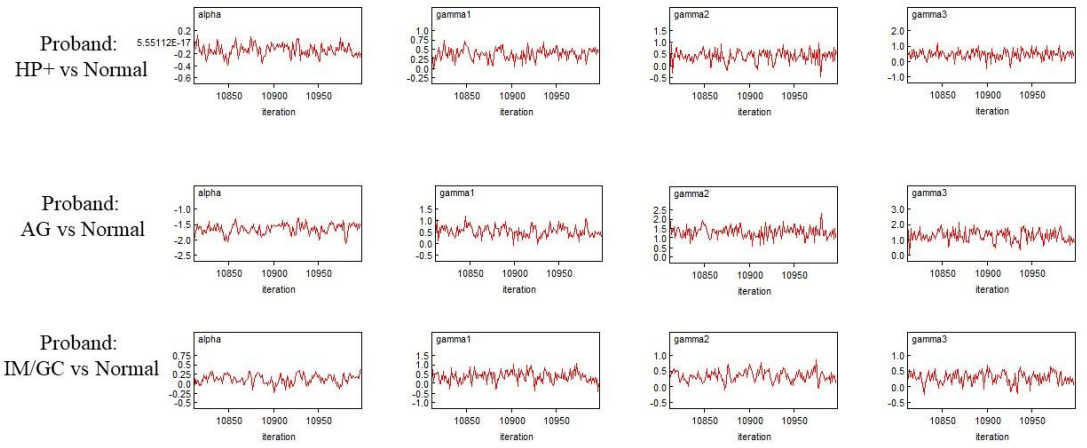


doi:10.6342/NTU201902662

Random sample

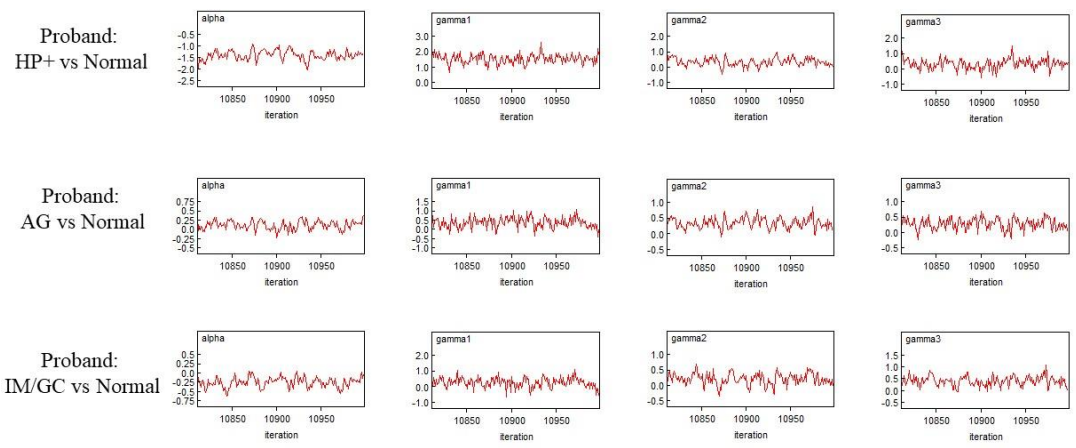


Random sample Period 1

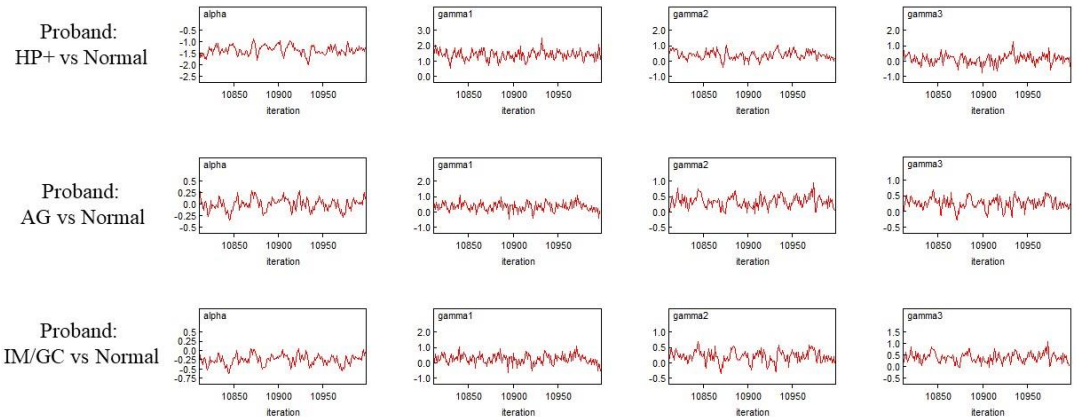


doi:10.6342/NTU201902662

Random sample Period 2



Random sample Period 3



doi:10.6342/NTU201902662

4.3 Family aggregation with Markov chain model

In table 4.3.1 the ratio of sum upper triangle and sum of lower triangle is higher means the markov chain with ergodicity theory base on Correa model have stronger family aggregation. Using full family data, the ratio is 2.1, shows the highly evidence of family aggregation and also the sampling data.

Furthermore, in the first period (1996) the △U/△L ratio is 2.24, shows the stronger family aggregation before intervention. However, after intervention of screening and chemoprevention the ratio decrease to 0.91 and 0.87 respectively for 2004 and 2008. In addition to the equilibrium distribution of HP infected is 0.40 in 1996, and after intervention the distribution change from 0.40 to 0.11 and 0.10 in period2 and period3.(Table 4.3.2)

doi:10.6342/NTU201902662

Table 4.3.1.1 Family aggregation using transition probabilities in Markov chain

Proband status	Index Status				△U /△L
Full Data					
	Normal	**HP**	**AG**	**IM/GC**	
Normal	0.41222	0.33034	0.35256	0.31010	
HP	0.31090	0.40638	0.28834	0.35192	2.10
AG	0.20882	0.18070	0.27130	0.33838	
IM/GC	0.06806	0.08258	0.08781	0.10453	
Random Sampling					
	Normal	**HP**	**AG**	**IM/GC**	
Normal	0.40379	0.35935	0.37250	0.30612	
HP	0.31073	0.38374	0.30250	0.34014	2.10
AG	0.21924	0.16260	0.26250	0.24490	
IM/GC	0.06625	0.09431	0.06250	0.10884	
Random Sampling in 1996 (Period 1)					
	Normal	**HP**	**AG**	**IM/GC**	
Normal	0.45356	0.39525	0.08855	0.06263	
HP	0.34945	0.45055	0.11868	0.08132	2.24
AG	0.29286	0.37857	0.21429	0.11429	
IM/GC	0.29487	0.41026	0.20513	0.08974	
Random Sampling in 2004 (Period 1)					
	Normal	**HP**	**AG**	**IM/GC**	
Normal	0.33529	0.08235	0.38235	0.20000	
HP	0.20741	0.23704	0.34074	0.21481	0.91
AG	0.26984	0.09524	0.42063	0.21429	
IM/GC	0.25667	0.09000	0.38333	0.27000	
Random Sampling in 2008 (Period 1)					
	Normal	**HP**	**AG**	**IM/GC**	
Normal	0.33140	0.08430	0.32849	0.25581	
HP	0.22137	0.22901	0.32061	0.22901	0.87
AG	0.26636	0.10280	0.37617	0.25467	
IM/GC	0.26596	0.07979	0.34309	0.31117	

△U：Summation of transition probabilities in the upper triangle

△L：Summation of transition probabilities in the lower triangle

doi:10.6342/NTU201902662

Table 4.3.1.2 Family aggregation using equilibrium distribution in Markov chain

Period	Disease status			
	Normal	**HP**	**AG**	**IM/GC**
1996	0.41	0.40	0.12	0.07
2004	0.26	0.11	0.39	0.24
2008	0.27	0.10	0.33	0.30

doi:10.6342/NTU201902662

4.4 Continues Markov multistate model

The estimated results based on the four-state Markov regression model with the case-control proband underpinning are listed in Table 4.3.2.1 to Table 4.3.2.3. Table 4.3.2.1 shows the estimated results on the three transition rates from normal status to H.P. infected (Normal HP (+)), further to AG (HP(+) AG), and then to IM (AG IM) by the status of index case. Given the index case with normal status, the three rates of transition were estimated as 1.82%, 2.78%, and 1.88% for normal HP (+), HP(+) AG, and AG IM, respectively. For the proband of HP(+) cases, the estimated rat of transition from normal to the status of HP infected is 2.37%, around 1.3 fold of those with normal index cases. A higher rates of turning into HP infected status can be noted for index case of AG (2.20%) and IM/GC (2.45%). The rate of transition from HP (+) to AG is highest for the proband with index case of AG (3.7%), around 1.3 fold for those with normal index case. The progression rate from AG to IM is around 2% regardless the status of index case, except for the index case of HP (2.4%).

Based on the Markov regression model, the information factors associate with the progression of gastric lesions such as gender and smoking behavior can be incorporated to see the role of index case status and these factors. Adjusting for gender and smoking, index cases with the status of IM (aHR: 1.35 95% CI: 1.14-1.59), HP(+) (aHR: 1.30, 95% CI: 1.18-1.44), and IM (aHR: 1.21, 95% CI: 1.07-1.36) result in a significantly higher transition rate from normal to HP infection. Given an index case of AG, the proband also have a higher risk of transition from HP(+) to AG (aHR: 1.33, 95% CI: 1.10-1.60). Regarding the transition from AG to IM/GC, the status of index cases have no significant role.

As the implementation of gastric cancer prevention programme in Matsu can be divided

doi:10.6342/NTU201902662

into the periods 1997-2004, and 2005-2008, we use the period before 1996 as a comparator to assess the period effect. It is well accepted that diet habits including salty foot intake and sea food intake may have influence on gastric lesion progression, these factors were evaluated using the Markov regression model with the results lists in Table 4.3.2.3. The effect of index case status on the risk of gastric lesion progression are consistent with that observed in Table 4.2.3.2 except that the IM index cases shows a significant influence for the progression from HP(+) to AG (aHR: 1.53, 95% CI: 1.20-1.96. Controlling for the index case status and period effect, diet habits demonstrate no significant effect on the transition of proband. Compared with early period (1996), both the period of 2004 and 2008 show a significantly higher risk of being the status of HP infection and further progress to AG with the aHR estimated around 1.5 and 6.7, respectively. This is due to the active detection of the implementation of gastric cancer prevention programme in Matsu. For the period of 2008, the risk of progress to IM and GC also shows an increasing but non-significant trend (aHR: 1.18), also due to the active detection for the IM/GC lesion through the introduction of endoscopic examination in the gastric cancer prevention programme.

doi:10.6342/NTU201902662

Table 4.4.1 Transition Rates by the index's disease status

Status of Index	**Transition Rate**		
	Normal → HP(+)	HP(+) → AG	AG → IM
Normal	1.82%	2.78%	1.88%
HP(+)	2.37%	2.05%	2.40%
AG	2.20%	3.70%	1.79%
IM/GC	2.45%	2.94%	2.15%

doi:10.6342/NTU201902662

Table 4.4.2 Estimates of relative risk in family aggregation adjusting for gender and smoking status

Probands' status		**Base line Transition Rate**	**95% CI**	
Normal → HP(+)		0.018	0.017	0.020
HP(+) → AG		0.028	0.024	0.031
AG → IM/GC		0.019	0.014	0.023
Probands' status	**Variable**	**exp(β)**	**95% CI**	
Normal → HP(+)	Index: HP(+)	1.30	1.18	1.44
	Index: AG	1.21	1.07	1.36
	Index: IM	1.35	1.14	1.59
HP(+) → AG	Index: HP(+)	0.74	0.62	0.88
	Index: AG	1.33	1.10	1.60
	Index: IM	1.06	0.82	1.37
AG → IM/GC	Index: HP(+)	1.28	0.94	1.75
	Index: AG	0.95	0.68	1.34
	Index: IM	1.14	0.72	1.81
Overall	Smoking	0.93	0.86	1.01

doi:10.6342/NTU201902662

Table 4.4.3 Estimates of relative risk in family aggregation adjusting for gender, smoking status, period and diet

Probands' status		**Base line Transition Rate**	**95% CI**	
Normal → HP(+)		0.025	0.020	0.030
HP(+) → AG		0.024	0.018	0.030
AG → IM/GC		0.038	0.025	0.051
Probands' status	**Variable**	**exp(β)**	**95% CI**	
Normal → HP(+)	Index: HP(+)	1.37	1.21	1.55
	Index: AG	1.32	1.18	1.47
	Index: IM	1.30	1.15	1.47
HP(+) → AG	Index: HP(+)	0.74	0.59	0.92
	Index: AG	1.34	1.08	1.66
	Index: IM	1.53	1.20	1.96
AG → IM/GC	Index: HP(+)	1.19	0.90	1.59
	Index: AG	0.88	0.71	1.10
	Index: IM	1.10	0.88	1.39
Overall	Female vs Male	0.78	0.72	0.85
	Smoking	0.87	0.79	0.95
	Sault Food	0.99	0.98	1.01
	Sea Food	0.99	0.97	1.02
Period	HP(+) 04 vs 96	1.53	1.36	1.71
	AG 04 vs 96	6.67	5.42	8.2
	IM/GC 04 vs 96	0.94	0.7	1.24
	HP(+) 08 vs 96	1.53	1.36	1.71
	AG 08 vs 96	6.65	5.41	8.18
	IM/GC 08 vs 96	1.18	0.88	1.57

doi:10.6342/NTU201902662

Chapter 5 Discussion

5.1 Quantitative approaches to studying family aggregation on HP infection and gastric neoplasm

Although intra-familial aggregation of *helicobacter pylori infection* has been well studied before, the effect size of family aggregation has been scarcely elucidated by using a well-designed genetic epidemiological study like case-control proband study. For example, in a 35 index Japanese pediatric patients familial study, two genetic typing procedures, MLST and random amplified polymorphicDNA (RAPD) fingerprinting have been used for detecting transmission route of familial aggregation. While this study found MLST is an useful indicator for detecting the transmission route of mother-to-child and sibling the magnitude of such a kind of intra-familial aggregation is still not exactly known.

While the familial aggregation of HP has been extended to include pre-cancerous lesions like AG, and IM, the effect size of familial aggregation study has been even barely addressed. This is partly because of intractable analysis of correlated multistate outcome and partly because of the inadequacy of applying the case-control proband study to the Correa's model of gastric neoplasm progression.

To address this request on the estimation of quantitative results, the current thesis developed a case-control proband study to estimate the effect size of familial aggregation of HP infection with the available information on DNA fingerprint using non-Bayesian and Bayesian modelling approach. The latter is the use of directed acyclic graphic (DAG) model to build up fully conditional distribution given the observed data and unknown quantity to estimate the effect size of family aggregation of HP infection making allowance for the correlation of HP infection across the same family members

doi:10.6342/NTU201902662

with the random effect model.

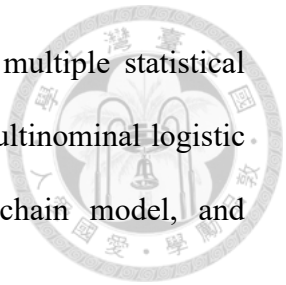
To model family aggregation of HP and gastric pre-cancerous neoplasm under the context of the Correa multistate model, the current thesis applied three approaches, the multi-nominal logistic regression model, the discrete-state and discrete-time Markov chain model, and continuous-time Markov process. These different approaches may have their own unique strengths and features when they are used for modelling familial aggregation of multistate outcome.

The multi-nominal logistic regression model is to extend two-state logistic regression model by treating the distribution of proband's disease status as a multi-nominal distribution that includes the multistate outcome of HP, AG, and IM to estimate the effect size of family aggregation index with three dummy variables with and without adjustment for other confounding factors.

The discrete-state and discrete-time Markov chain model is the application of Markov chain with the ergodicity theory to see whether the transition between the disease status of index cases and the disease status of probands can reach the equilibrium distribution. The matrix of transition probabilities is therefore used for assessing the extent of familial aggregation by comparing the summation of transition probabilities in the upper triangle with that in the lower triangle. The unique feature of this approach is to model familial aggregation with the disease status reaching the equilibrium after several generations as seen in the field of Hardy-Wen equilibrium distribution.

The continuous-time Markov process is to model the family aggregation of the transitions between multistate outcomes of probands taking into account time to each outcome i.e. age at diagnosis of different statuses. Such a kind of model enables one to assess how familial aggregation can affect the dynamic transition of HP and pre-cancerous lesion under the context of the Correa model.

doi:10.6342/NTU201902662



To sum up, in contrast to previous studies, we made use of multiple statistical methods to quantify the effect of familial aggregation , including multinominal logistic regression model, the discrete-state and discrete-time Markov chain model, and continuous-time Markov process.

In this study, we tried to use Japanese familial aggregation study and a community-based screening data to investigate the effect size of familial aggregation. The strength of this study is the quantitative estimation of familial aggregation in consideration of HP genotyping and the dynamic transition of pre-cancerous lesion in stomach. The results using Japanese study on genotyping associated with HP infection identified a strong tendency of familial aggregation for the matched sequence of RAPID or MLST compared with the unmatched sequence. The results are consistent with the study proposed by Yokota et al. In addition, considering the familial aggregation for the dynamic transition of HP, the results of multinominal logistic regression model and Markov chain model have suggest a significant familial aggregation (effect size in the range of 1.39-2.02) as well as a relatively strong familial aggregation in the early stage of re-cancerous lesion (transition from normal to HP infection).

5.2 Generalized estimating equation model (GEE)

We have already developed the case-control proband study design to estimate the effect size of family aggregation associated with the same DNA sequence (MLST only, RAPD only and MLST in conjunction with RAPD) with considering the correlation between the family members by using GEE. However, for the assumption of working matrix in GEE model, there are some widely used covariance structures such as (1) unstructured, which make allowance for the different correlation between distinct family members, (2) compound symmetry (exchangeable), which means the correlation with any

doi:10.6342/NTU201902662

other family members are the same and (3) autoregressive, which means the correlation with other members is dependent on the distance. When we used the unstructured correlation structure to estimate covariance, the degree of freedom was insufficient due to small sample size and the parameters could not be estimated. Therefore, we selected the compound symmetry in our analysis.

5.3 Bayesian DAG (directed acyclic graphic model)

With the application of the DAG model with Bayesian approach, we used the deviance information criterion (DIC) for model selection when comparing MLST with RAPD. Based on the DIC value, we discovered that RAPD (Model 2) was better than MLST (Model 1) and the combination of either MLST or RAPD (Model 3) was the same as RAPD only (Model 2), therefore, RAPD only (Model 2) might be the most appropriate method to estimate the effect of family aggregation of HP infection. By using the RAPD method, the effect size was enlarged for index subject with same trait and shrunk for index subject with different trait in comparison with MLST method. In addition, the effect size was greater with the DAG model than the GEE model, although the corresponding 95% CI became wider. Nevertheless, the advantage of using Bayesian approach is that we could take prior information into account and updated the results.

5.4 Multi-nominal logistic regression model

We have already used multi-nominal logistic regression model to assess the dynamic transitions between HP, AG, and IM under the context of the Correa model.
According to the results, if index status has HP infection or occurrence of AG or IM then the risk that his or her proband got HP, AG or IM would be increased.

With the application of multi-nominal logistic regression model to the Matsu data, the category ‘normal state’ is chosen as the reference category. So we can only compare

doi:10.6342/NTU201902662

infection HP, occurrence of AG and IM to normal state. Because of these states (HP and pre-cancerous lesions like atrophy gastritis (AG) and intestine metaplasia (IM)) have a progressive property, different situations such as “AG vs. HP+Normal”, “IM vs. AG+HP+Normal”, etc can be compared. This means we can use the cumulative logistic regression model to assess these situations in the future.

5.5 Family aggregation with Markov Chain Model

In addition to generalized linear models with either frequentist or Bayesian approach, stochastic models like Markov chain and Markov process can also be used to estimate the effect of family aggregation. By applying the limiting theorem and properties of Markov Chain model, we used two different methods to do estimation. First, from the equilibrium distributions of disease status (Table 4.3.1.2), we can see that the estimated long-term probability of HP infection decrease from 0.4 in 1996 to 0.10 in 2008. This result shows the success of HP eradication program in Matsu. Second, the ratio of summation of transition probabilities in the upper triangle to that of in the lower triangle also proves that the phenomenon of family aggregation disappears after intervention program. In the future, these two methods can also be applied to other kinds of aggregate data to do the evaluation of intervention program.

5.6 Continuous-time Discrete-state Markov Model

By using the four-state Markov regression model, we quantified the effect of family aggregation on each transition rate of evolution of gastric neoplasm, making allowance for gender and habit of cigarette smoking of proband in the same time.

Several novel findings were addressed. We elucidated the state-specific role of family aggregation on evolution of normal to HP infection and further progression to gastric neoplasm. The result shows that family aggregation was highly associated with the

doi:10.6342/NTU201902662

dynamic transition from normal to HP infection of the Correa model. As far as the severer status of gastric neoplasm in terms of AG, IM, or GC is concerned, the family aggregation shows a less effect on the transition rate of the dynamics Correa model. Applying the continuous-time Markov regression model, we successfully assessed how the family aggregation contributes to the dynamics of natural evolution of HP infection and gastric neoplasm.

doi:10.6342/NTU201902662

REFERENCE

Aziz RK, Khalifa MM, Sharaf RR. Contaminated water as a source of Helicobacter pylori infection: A review. J Adv Res. 2015;6(4):539-47.

Bamford KB, Bickley J, Collins JS, Johnston BT, Potts S, Boston V, Owen RJ, Sloan JM. Helicobacter pylori: comparison of DNA fingerprints provides evidence for intrafamilial infection. Gut 1993;34:1348–50.

Blaser MJ. The bacteria behind ulcers. Sci Am 1996;274:104–7.

Chiu YH, Lin WY, Wang PE, Chen YD, Wang TT, Warwick J, Chen TH. Population-based family case-control proband study on familial aggregation of metabolic syndrome: finding from Taiwanese people involved in Keelung community-based integrated screening (KCIS no. 5). Diabetes Res Clin Pract. 2007;75(3):348-56. Epub 2006 Aug 28.

Chiu SY, Chen LS, Yen AM, Chen HH. Population-based proband-oriented pedigree information system: application to hypertension with population-based screening data (KCIS No. 25). J Am Med Inform Assoc. 2012;19(1):102-10.

Cox, D.R. and Miller, H.D. (1965) The Theory of Stochastic Process. Chapman & Hall, London.

Dominici P, Bellentani S, Di Biase AR, Familial clustering of Helicobacter pylori infection: population based study. BMJ. 1999 Aug 28;319(7209):537-40.

Fukase K, Kato M, Kikuchi S, et al. Effect of eradication of Helicobacter pylori on incidence of metachronous gastric carcinoma after endoscopic resection of early gastric cancer: an open-label,randomised controlled trial. Lancet 2008;372:392–7.

Goodman KJ, Correa P. The transmission of Helicobacter pylori. A critical review of the evidence. Int J Epidemiol 1995;24:875–87.

Hashinaga M, Okimoto T, Kodama M, Mizukami K, Ogawa R, Okamoto K, et al. count report – of present conditions – 2013–2014 year resistant bacteria surveillance of drug-resistant Helicobacter pylori in our country [Japanese]. Jpn J Helicobacter Res. 2016;17:45–9.

IARC Helicobacter pylori Working Group. Helicobacter pylori Eradication as a Strategy for Gastric Cancer Prevention. Lyon, France: International Agency for Research on Cancer (IARC Working Group Reports, No. 8). International Agency for Research on Cancer 2014. 978-92-832-2454-9

Japanese Society for Helicobacter Research. Guidelines for the management of Helicobacter pylori infection in Japan. 2016 ed, revised. Tokyo: Sentan Igaku-Sha; 2016.

Kivi M, Tindberg Y. Helicobacter pylori occurrence and transmission: a family affair? Scand J Infect Dis. 2006;38(6-7):407-17.

doi:10.6342/NTU201902662

Kobayashi I, Murakami K, Kato M, Kato S, Azuma T, Takahashi S, et al. Changing antimicrobial susceptibility epidemiology of Helicobacter pylori strains in Japan between 2002 and 2005. J Clin Microbiol. 2007;45(12):4006–10.

Kobayashi I, Azuma T, Ikeda F, Uemura T, Kato S, Kato M, et al. Count report of the present conditions 2010-2011 year resistant bacteria surveillance of drug-resistant Helicobacter pylori in our country [Japanese]. Jpn J Helicobacter Res. 2013;14:102–6.

Lee YC, Chiang TH, Chou CK, Tu YK, Liao WC, Wu MS, Graham DY. Association Between Helicobacter pylori Eradication and Gastric Cancer Incidence: A Systematic Review and Meta-analysis. Gastroenterology. 2016;150(5):1113-1124.e5.

Liang KY, Pulver AE. Analysis of case-control/family sampling design. Genet Epidemiol. 1996;13(3):253-70.

Mamishi S, Eshaghi H, Mahmoudi S, Bahador A, Hosseinpour Sadeghi R, Najafi M, Farahmand F, Khodadad A, Pourakbari B. Intrafamilial transmission of Helicobacter pylori: genotyping of faecal samples. Br J Biomed Sci. 2016;73(1):38-43.

McMillan M, Mackay WG, Williams CL, Shepherd AJ, Malcolm C, Weaver LT. Intrafamilial Genotyping of Helicobacter pylori from Faecal DNA. Gastroenterol Res Pract. 2011 2011:491035

Nahar S, Kibria KM, Hossain ME, Sultana J, Sarker SA, Engstrand L, Bardhan PK, Rahman M, Endtz HP. Evidence of intra-familial transmission of Helicobacter pylori by PCR-based RAPD fingerprinting in Bangladesh. Eur J Clin Microbiol Infect Dis. 2009;28(7):767-73.

Nomura A, Stemmermann GN, Chyou PH, Kato I, Perez-Perez GI, Blaser MJ. Helicobacter pylori infection and gastric carcinoma among Japanese Americans in Hawaii. N Engl J Med. 1991;325(16):1132-6.

Osaki T, Okuda M, Ueda J, Konno M, Yonezawa H, Hojo F, Yagyu K, Lin Y, Fukuda Y, Kikuchi S, Kamiya S. Multilocus sequence typing of DNA from faecal specimens for the analysis of intra-familial transmission of Helicobacter pylori. J Med Microbiol. 2013;62(Pt 5):761-5.

Osaki T, Konno M, Yonezawa H, Hojo F, Zaman C, Takahashi M, Fujiwara S, Kamiya S. Analysis of intra-familial transmission of Helicobacter pylori in Japanese families. J Med Microbiol. 2015;64(Pt 1):67-73.

Parsonnet J, Friedman GD, Vandersteen DP, Chang Y, Vogelman JH, Orentreich N, Sibley RK. Helicobacter pylori infection and the risk of gastric carcinoma. N Engl J Med. 1991;325(16):1127-31.

Pryczynicz A, Guzińska-Ustymowicz K, Kemona A, Czyzewska J. Helicobacter pylori infection and expressions of EGF, EGFR and c-erbB-2 proteins in gastric carcinoma. Folia Histochem Cytobiol 2009; 47: 447-51.Suerbaum S, Michetti P. Helicobacter pylori

doi:10.6342/NTU201902662

infection. N Engl J Med 2002;347:1175-86.

Puz S, Innerhofer A, Ramharter M, Haefner M, Hirschl AM, Kovách Z, Rotter M, Makristathis A. A novel noninvasive genotyping method of Helicobacter pylori using stool specimens. Gastroenterology. 2008;135(5):1543-51.

Rajindrajith S, Devanarayana NM, de Silva HJ. Helicobacter pylori infection in children. Saudi J Gastroenterol. 2009;15(2):86-94.

Roma-Giannikou E, Karameris A, Balatsos B, Panayiotou J, Manika Z, Van-Vliet C, Rokkas T, Skandalis N, Kattamis C. Intrafamilial spread of Helicobacter pylori: a genetic analysis. Helicobacter. 2003;8(1):15-20.

Suerbaum S, Michetti P. Helicobacter pylori infection. N Engl J Med. 2002 10;347(15):1175-86.

Vale FF, Vítor JM. Transmission pathway of Helicobacter pylori: does food play a role in rural and urban areas? Int J Food Microbiol. 2010;138(1-2):1-12.

Watari J, Chen N, Amenta PS, et al. Helicobacter pylori associated chronic gastritis, clinical syndromes, precancerous lesions, and pathogenesis of gastric cancer development. World J Gastroenterol. 2014 14;20(18):5461-73.

Yokota S, Konno M, Fujiwara S, Toita N, Takahashi M, Yamamoto S, Ogasawara N, Shiraishi T. Intrafamilial, Preferentially Mother-to-Child and Intraspousal, Helicobacter pylori Infection in Japan Determined by Mutilocus Sequence Typing and Random Amplified Polymorphic DNA Fingerprinting. Helicobacter. 2015;20(5):334-42.

doi:10.6342/NTU201902662