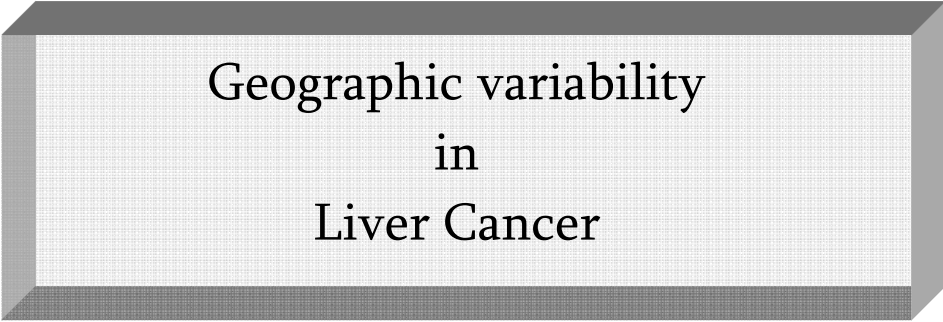Departament de Pediatria, d'Obstetricia i Ginecologia

i de Medicina Preventiva i Salut Pública

Programa de doctorat

Salut Pública i Metodologia de la Recerca Biomèdica

Facultat de Medicina

Universitat Autònoma de Barcelona

Geographic variability
in
Liver Cancer

TESI DOCTORAL presentada per Ramon Clèries Soler per a l'obtenció

del grau de doctor en Medicina Preventiva i Salut Pública

per la Universitat Autònoma de Barcelona

DIRECTORS DE LA TESI

Víctor Moreno Aguado

Josepa Ribes Puig

Barcelona, 2006

El Dr Víctor Moreno Aguado professor titular de la Universitat Autònoma de Barcelona i la Dra Josepa Ribes Puig epidemiòloga del Servei d'Epidemiologia i Registre del Càncer de l'Institut Català d'Oncologia.
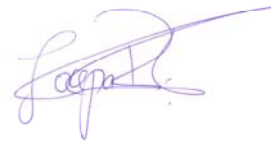
Certifiquen que: Ramon Clèries Soler, llicenciat en Ciències i Tècniques Estadístiques per la Universitat Politècnica de Catalunya, ha dut a terme sota la seva direcció la Tesi Doctoral titulada:

Geographic Variability in Liver Cancer.

I que aquesta Tesi compleix els requisits per tal de ser presentada i defensada.

Víctor Moreno Aguado   Josepa Ribes Puig

Barcelona, 19 de Juliol de 2006

Agraïments (Acknowledgments)

Primer de tot, vull agrair al Dr Víctor Moreno la seva ajuda i consells a l'hora de introduir-me en un programa de doctorat i poder dur a terme aquesta tesi. Ell ha estat una de les persones que al llarg d'aquests anys a l'ICO m'ha proporcionat l'interés per la recerca i per continuar aprenent. Mitjançant el seu suport he dut a terme la meva introducció a l'inferència bayesiana, tal com queda reflectit en aquest treball.

Als meus pares, Miquel i Carme, i al meu germà, Pere, i a la familia en general per el seu suport al llarg d'aquests anys. També al meu tiet, Roc Clèries, per tots aquests anys d'aprenentatge i empenta.

Als de la "colla" i "de vacances", per les sortides i caps de setmana, que sempre van bé per "cambiar d'aire": Siscu Surroca, Siscu Valls, Toni Surroca, David Estrada (Per cert, no haviem d'anar al "Serrallo"?) Santi, Judit, Eduard+Neus=Aina (*.tmp), Josep ("Larri"), Adela, Marcel, Anna, Joan Torrents, Alfred("-di"), Carles ("Litus"), Marialba, Jordi i Belén->Oriol, Toni ("Torelo") i Mireia, Pep Borràs, Ramon ("Xenxe"), Jordi Campos, Toni Ribera, Ferran, Xavi ("Xu"), Josep Albert ("JA"), Xavi ("Santiago") i al Pepe "Sepu".

Als meus companys de feina de l'ICO-SERC: Gina+Ester, Toni (TB), Guillem (WP), Xavi Castellsagué (XC), Carlos CAG, FX Bosch, Eli Luquín, Silvia (SS), Raquel Ib, Elena, Vilma, Roberto ("viva México"), Xavi Solé, Eli Guinó, Carly, Mireia, Jesús (Person-years: recordes el mític programa?), Paula, Erna ("Go Ahead Eagles") i Raquel In,

Cris, Meri, Yoli i Rebe per la seva disponibilitat a ajudar sempre que ha fet falta i per la seva empenta i ganes de treballar (A més, algun dia hem de dinar junts però que no sigui a la biblio....) i A na Laia Cabrera, na Sara Tous, na Gemma Prieto, na Laia Bruni i na Laura Esquius, les noies del petit cau de l'entrada del SERC perquè sempre he pogut compartir un somriure amb vosaltres quan ha fet falta. Voldria parlar de sopars però algú pensarà que sóc un pesat.....

A les meves excompanyes de pis, Magui, Laura i Núria i a en Carlos Crespo, Juanki, Cristinita, Juanbi, Néstor, Bea, Ester, Raquel i Aleix per els bons moments viscuts durant tots aquests anys que he compartit amb vosaltres, entre la llicenciatura d'estadística i cursos de doctorat.

Al Dr Toni Agudo per la seva ajuda en quant a la revisió d'aquesta tesi i per la seva disponibilitat en les consultes fetes al llarg d'aquests anys. Sempre recordarem les mítiques simulacions (perquè al SERC potser al cap de setmana també hi ha algú treballant). Però tinguem present que sempre podrem recordar la veu d'aquell que diu

"How much I know" (al costat de la màquina de café) o sempre trobarem un "some evidence suggests...".

I would like to thank to Dr Michelle Méndez for the English and Epidemiology review on this thesis. There was a thoughtful insight into every one of your comments and suggestions that made me think about epidemiologic concepts that I should review.

I would also thank to Lis Ellison for her helpful comments on the english review. But here is my suggested trip: Don't forget to visit Nuria, Queralbs and Ripoll.

To Tim Hahn Taylor, for the English review and also because he is a "unique" English teacher. Thanks for your classes and, I will try to "wake up" early next year..... (I still think that Peter Cushing has been one of the best actors ever, remember "The Man in the iron Mask" and, of course, the Hammer films)

Amongst many others I would like to thank especially the following people from Finland for their support and useful remarks, specially during summer of 2005 in Helsinki: Dr Eero Pukkala, Dr Risto Sankila and Dr Vesa Kanniainen (KATAL).

But my very special thanks for the great time during my short stage at the Finnish Cancer Registry is for Prof. Timo Hakulinen and Dr Tadek Dyba. We have not finished the trekking trails!!!! Next time in Catalonia?

Als Navarro, na Matilde i en Valentí, i a en Ramon Galiana per la vostra paciència i perquè espero dedicar-vos més temps d'ara en endavant.

A l'Àngels per la seva ajuda en l'elaboració de la tesi i per estar sempre disposada a donar un cop de mà. El seu treball ha estat fonamental per poder dur a terme aquest treball. Visca Alaior!!!!!!!!

També vull agraïr a na Laura Esteban (ja tenim PFC) per la seva ajuda en la preparació de bases de dades i per les subanàlisi que ha portat a terme sempre que li he demanat. Visca Tàrrega, però compte amb la Wonder!!!!

A na Laura Pareja perquè sempre que ha fet falta, tant en la revisió d'articles com en aquesta tesi, ha guardat una part del seu temps per ajudar-me. A na Lluïsa, perquè s'ha integrat dins l'equip Registre i perquè sempre puguem parlar amb la "rastes" a l'hora de dinar. A en Jordi Gàlvez, perquè quan ha fet falta ha programat i preparat la base de dades més complicada, i d'aquesta forma n'ha facilitat l'us. Vàrem iniciar conjuntament la pàgina de l'aplicació WAERS i ara anem per una nova versió que sense la seva ajuda seria impossible. Com no a en Xavi Sanz (SNOMED MEN) per ser un membre del Registre (perquè sempre m'ha recuperat els fitxers desapareguts) i mantenir WAERS.

Als qui em van introduir al món de la recerca ja fa més de 6 anys: En Joan Cos Trullàs i na Marta Pérez-Cassany. Els moments viscuts en aquell PFC a la Diplo em van servir per agafar aquest camí i espero que puguem continuar-lo junts ben aviat.

A en Pedro Delicado per les seves classes i per introduir-nos al món de l'estadística Bayesiana en el primer curs de doctorat que varem iniciar a la UPC. Els dilluns era un dia que ens reuníem na Lourdes Rodero, na Rosa Lamarca, l'Alfonso Buil i en José Miguel Martínez per tal de resoldre els problemes d'aquestes classes. A partir d'aquests problemes varem iniciar la nostra amistat.

To Joan Benach, Maria, Nano, Carles Muntaner, Pat O'Campo, Gabriel & Daniel, Yvonne Hinds and Dr. Samuel Noh for the great time at Toronto when I initiated   the first chapter of this thesis.

A na Cris i al meu amic i company de Doctorat, "lo" Joan Valls i a lo Pau. Amb ell vaig iniciar els cursos de doctorat a la UAB, i amb qui millor per inciar-los…ja que com diu la l'himne del CB Lleida "…és que la gent de Lleida és la millor".

A na Mar Torné i especialment al meu amic José Miguel Martínez, perquè sempre hi la seva ajuda i perquè sempre hi és quan el necessites. Sempre és agradable compartir coneixements i punts de vista sobre un tema en el qual hi estàs treballant, però si amb qui pots aprofundir en aquests coneixements és un amic, encara millor. Ara és la nostra!!!

A en Klaus Langohr per ser una persona pacient i per ajudar-me en el pas previ a aquesta tesi. Sempre ha sigut un plaer treballar i col·laborar amb tu i espero que poguem continuar treballant junts.

Vull agraïr molt especialment (encara que em repeteixi, once again) a na Cris, Rebeca i Pitus perquè quan ha fet falta (recordem el maleït RefMan) hi heu estat.

Als meus companys del departament d'Economia de l'Empresa de la UAB: Lina, César, Jordi i Dolors, per confiar en mi a l'hora de tirar endavant amb les classes que conjuntament he combinat amb la feina l'ICO-SERC i els cursos de doctorat. Ja fa uns quants anys que vareu confiar en mi i em vareu donar la oportunitat que sempre us agrairé.

I la Pitus? Bé, vull agraïr a la "boss", com li diem els del Registre, la Dra. Josepa Ribes Puig (Pitus) per els consells, la paciència i la confiança que ha tingut amb mi al llarg d'aquests anys. Aquests s'han repartit amb l'inici del RCC, l'ASEDAT, WAERS, PFCs i com no, "LA TESI". L'aprenentatge al llarg de tots aquests anys al costat seu m'han aportat l'interés per a continuar aprenent i estudiant. L'elaboració d'aquesta tesi no hagués estat possible sense la seva ajuda i sobretot paciència. Mai un moltes gràcies serà suficient.

Table of contents

List of abbreviations

AAIRs:	Age-adjusted incidence rates.
AAMRs:	Age-adjusted mortality rates.
Anti-HCV:	Hepatitis C virus Antibody.
ALT:	Alanine Aminotransferase.
AC:	Age-Cohort model.
AD:	Age-Drift model.
AE:	Age-Effect model.
AP:	Age-Period model.
APC:	Age-Period-Cohort model.
APC1:	Age-Period-Cohort model with autoregressive smoothing for the age, period and cohort parameters.
APC2:	Age-Period-Cohort model as APC1 with constraint on the second order differences related only with the age parameter.
APCH:	Annual Percent Change.
CI:	Confidence Interval.
CRI:	Credibility Interval.
DIC:	Deviance Information Criterion.
DPP:	Dirichlet Prior Process.
GP:	General Population.
HBcAg:	Hepatitis B core antigen.
HBeAg:	Hepatitis B "e" antigen.
HbsAg:	Hepatitis B surface antigen.
HBV:	Hepatitis B Virus.
HCC:	Hepatocellular Carcinoma.
HCV:	Hepatitis C Virus.
HIV:	Human Immunodefficiency virus.
IARC:	International Agency for Research on Cancer.
ICD:	International Classification of Diseases.
IFN:	Alfa Interferon.
IMRR:	Increase, in multiplicative scale, of the Relative Risk.
NPMLE:	Non-Parametric Maximum Likelihood Estimator.
U.S.:	United States of America.
U.K.:	United Kingdom.
RR:	Relative Risk.
OC:	Oral Contraceptives.
OR:	Odds Ratio.
pD:	Parameter dimension and "number of effective parameters".
PLC:	Primary Liver Cancer.
RR:	Relative Risk.
RR_Pooled:	Relative Risk Pooled.

RNA:	Ribonucleic acid.
SIR:	Standardized Incidence Ratio.
SMR:	Standardized Mortality Ratio.
SRF:	Scale Reduction Factor.
WBD:	Workers or Blood Donnors.
ZIP:	Zero Inflated Poisson Model.

List of tables

List of figures

Introduction

Liver cancer

Liver cancer is a disease in which cancer cells grow in the liver. There are two main types of liver cancer: hepatocellular carcinoma and cholangiocarcinoma[1].

The most frequent and important hepatic neoplasm is hepatocellular carcinoma (HCC), which is a malignant tumour derived from hepatocytes. In many parts of the world, in particular Africa and Asia, HCC poses a significant disease burden[2,]. In these high incidence regions, chronic infection with hepatitis B virus (HBV) is the principal underlying cause, with the exception of Japan which has a high prevalence of hepatitis C virus (HCV) infection. HBV vaccination has become a powerful tool in reducing cirrhosis and HCC, but implementation is still suboptimal in several high risk regions[2]. In Western countries, chronic alcohol abuse is also a major etiological factor[2].

The less frequent type of liver cancer, hepatic cholangiocarcinoma, has a different geographical distribution, with peak incidences in Northern Thailand. There, it is caused by chronic infection with the liver fluke, *Opisthorchis Viverrini*, which is ingested through infected raw fish[1].

Most cases of liver cancer are actually cancers that started in another organ (metastases). Because of its very high blood flow and many biological functions, the liver is one of the most common places for metastases to grow. Tumors that originally arise in the colon, pancreas, stomach, lung or breast can spread to the liver. In this instance, these tumors are the primary source of the liver cancer.

For that reason, the international variability in the diagnostic ability as well as in the coding and registration practices for liver cancer (primary, intra-hepatic biliary ducts,

metastases and liver tumors of uncertain nature whether primary or secondary) makes the interpretation of liver cancer time trends difficult[2].

National cancer registries generally list HCC as "primary liver cancer". Therefore, the term "primary liver cancer" (PLC) will be used through the thesis to refer to HCC.

Hepatitis B virus

In 1965, Dr. Blumberg who was studying haemophilia, found an antibody in two patients which reacted against an antigen from an Australian Aborigine. Later, the antigen was found in patients with serum type hepatitis and was initially designated as the "Australian antigen". Subsequent study has shown the Australia Antigen to be the hepatitis B surface antigen. Dr. Blumberg was subsequently awarded the Nobel Prize for his discovery. Initially there appeared to be three particles associated with hepatitis B infection: a large "complete" particle called the "Dane particle", a small circular 20nm particle and an oblong 40nm particle. Further research identified the Dane particle as the hepatitis B virion and the other two particles as excess surface protein. This former terminology is no longer used and the virus is described according to its structure[3].

Biology

HBV is a small DNA virus belonging to the group of hepatotropic DNA viruses known as hepadnaviruses. HBV consists of an outer envelope (see Figure 1), composed mainly of Hepatitis B surface antigen (HBsAg), and an internal core (nucleocapsid), which contains hepatitis B core antigen (HBcAg), hepatitis B "e" antigen (HBeAg), a DNA polymerase/reverse transcriptase, and the viral genome. The genome consists of a partly double-stranded circular DNA molecule of about 3200 base pairs with known sequence and genetic organisation. In recent years, HBV variants with mutations in viral genes and in some regulatory genetic elements have been detected in patients with HBV infection[4].

Figure 1. Hepatitis B Virion, Dane particle and HBsAg.

Hepatitis B surface antigen M — L, S
DNA polymerase
DNA (mostly double stranded), 3200 bp
Core antigens HBc, HBe
Protein kinase

42 nm

Dane particle

22 nm

22 nm

100–700 nm

HBsAg
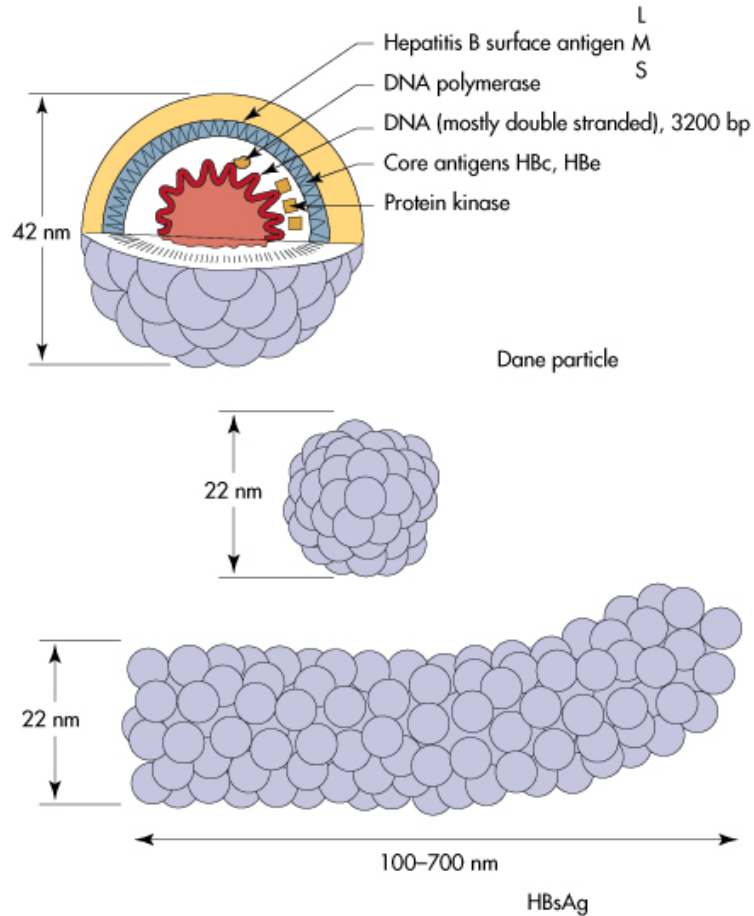
Source: Murray et. al., 2005[4]

Geographic variability

It has been estimated that 350 million people worldwide are infected with HBV. In areas of Africa and East Asia, 50% of the population may be seropositive, and between 5 to 15% may be chronically infected (carriers)[5]. All these subjects are at high risk to develop hepatocellular carcinoma[6;5].

Figure 2 depicts a map of the worldwide distribution of HBsAg prevalences. High endemic areas for HBV are found in East and Southeast Asia and Sub-Saharan Africa.

Europe is a low endemic area, with the exception of Southern and Eastern European countries, which are intermediate endemic areas.

Figure 2. Worldwide prevalence of HBsAg.
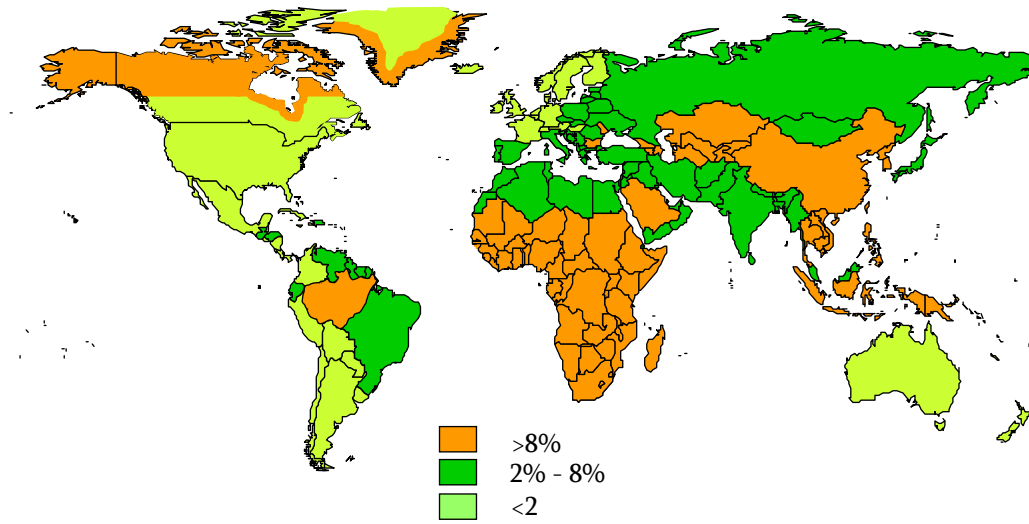
Source: Bosch et al, 2004[7]

Patterns of transmission

The pattern of transmission of HBV varies depending on the geographical area. In areas where persistent infection is highly endemic, transmission is mainly either perinatal, from a carrier mother to her newborn, or through close contact between children (horizontal transmission). In Asia approximately 40% of HBV carrier women of childbearing age are also positive for the HBeAg and these mothers have a 70% to 90% chance of infecting their newborn perinatally[6].

In many developed countries (e.g. Western Europe and North America), the pattern of transmission is different. In these countries, mother-to-infant and child-to-child transmission accounted for up to one third of chronic infections before childhood hepatitis B vaccination programmes were implemented. However, the majority of infections in these countries are now acquired during young adulthood by sexual activity, and injecting drug use.

Natural History

Infection acquired perinatally and in early childhood is usually asymptomatic. Approximately 30% of infection among adults present as icteric hepatitis and 0.1–0.5% develop fulminant hepatitis. Infection resolves in > 95% of adults with loss of serum HBsAg and the appearance of anti-HBs. Chronic infection is characterised by the persistence of HBsAg and anti-HBc, and by serum HBV-DNA levels detectable for more than 6 months using non-polymerase chain reaction (PCR) based assays[8].

Chronic HBV infection presents as one of three potentially successive phases known as *immunotolerant, immunoactive*, and *low-* or *non-replicative*. In the *immunotolerant phase*, serum HBsAg and HBeAg are detectable; serum HBV-DNA levels are high; and serum aminotransferases normal or minimally elevated. In the *immunoactive phase*, serum HBV-DNA levels decrease and serum aminotransferase levels increase. During this phase, symptoms may appear and flare-ups of aminotransferases may be observed. In some patients, these flare-ups are followed by HBeAg-anti-HBe seroconversion. The *non-replicative* phase follows with HBeAg-antiHBe seroconversion. HBV replication persists but at very low levels being suppressed by the host immune response. This phase is also termed the 'inactive HBsAg carrier state'. It may lead to resolution of HBV infection where serum HBsAg becomes undetectable and anti-HBs is detected. In some patients HBeAg seroconversion is accompanied by the selection of HBV variants that are unable to

produce HBeAg. A proportion of these HBeAg negative patients may later develop higher levels of HBV replication and progress to HBeAg negative chronic hepatitis[8].

The 'inactive HBsAg carrier state' is characterised by HBsAg and anti-HBe in serum, undetectable HBeAg low or undetectable levels of HBV DNA, and normal serum aminotransferases. Histology shows little or no necro inflammation and mild or no fibrosis (although inactive cirrhosis may be present if transition to an inactive carrier state occurred after many years of chronic hepatitis). The prognosis of the carrier state without cirrhosis is usually benign; but 20–30% of patients may undergo reactivation of hepatitis B. Acute flares of hepatitis are usually due to reactivation of HBV replication but can occur with superinfection with other hepatotropic viruses (such as HCV) or other causes of acute liver disease (e.g. drug toxicity, alcohol abuse). Some patients may develop HCC without cirrhosis, albeit less frequently.

In Western countries, about 1–2% of carriers become HBsAg negative each year; in endemic areas the rate of HBsAg clearance is lower (0.05–0.08% per year)[8].

Treatment

Hepatitis B can present as an acute, fulminant disease or in an asymptomatic chronic carrier. There is no recommended therapy for acute hepatitis but Lamivudine has been used in some cases. The treatments discussed here are for chronic hepatitis B which by definition is a persistently positive HBsAg for greater than six months.

Treatment is recommended for people with positive HBsAg and elevated alanine aminotransferase (ALT) and viral DNA levels. The treatment goal is to achieve seroconversion of HBsAg, which is rare, or loss of HBeAg which would mean less viral infectivity. Hopefully, this in turn would lead to lower risk of cirrhosis and liver

cancer. The two most commonly used treatments are Alfa Interferon (IFN) and Lamivudine[9].

IFN is a family of naturally occurring small proteins and glycoproteins which are products of immune cell response to a viral infection. The mechanism of action is unknown. It is thought that it inhibits viral replication, inhibits viral attachment, induces proteases or amplifies cytotoxic T-cell levels. Therefore, people lacking competent or with under developed immune systems do not response well to IFN. Subjects with high ALT and low pre-treatment viral DNA levels reflecting a good endogenous immune responses have a good prognosis with IFN. There is a loss of HBeAg and viral DNA in 20-40% of cases, and loss of HBsAg in 5-10%[9].

The other main treatment for HBV consists of nucleoside analogues, such as Lamiduvine. This treatment last has been tested in patients with chronic HBV in long term trials. There are now data on use of Lamivudine for up to four years[9]. In the initial study,100 mg per day of Lamivudine was given for one year. There was 72% normalization of ALT, 16% HBeAg loss or conversion and 55% improvement in histology. A two year study revealed 27-38% HBeAg loss with 52% undectable DNA. A three year study revealed 40% HBeAg loss, and a four year study revealed 47% HBeAg loss. The seroconversion of HBeAg is greater if ALT is more than 2-fold times normal levels. The side effects of long-term use are minimal with respect to pancreatitis and lactic acidosis[9].

Vaccination

Hepatitis B vaccine, if given before exposure, can prevent infection and disease in almost all individuals. The vaccine is highly effective when included in the infant immunization schedule, although it can be used at any age. By the year 2003, 138

countries (43% of the globe) had included HBV vaccination in their national infant immunization programs (Figure 3)[10].

Two types of hepatitis B vaccine exist, namely plasma-derived vaccine and recombinant DNA-derived vaccine[11]. Immunity in individuals given the recombinant DNA vaccine can be boosted with plasma-derived vaccine and vice versa. Seroconversion rates with the two vaccines are comparable. Plasma-derived vaccines have been shown to be safe, and transmission of HBV and other viruses, including the human immunodeficiency virus (HIV) and hepatitis C virus (HCV), has not been documented following extensive surveillance[11].

Figure 3. Countries using hepatitis B vaccines in their national infant immunization system (December 2003).
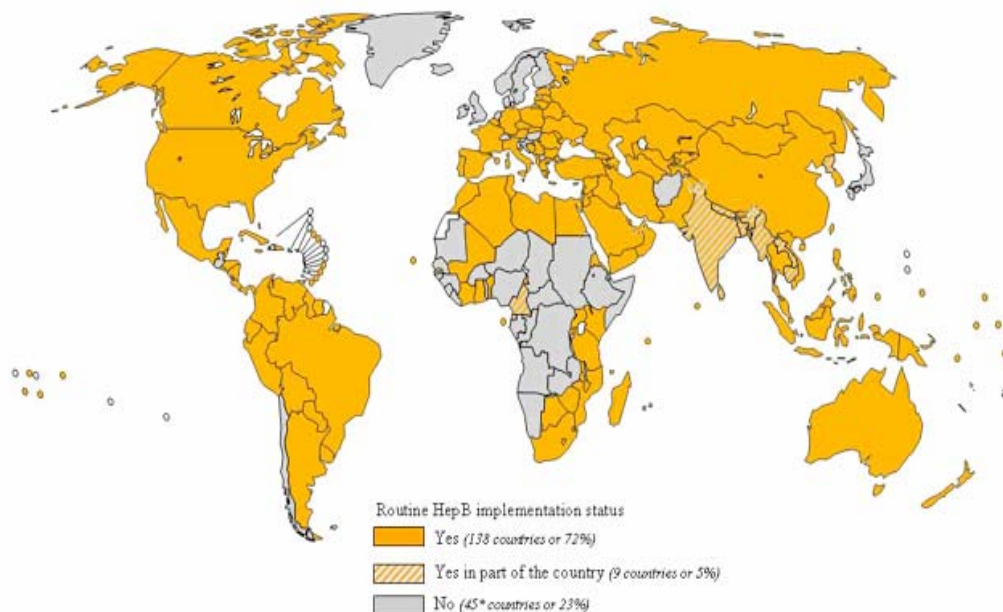
Source: World Health Organization. Infant Hepatitis B immunization programs.
http://www.who.int/vaccines-surveillance/graphics/htmls/hepb.htm[10]

In industrialized countries, hepatitis B virus is the major infectious occupational hazard of health workers, and most health care workers have received hepatitis B

vaccine[12]. Most industrialized countries screen every pregnant women for HBsAg, and treat infants of carrier mothers with specific hyperimmune globulin and hepatitis B vaccine[13,14].

Hepatitis C virus

After the discovery of the HBV virus, in the early 80s the hepatitis A virus was identified and testing for the antibodies became available. There were many patients with neither hepatitis A nor B who were then called "non-A non-B" hepatitis. Since some patients seemed to develop hepatitis after a blood transfusion, some health care workers referred to this type of hepatitis as "post transfusion" hepatitis. The HCV was cloned in 1989 and testing became available in early 1990. Many of the patients previously diagnosed as either non-A non-B or post transfusion hepatitis were retested, (or stored serum was tested) and found to have hepatitis C[3]. At present, hepatitis C is recognized as one of the most common types of hepatitis with up to 2% of the population being seropositive[15].

Biology

HCV is a single-stranded RNA virus in the Flaviviridae family. The genome is approximately 10,000 nucleotides and encodes a single polyprotein of about 3,000 amino acids. The structure of the hepatitis C virus is unknown; however, based on 3-D structures of related viruses, it is hypothesized that the virion is composed of an icosahedral lipid membrane with 2 glycoproteins (termed E1 and E2) that form heterodimers[4] (see Figure 4).

Figure 4. Proposed structure of the hepatitis C virus.

MODEL OF THE HUMAN HEPATITIS C VIRUS
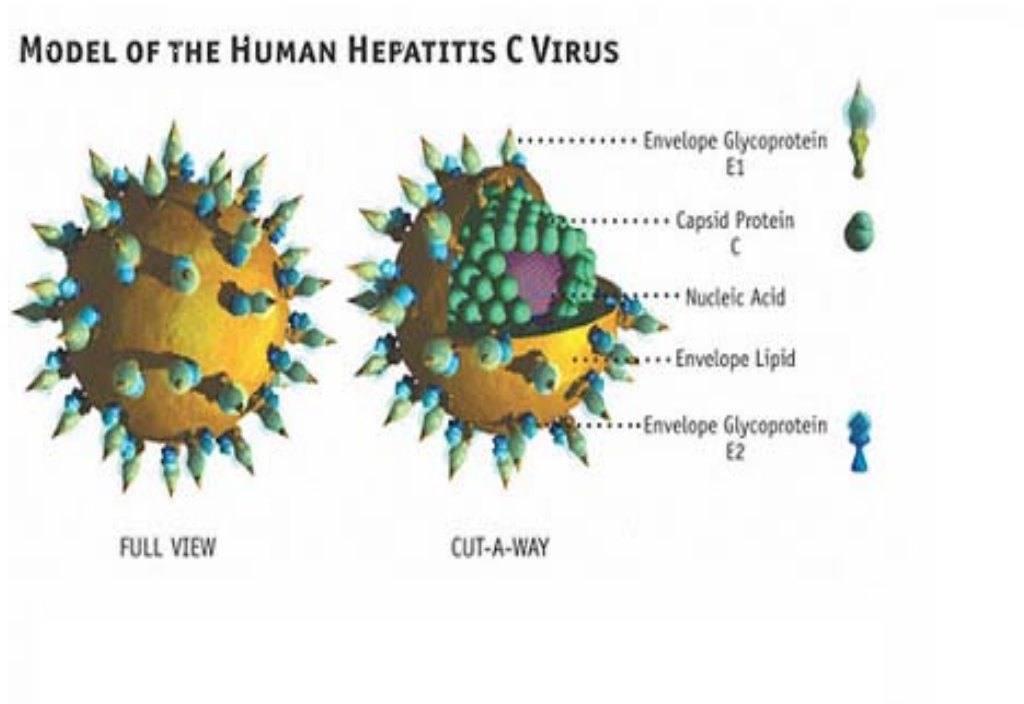
Source: Murray et. al., 2005[4]

Inside the viral membrane is thought to be an icosahedral nucleocapsid. The polyprotein is processed by host cell and viral proteases into three major structural proteins and several non-structural protein necessary for viral replication[4].

Geographic variability

It has been estimated that approximately 123 million people worldwide are infected with HCV[15]. Anti-HCV antibodies are found in 15-80% of HCC patients, depending on the patient population studied. HCV appears to be a major cause of HCC in Japan, Italy and Spain, but it seems to play a less important role in South Africa and Taiwan[1].

Although HCV is endemic worldwide, there is a large degree of variability in its geographical distribution, as Figure 5 shows. Countries with the highest reported

prevalence rates are located in Africa and Asia; areas with lower prevalence include the industrialised nations in North America, northern and western Europe, and Australia. Populous nations in the developed world with relatively low rates of HCV seroprevalence include Germany (0.6%), Canada (0.8%), France (1.1%), and Australia (1.1%)[15]. Low, but slightly higher seroprevalence rates have been reported in the USA (1.8%), Japan (1.5–2.3%), and Italy (2.2%)[15].

Figure 5. Worldwide prevalence of HCV infection.
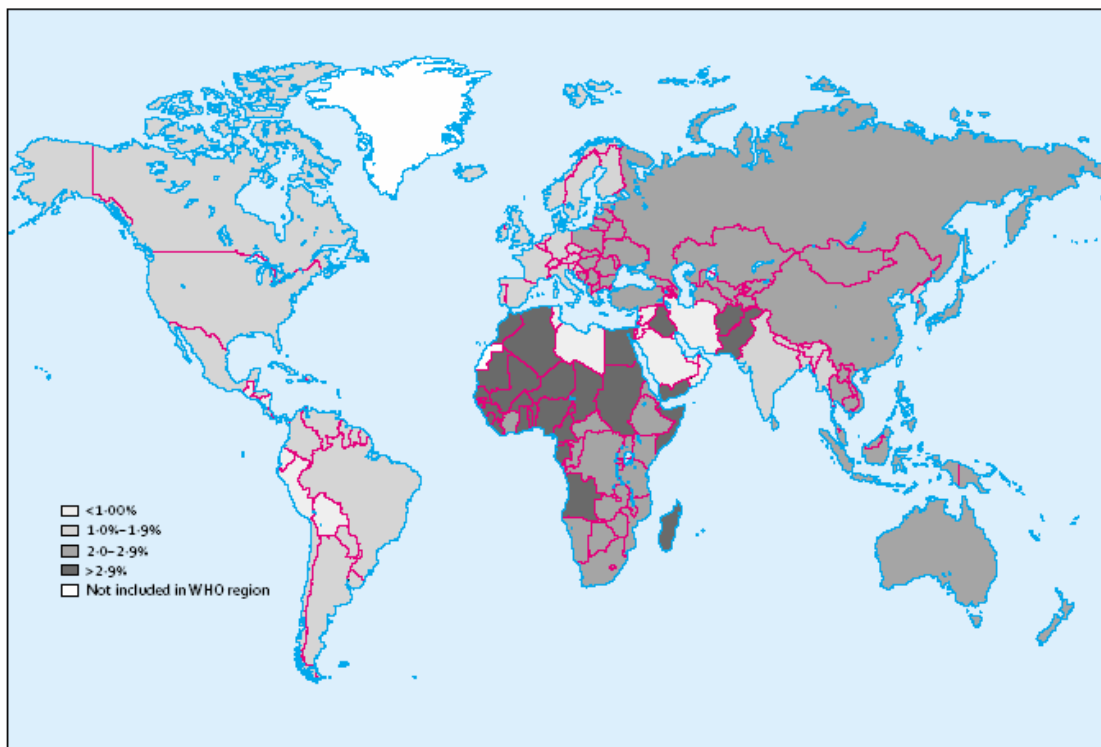
Figure: Estimated prevalence of HCV infection by WHO region

Source: Sheppard et al, 2005[15].

Patterns of transmission

HCV and HBV share similar routes of transimission. HCV is spread primarily by direct contact with human blood. Transmission through blood transfusions that are not screened for HCV infection, through the reuse of inadequately sterilized needles, syringes or other medical equipment, or through needle-sharing among drug-users, is well documented. Sexual and perinatal transmission may also occur, although less frequently[16].

In both developed and developing countries, high risk groups include injecting drug users, recipients of unscreened blood, haemophiliacs, dialysis patients and persons with multiple sex partners who engage in unprotected sex. In developed countries, it is estimated that 90% of persons with chronic HCV infection are current and former injecting drug users and those with a history of transfusion of unscreened blood or blood products[15].

In many developing countries, where unscreened blood and blood products are still being used, the major means of transmission are unsterilized injection equipment and unscreened blood transfusions[16]. In addition, people who use traditional scarification and circumcision practices are at risk if they use unsterilized tools.

Natural History

HCV-associated HCCs typically develop after 20-30 years of infection and are generally preceded by liver cirrhosis[1]. However, the natural history of chronic HCV infection can vary dramatically between individuals. Some will have clinically insignificant or minimal liver disease and never develop complications[17]. Others will have clinically apparent chronic hepatitis. Of these, some go on to develop cirrhosis, although the proportion is unknown. About 20% of individuals with hepatitis C who

develop cirrhosis will develop end-stage liver disease. Cirrhosis caused by hepatitis C is presently the leading indication for liver transplantation in the United States. Individuals with cirrhosis from hepatitis C are also at an increased risk of developing hepatocellular carcinoma[1].

About 85% of individuals acutely infected with HCV become chronically infected. Hence, HCV is a major cause of chronic (lasting longer than six months) hepatitis. Once chronically infected, the virus is almost never cleared without treatment. In rare cases, HCV infection causes clinically acute disease and even liver failure, however, most instances of acute infection are clinically undetectable[3,18,19].

Treatment

Antiviral drugs such as interferon taken alone or in combination with ribavirin, can be used for the treatment of persons with chronic hepatitis C, but the cost of treatment is very high[19,18]. Treatment with interferon alone is effective in about 10% to 20% of patients. Interferon combined with ribavirin is effective in about 30% to 50% of patients. Ribavirin does not appear to be effective when used alone[18,16].

There is no vaccine against HCV. Research is in progress but the high mutability of the HCV genome complicates vaccine development. Lack of knowledge of any protective immune response following HCV infection also impedes vaccine research. It is not known whether the immune system is able to eliminate the virus. Some studies, however, have shown the presence of virus-neutralizing antibodies in patients with HCV infection[17].

Several different genotypes of HCV with slightly different genomic sequences have been identified that correlate with differences in response to treatment with interferon [17].

Thesis motivation

At the beginning of the 21st century, PLC remains as the fifth most common malignancy in men worldwide, and ranks eighth in women[20]. At the end of the 20th century increase of both incidence and mortality from PLC was detected in some developed countries, which was related to a dominant cohort effect associated with previous HCV exposures[2]. However, the role of HBV infection in the development of PLC in these areas has been also established through cohort studies among HBV carriers[2].

In Catalonia, Drs. Ribes and Bosch of the Catalan Institute of Oncology carried out in the 90s a retrospective cohort study of 2,206 HBV carriers in the area of Barcelona, for which I was incorporated as a statistician in 1999. By 2003, with a mean follow-up of 20 years, there was determined an excess risk of mortality from PLC in men (Standardized Mortality Ratio (SMR): 14.1) and from liver cirrhosis for both sexes (SMR men: 10.5; SMR women: 7.2)[21]. There was high variability in relative risks (RR) of death from PLC among cohort studies carried out in different geographic areas (RRs range of variability: from 5.3 to 148)[2]. In order to explain the heterogeneity in RRs, an investigation into the geographic variability of PLC was initiated. This analysis was the beggining of this thesis.

The first objective of this thesis was to determine the sources of heterogeneity which led to variability in PLC risk in cohort studies among HBV carriers. Three of these studies were carried out in different European countries (U.K.[22,23], Italy[24] and Spain[21]). Geographic variability in the PLC incidence and mortality among European countries had been previously described[25], although the joint effect of both HBV and HCV prevalences in the incidence and mortality from PLC in Europe had not been established.

The second study of this thesis was performed with the aim to determine the role of HBV and HCV prevalences in the heterogeneity of PLC rates among European countries. High incidence and mortality rates from PLC were observed in some of these countries during the period 1978-92[26]. The increase of incidence rates in France and Italy, two Southern European countries, was related with HCV infection[26]. However, in Spain, a Southern European country with similar incidence rates and with similar HBV and HCV prevalences as those countries, no increase in PLC incidence was detected in that period[26].

The third study was performed under the hypothesis of similar pattern of PLC risk in Southern European countries. The aim of this last study was to evaluate the time trends of mortality and incidence rates due to PLC and chronic liver disease in Spain, during the most recent time period for which incidence data were available (period 1983-97).

Thesis

This thesis presents three studies for which the main objective is to describe PLC incidence and mortality in different geographic areas. Each one of these studies covers both epidemiological and methodological aspects. For each study, different statistical methods on the basis of the Bayesian inference have been proposed, evaluated and discussed.

A review of liver cancer epidemiology is presented in Chapter 1, providing the basis for the objectives and aims for each of the studies, which are detailed in Chapter 2.

Chapter 3 presents the first study, entitled "Meta-analysis of cohort studies of risk of liver cancer death among HBV carriers". This study has evaluated the effect of geographic area and study design on the variability in PLC mortality reported in several cohort studies of male HBV carriers. The statistical methods of this study have been focused on mixtures of probability distributions. Those methods have allowed the identification of the sources of heterogeneity in RRs.

Chapter 4 presents the second study entitled "Geographic distribution of primary liver cancer in Europe in 2002". This study is a geographic analysis of risk of PLC incidence and mortality in 38 European countries during 2002. By means of random effects models, PLC maps of incidence and mortality risks in Europe were obtained, taking into account the joint effect of both HBV and HCV seroprevalences on these risks.

The last study, in Chapter 5, entitled "Time trends in liver disease in Spain during the period 1983-97", describes incidence and mortality trends in hepatocellular carcinoma and cholangiocarcinoma jointly with mortality trends in liver cirrhosis in Spain. The statistical analysis involves autoregressive age-period-cohort models.

A general discussion of the three studies is presented in Chapter 6, whereas Chapter 7 concludes the thesis with the global conclussions of the three studies.

WinBUGS code for the statistical models, simulation analyses for prior distributions of the models and a subanalysis of the time trends of PLC in Spain have been also included in the Appendix.

Chapter 1

1. Epidemiology of primary liver cancer

1.1. Incidence of primary liver cancer

For the year 2002 it has been estimated that some 626,241 new cases of primary liver cancer (PLC) occurred worldwide, corresponding to 442,149 in men and 184,092 in women[20].

Table 1.1. Estimated number of new cases and PLC incidence rates by sex and geographical area for the year 2002.

Geographical Area	Men		Women	
	N	AAIRs	N	AAIRs
World	442,149	15.7	184,092	5.8
Developed countries	74,253	8.5	36,151	3.0
Developing countries	365,923	18.4	147,210	7.1
Africa				
Eastern	14,012	21.1	6,267	8.6
Central	7,744	27.8	4,571	13.4
Northern	2,351	4.2	1,442	2.2
Southern	1,072	7.0	469	2.5
Western	10,637	15.3	4,162	5.6
Asia				
Eastern	297,014	36.9	113,172	13.3
South-Eastern	35,691	18.3	12,221	5.7
South-Central	14,536	2.6	8,374	1.4
Western	3,051	4.6	1,480	2.0
Pacific Islands [a]	400	12.5	196	5.1
Europe				
Eastern	9,674	5.3	6,754	2.4
Northern	2,531	3.4	1,663	1.7
Southern	14,021	11.6	6,497	4.0
Western	9,077	6.2	3,401	1.7
Americas				
Caribbean	1,433	8.2	911	4.5
Central	2,197	4.9	2,503	4.9
Southern	5,036	3.7	4,616	2.8
Northern [b]	11,058	5.3	5,152	1.9
Australia & New Zealand	622	3.9	239	1.3

Source of data: Globocan 2002[20]

[a] Melanesia, Micronesia and Polynesia; [b] United States and Canada; AAIRs: Age Adjusted Incidence Rates (world-standard population) per 100,000 person-years;

Table 1.1 presents the estimated number of PLC cases per year and the age-adjusted incidence rates (AAIRs) per 100,000 person-years by sex and geographic area. The data have been obtained through Globocan 2002[20], which presents a combined analysis of reports from the population-based cancer registries and the World Health Organization Mortality Databank[27]. In men, the highest AAIRs are located in Eastern Asia (AAIR: 36.9), followed by Central (AAIR: 27.8) and Eastern (AAIR: 21.1) Africa, whereas the lowest rates are found in South-Central Asia (AAIR: 2.6), Northern Europe (AAIR: 3.4) and Southern America (AAIR: 3.7). Southern Europe, where Spain is located, presents intermediate rates (AAIR: 11.6). The corresponding distribution of AAIRs among women follows a similar geographical pattern.

Figure 1.1 depicts the worldwide distribution by country of the AAIRs as annual averages per 100,000 in men[20]. The standard population used for   adjustment is the world standard, which tends to increase rates in countries with young populations and to decrease rates in countries with older populations with respect to the crude rates.

Figure 1.1. Geographical distribution of the PLC AAIRs among men.

AAIRs: Age Adjusted Incidence Rates of PLC per 100,000 person-years.
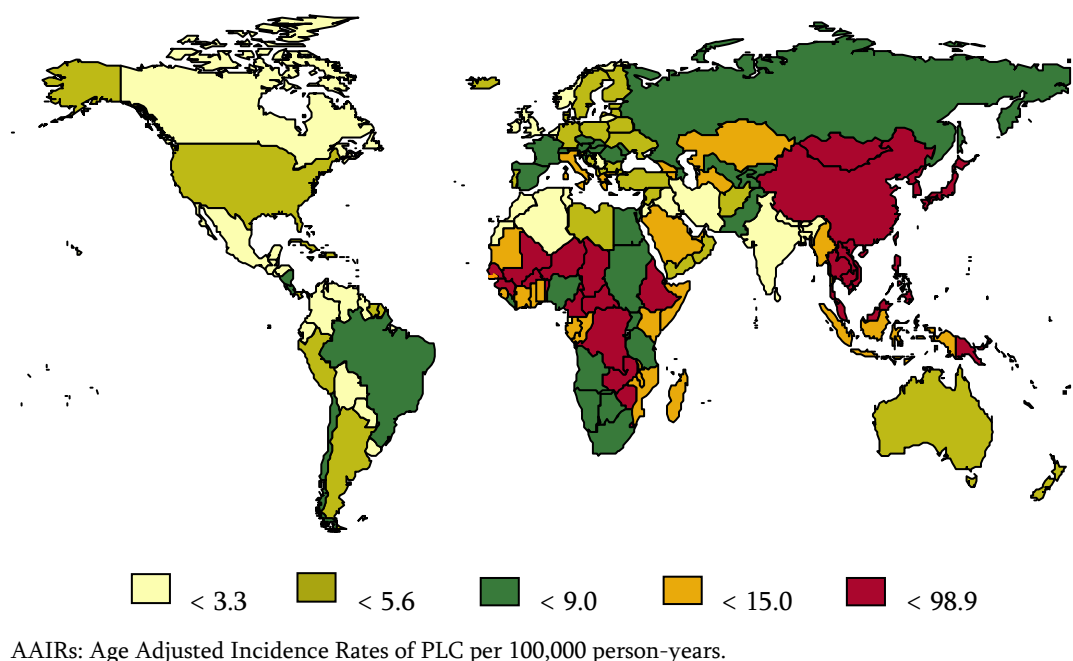
1.1.1. Age-specific incidence rates

In the most high risk areas, such as Southeast Asia (Qidong in China) or the West Coast of Africa (Bamako in Mali), PLC rates increase after 20 years of age and peak or stabilize at the age 50 and above (*Figure 1.2*). In these countries PLC is not a rare event at ages 20 to 35. Still the incidence in Qidong is substantially higher at each age group than the corresponding incidence in Mali -a high risk country in Africa- and the mean age of occurrence is significantly shifted towards younger age groups[28;7].

Figure 1.2. PLC AAIRs in men in selected high risk populations.
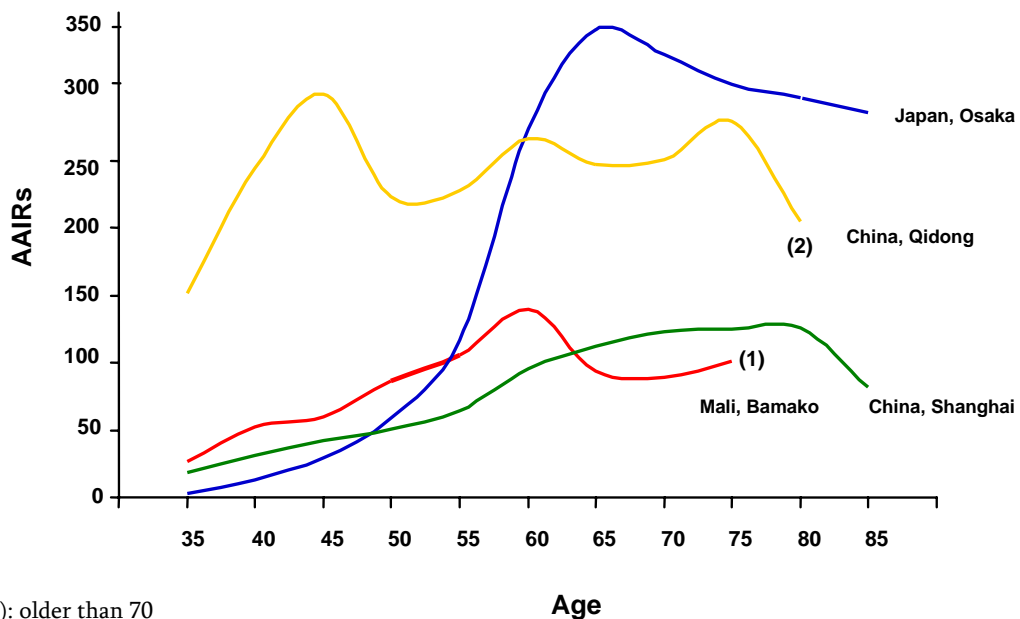
(1): older than 70
(2): older than 80
AAIRs: Age-adjusted incidence rates of PLC per 100,000 men-years

A second pattern is observed in high risk countries or regions that have recently experienced substantial degrees of development. For example in the Shanghai Cancer Registry, rates are lower than the corresponding rates in Qidong and the age-specific incidence increases steadily with age, a pattern similar to that usually observed in low

risk areas. In the Osaka Cancer Registry in Japan, the incidence among men increases after the age range 45 to 50 and reaches a plateau at 65 years of age[28] and above.

Figure 1.3 shows the AAIRs of PLC among men in selected European countries and ethnic groups in the U.S. The higher global risk in the Mediterranean countries in Europe is reflected in all age groups, most notably in the elderly. The age specific pattern among non-Hispanic White populations in the U.S. is similar to that observed in Europe. Hispanic Whites in the U.S. show a peculiar trend with steep increases after age 60, without a clear explanation of the risk factors that operate in these groups of older migrants[28,29].

Figure 1.3. AAIRs of PLC in men in selected populations in Europe and in the U.S.
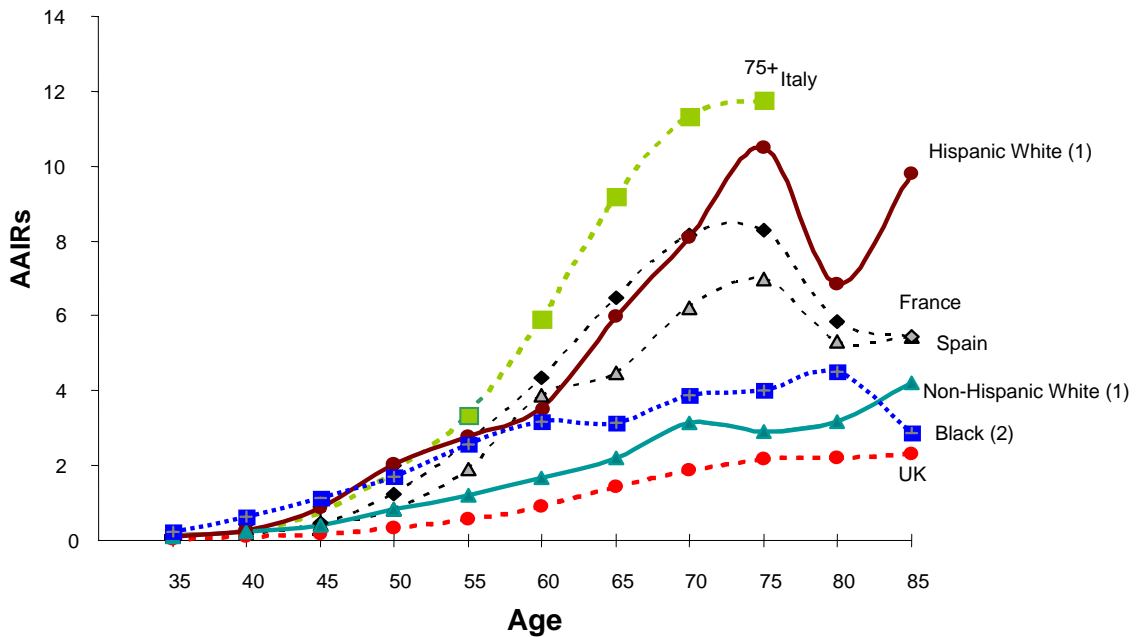
(1) California (Los Angeles), California (San Francisco), New Mexico;
(2) SEER program: Connecticut, Iowa, New Mexico, Utah, Hawaii, California (San Francisco Bay), Michigan (Detroit), Georgia (Atlanta), Washington (Seattle)

1.1.2. Sex ratios

It has been reported that the correlation between AAIRs of PLC in men and women is extremely high in European cancer registries (Correlation Coefficient = 0.953, p<0.001, see Figure 1.4)[30]. Worldwide, the range in the sex ratio of AAIRs is 1.3 to 3.6 [20], which reflects the excess of PLC incidence among men compared to women. In high risk countries, sex ratios tend to be higher, and the male excess is more pronounced below 50 years of age. Migrant populations also show a shift in the sex ratio values. Japanese populations in the U.S. show a fairly steady sex ratio between 2 and 3 in the age groups above 50, whereas Japanese populations in Japan show excess risk between 3 and 5 in the same age groups. In the age groups below 50, the sex ratios of PLC among Japanese in Japan range between 7 and 10. Among Japanese in the U.S., the incidence of PLC among women below age 50 is very rare and the sex ratio calculations are unreliable[28]. In populations with low incidence, the highest sex ratios occur later, at around 60-70 years of age[31].

Figure 1.4. Correlation between estimated PLC AAIRs in men and women.
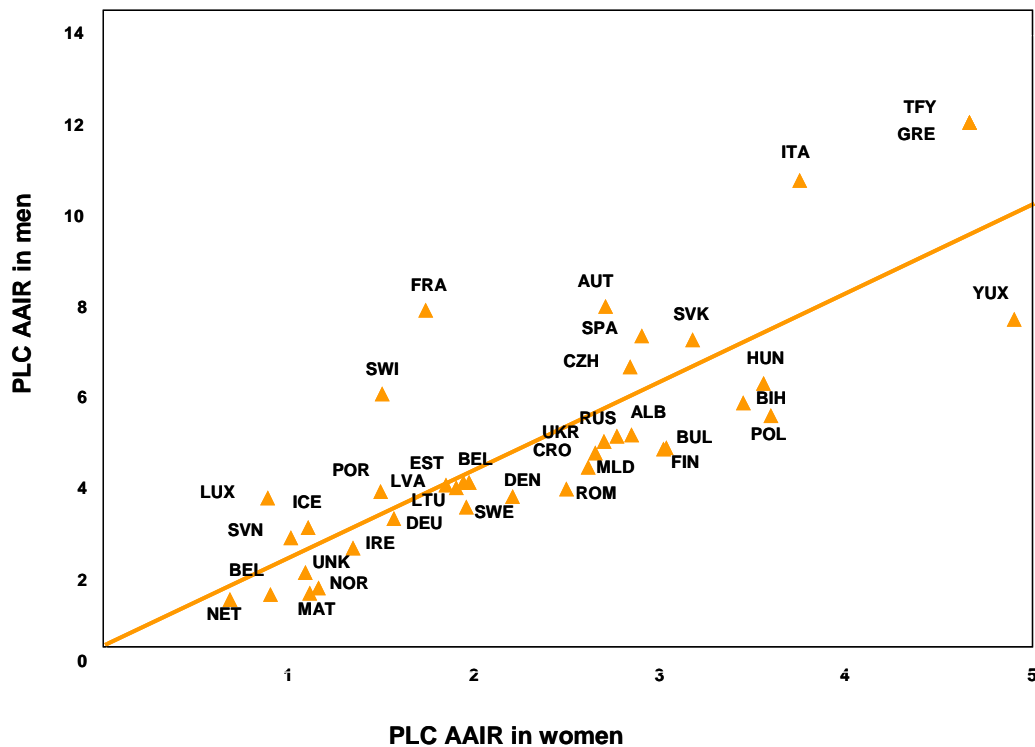
A satisfactory biological explanation for the observed sex ratio has not been identified. Several hypotheses have been investigated, including the interaction of testosterone with the HBV cycle or the impact of sex-specific exposures such as alcohol or tobacco in some cultural environments[32,33].

1.2. Mortality from liver cancer

Table 1.2 shows age-adjusted PLC mortality rates (AAMRs) obtained through Globocan 2002[20]. In spite of the limitations of mortality data, AAMRs follow a geographical pattern consistent with incidence data[20,27] (*Table 1.1*). PLC is a highly lethal tumor with an annual fatality ratio around 1, indicating that most cases do not survive one year. In Europe for the period 1990-1999[34] and in the U.S. for the period 1995-2000[35], population-level PLC survival rates were analyzed using data from population-based cancer registries. The five-year relative survival rates (mortality from PLC adjusted for mortality from competing causes) were 6.5% and 8.3% respectively. There is little difference in survival rates according to sex, suggesting a similar distribution of stage at diagnosis. In developing countries PLC is inevitably fatal.

Table 1.2. Estimated number of deaths and PLC mortality rates by sex and geographical area for the year 2002.

Geographical Area	Men		Women	
	N	AAMRs	N	AAMRs
World	**416,926**	**14.9**	**181,486**	**5.7**
Developed countries	71,153	8.0	38,083	3.0
Developing countries	343,956	17.4	142,728	6.9
Africa				
Eastern	13,805	20.8	6,180	8.5
Central	7,613	27.3	4,495	13.2
Northern	2,318	4.1	1,420	2.2
Southern	1,026	6.7	450	2.4
Western	10,454	15.1	4,093	5.5
Asia				
Eastern	272,778	33.9	104,715	12.3
South-Eastern	33,514	17.2	11,555	5.4
South-Central	13,873	2.5	7,969	1.4
Western	2,914	4.4	1,466	2.1
Pacific Islands [a]	190	4.9	379	12.1
Europe				
Eastern	10,670	5.8	7,839	2.7
Northern	2,401	3.1	1,809	1.8
Southern	12,518	10.1	7,011	4.1
Western	10,546	6.9	4,692	2.2
Americas				
Caribbean	1,499	8.5	1,049	5.2
Central	2,975	6.7	3,420	6.7
Southern	7,845	5.8	7,519	4.6
Northern [b]	9,229	4.4	5,319	1.9
Australia & New Zealand	573	3.5	292	1.5

[a]Melanesia, Micronesia and Polynesia; [b] United States and Canada;
AAMRs: Age adjusted mortality rates (world-standard population) per 100,000 person-years.

1.3. Liver cancer in ethnic groups and migrant populations

Cancer registries in the U.S. report the incidence of PLC by ethnic origin. The lowest incidence rates are consistently found among Caucasian Whites (3.8 in men and 1.4 in women). Gradually increasing rates are found in Japanese (5.5 in men and 4.3 in women), Black (7.1 in men and 2.1 in women), Hispanic White (9.8 in men and 3.5 in women), Filipino (10.9 in men and 2.4 in women), Chinese (16.2 in men and 5.0 in women) and Korean (20.7 in men and 10.4 in women) ethnic groups. Among women, the high risk ethnic groups present a 2 to 5-fold higher PLC AAIRs than the rates observed in non-Hispanic White women[28]. Several studies have been conducted among migrant populations comparing their PLC mortality rates to the rates of the host population. *Table 1.3* summarizes most of the published studies in which country of origin is usually a high risk area and the host population rate is taken as a reference[36-41].

Table 1.3. Relative risks of death due to PLC in migrant populations as compared to host populations.

Host Region	Region of Origin	RR men	RR women	RR both sexes
U.S.	China	3.3 - 10.9	1.1 - 4.3	-
Canada	China	7.7– 10.5	2.5 - 3.0	-
Australia	China	5.2	2.6	-
	East Asia	-	-	8.9
	Southeast Asia	-	-	10.0
	Near East	2.5	0.3	-
France	East Asia	1.3	3.3	
	Southeast Asia	2.5	1.6	-
England & Wales	West African	31.6	5.4	-
	East African	1.1	1.8	-
	Caribbean	5.3	3.2	-

RR: Relative Risk
Source of data: Hanley AJ *et al.* 1995[36], Khlat M *et al.* 1993[37], Bouchardy C *et al.* 1994[38], Fang J *et al.* 1996[39], Grulich AE *et al.* 1992[40], McCredie M *et al.* 1999[41]

Among men, a 1.3 to 10.9-fold excess of PLC mortality is observed among Asian migrants to North America, Australia or Europe. A 31.6-fold higher risk has been reported among West African migrants to England & Wales and a 5.3-fold higher mortality rate among Caribbean migrants. Among women, perhaps due to the small number of deaths, the results of these studies are not consistent, and the excess of PLC mortality among migrants ranges from 1.1 to 5-fold[36-39,41].

1.4. Trends in PLC incidence and PLC mortality rates

During the last two decades, increases in PLC incidence rates have been reported in Australia, Central Europe[26,42], the United Kingdom[43], Japan[44], and North America[45,46]. Decreasing trends were reported from Chinese populations in Singapore and Shanghai, India, Sweden and Spain[26]. In a study conducted during 1976 to 2003 in the U.S., the consistency in PLC trends was investigated using three different sources of information: liver cancer hospitalization rates, incidence rates from cancer registries and mortality data[45,46]. In this study significant increasing trends were found for Black, White and Hispanic populations. Analyses of the components of the time trends suggested a predominant cohort effect[45,46].

International trends in PLC mortality have also been evaluated. Among males, increases in mortality from PLC have been reported in the U.S., Japan, Australia, Scotland, France and Italy, while decreasing trends have been reported in the U.K.[47,48]. Trends among women are largely similar. Increases in cholangiocarcinoma have also been reported in the U.S., Japan, England and Wales, Australia, Spain and Scotland, and among women in the U.S., Australia and England and Wales[47,49,50].

However, international variation in diagnostic ability as well as in the coding and registration practices for PLC (primary, intrahepatic biliary ducts, metastases and liver tumors of uncertain nature as if primary or secondary) makes the interpretation of long-term time trends difficult. *Table 1.4* shows coding recommendations for liver disease from the 8th to 10th revisions of the International Classification of Diseases (ICD). For PLC and liver metastasis those changed slightly from the 8th to the 9th revisions of the ICD implemented after 1965 and 1975 respectively[51,52]. In 1992 the 10th revision[53] was introduced with major changes including codes for hepatoblastoma (C22.2), angiosarcoma (C22.3), other liver sarcomas (C22.4) and other inespecified carcinomas (C22.7).

Table 1.4. Coding recommendations for liver disease from 8th to 10th revisions of the International Classification of Diseases (ICD)[51-53]

	ICD-8	ICD-9	ICD-10
Primary liver cancer	155.0	155.0	C22.0
Intrahepatic bile duct (cholangiocarcinoma)	155.1	155.1	C22.1
Hepatoblastoma	-	-	C22.2
Angiosarcoma	-	-	C22.3
Other liver sarcomas	-	-	C22.4
Other unspecified carcinomas	-	-	C22.7
Liver tumor unspecified if primary or secondary	197.8	155.2	C22.9
Gallbladder	156.0	156.0	C23
Extrahepatic bile duct	156.1	156.1	C24.0
Ampulla of Vater	156.2	156.2	C24.1
Overlapping lesion of billiary tract	156.8	156.8	C24.8
Billiary tract, NOS	156.9	156.9	C24.9
Liver, specified as secondary	197.7	197.7	
Liver cirrhosis	571	571	K74

NOS: Not specified ; **ICD-8:** International Classification of Diseases, 8th revision; **ICD-9:** International Classification of Diseases, 9th revision; **ICD-10:** International Classification of Diseases, 10th revision.

Therefore, if properly done, in countries with developed health systems the predictable impact of changing codes for liver cancer should be, if any, a small reduction in incidence as a result of decreasing the number of liver metastases misclassified as PLC. For the majority of high PLC risk countries, the relative impact of diagnostic ability, including access to medical care as well as technology, should be considerably more important than the variability attributed to coding practices.

Another source of variability in the health statistics is inter-country differences with regard to the number of entries from death certificates that are routinely extracted and coded. In some countries several diagnoses are processed and additional rules apply in the assignment of the cause of death[54]. This is particularly important in PLC because of the high frequency of concurrent liver cirrhosis (over 90% in most populations) and its clinical complications, many of which can lead to an immediate

cause of death. These sources of variability should be considered when time trends are calculated and international comparisons are to be made.

Of particular concern in some countries is the likely impact of migrants from high risk countries. These populations are often visible in the health system at the time of diagnosis, but are less likely to be counted in the census, and constitute an important component of the number of PLC cases[36,37,39-41].

1.5. Risk factors for liver cancer

The etiology of PLC has been largely established, and *Table 1.5* shows current estimates of the attributable fractions for the main risk factors by three geographic areas[7,55]. In developing countries a predominant role of HBV infection in the development of PLC has been described, whereas in developed countries PLC arises in cirrhotic livers due to HCV infection or alcohol intake[56]. The role of each risk factor will be described in subsections that follow.

Table 1.5. Risk factors of PLC and estimates of the univariate attributable fractions (%).

	Europe & U.S.		Japan		Africa & Asia	
	Estimate	Range	Estimate	Range	Estimate	Range
Hepatitis B Virus	22	4-58	20	18-44	60	40-90
Hepatitis C Virus	60	12-72	63	48-94	20	9-56
Alcohol	45	8-57	20	15-33	--	11-41
Tobacco	12	0-14	40	9-51	22	--
Oral Contraceptives	--	10-50	--	--	8	--
Aflatoxin	Limited exposure		Limited exposure		Important exposure	
Other risk factors	<5	--	--	--	<5	--

Sources of data: Llovet *et al*[55] and Bosch *et al*[7]

1.5.1. Hepatitis viruses

Overall 75 to 80% of the PLC cases can be related to persistent viral infections with either HBV (50-55%) or HCV (25-30%)[57]. Worldwide, strong geographic correlations have been found between the incidence of PLC and the prevalences of hepatitis B surface antigen (HBsAg) (correlation coefficient: 0.67, $p<0.001$) or hepatitis C virus antibody (anti-HCV) (correlation coefficient: 0.37, $p<0.001$)[30].

The role of chronic infection with HBV in the etiology of PLC is well established. Cohort studies conducted worldwide yielded relative risk estimates of 5.3 – 148, with the highest values reported from China and Taiwan[21-23,58-72]. Age at HBV infection is a key determinant of the risk. In developed areas at low risk of PLC, HBV acquisition at birth or during childhood is rare and most of the viral infections are acquired in adolescence or adulthood, through sexual contacts, blood transfusions or other invasive procedures under non sterile conditions[73]. In these populations, the impact of cofactors such as alcohol and tobacco is reflected only in the incidence of PLC in the most advanced age groups. In contrast, mother-to-child HBV transmission and HBV exposures in the first years of life are typical of most of the high risk countries.

More specifically, one possible explanation for the differences in incidence rates in China and West Africa, may be related to age at infection. In many Asian countries, there is a high prevalence of HBsAg carrier mothers who also express Hepatitis B "e" antigen (HBeAg) and remain HBV-DNA positive throughout the reproductive years[74-76]. In high risk areas in Africa, the mother-to-child transmission rate is somehow lower and the child-to-child mode of transmission predominates. The impact of co-factors, perhaps with the exception of Aflatoxin, is likely to be marginal in the high risk countries.

The reduction of risk observed in migrants to lower-risk host populations probably reflects a combination of the reduction in the mother-to-child transmission of HBV linked to mixing populations, the reduction of exposure to contaminated blood products during medical interventions or other invasive medical or ritual procedures, the lower prevalence of HBV (and possibly HCV) among siblings and, more recently, the impact of the massive introduction of HBV vaccines.

Iatrogenic exposures to HCV during massive public health interventions have been demonstrated and the subsequent risk of PLC is now being expressed. *Figure 1.5* shows the prevalence of HCV antibodies in three populations by age groups: the

Japanese population, the U.S. population and in the general population in Egypt[77,78]. The sets of data are consistent with predominant cohort effects. In Japan, increasing PLC incidence and mortality trends since the early 1970s have been related to exposure of the population to HCV. In Japan the cross sectional HCV antibody prevalence decreases with younger age (cohort effect)[78]. In the U.S., the predominant cohort component of the increases in PLC incidence and mortality have been interpreted as a long term consequence of important HCV exposure in the period 1960-70 through contact with contaminated blood and syringe exchange in the relevant generations[79]. It is predictable that the number of PLC cases generated by the pool of HCV carriers will continue to increase for some time in the U.S.. The impact of the introduction of generalized HCV testing of blood products and the public awareness of the risks of HCV linked in part to the Acquired Immune Deficiency Syndrome (AIDS) education campaigns, should result in lower HCV related liver cancer in the more distant future.

In Egypt, a documented epidemiological study traced the source of the infection to the massive treatment campaigns against schistosomal infestation conducted in the interval 1920-1970 in the general population. The treatment typically included several courses of intramuscular or intravenous drugs against schistosoma that were delivered under non-sterile conditions[80]. The HCV antibody prevalence in the population is one of the highest registered and is consitently high accross age-groups (*Figure 1.5*)[77].

Figure 1.5. Hepatitis C virus antibody prevalence by age groups in three countries.
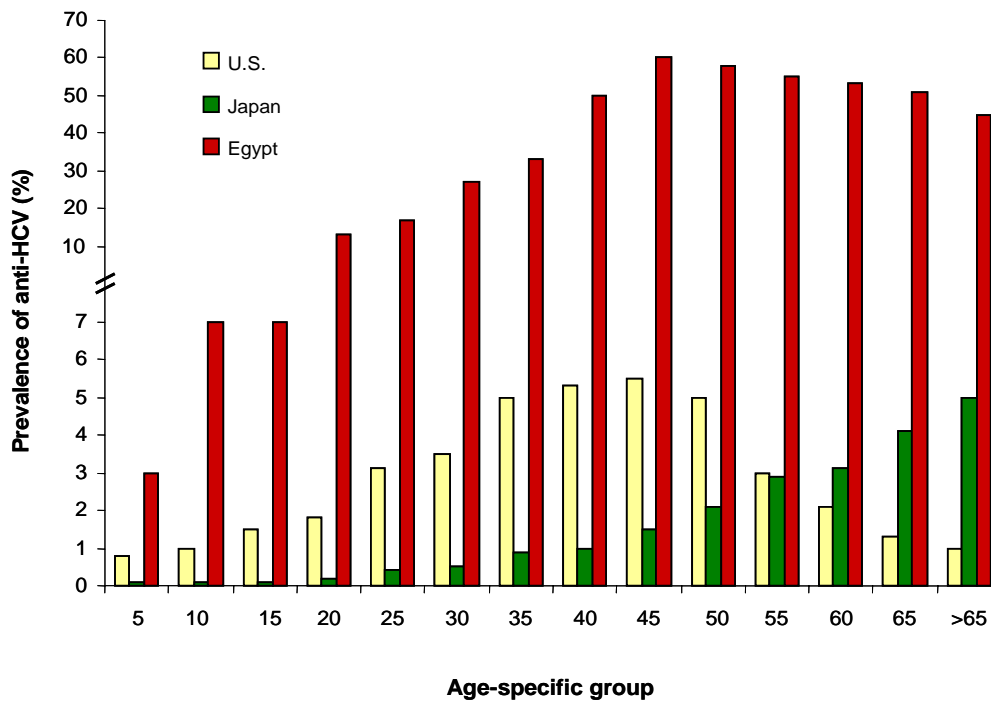
Figure 1.6 shows the impact of the two viral types in liver cancer in Japan and the United States. *Figure* 1.6A shows the increased number of deaths from PLC in the interval 1975-92 in Japan. While the prevalence of HBV related PLC remains fairly constant over the period, the bulk of the increase seems to be related to HCV-linked PLC[81]. In the United States the number of hospitalized PLC cases linked to HCV has increased dramatically compared to HBV-related or alcohol-related PLC cases in the period 1993-1998 *(Figure 1.6B)*[82].

Figure 1.6. Temporal trends in age-adjusted PLC hospitalization rates in the U.S. and time trends in PLC mortality rates in Japan by cause.

A. Japan

Figure reproduced from Okuda K. 1997[61]

B. U.S.

Reproduced from El-Serag HB *et al.* 2000[62]
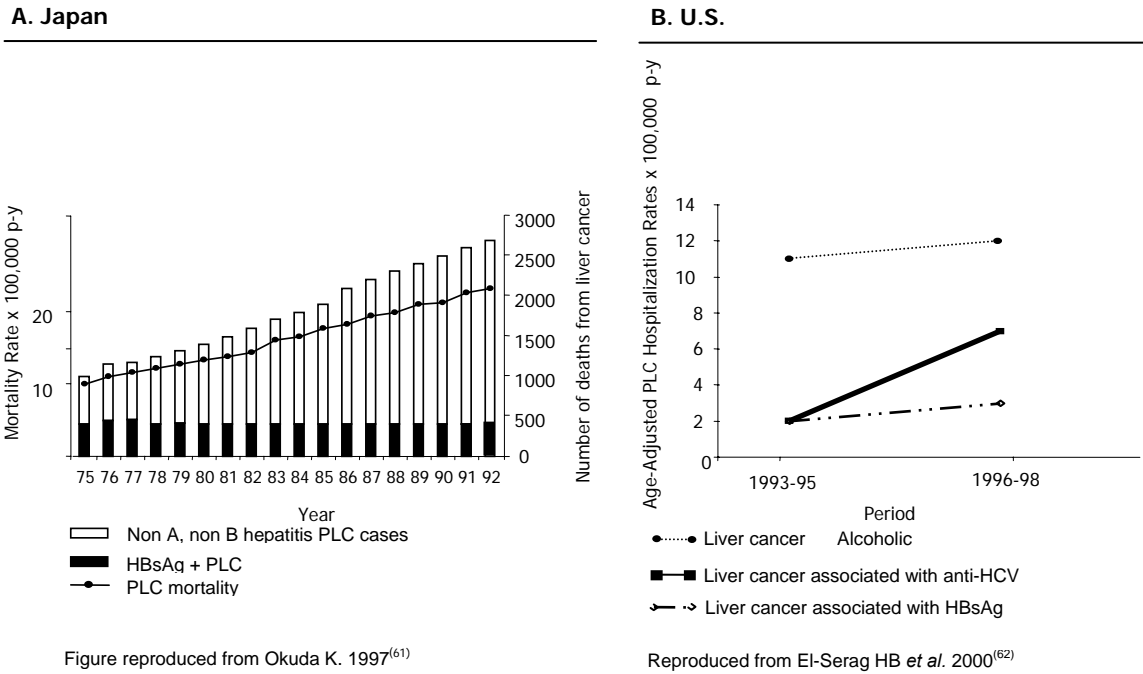
HCV: Hepatitis C Virus; **PLC:** Primary liver cancer; **HBsAg:** Hepatitis B Surface Antigen; **p-y:** person-years

More precise estimates of the relative contribution of viral infections to PLC incidence are provided by the introduction of more advanced detection methods in epidemiological studies. A large collaborative study in Europe, using polymerase chain reaction technology found that among those PLC cases that were negative for HBsAg or anti-HCV an additional 33% were positive for HBV DNA and 7% positive for HCV RNA[83]. This observation has been repeatedly confirmed and the trend suggests that in countries where HBV is common, the presence of HBV DNA among HBsAg negative PLC cases is higher than that found in European cases. A meta-analysis on viral factors and PLC reported summary odds ratios (OR) for HBsAg status (positivity or negativity) combined with anti-HCV or HCV RNA status (Table 1.6), suggesting a synergism of the two viral infection in the causation of PLC[84].

50

Table 1.6. Odds Ratios and their 95% confidence intervals for viral factors and PLC risk.

		anti-HCV or HCV RNA	
		Negative OR (95% CI)	Positive OR (95% CI)
HBsAg	**Negative**	*Reference*	17.3 (13.9 - 21.6)
	Positive	22.5 (19.5 - 20.6)	165 (81.2 – 374)

OR: Odds Ratio; 95% CI: 95% Confidence Interval;
Source: Donato *et al.,* 1998[84]

The impact of variants of HBV or HCV is now being described. There is substantial international variability in the prevalence of specific viral genotypes, and there is some evidence that genotypes may modulate the risk of progression to severe liver disease and probably PLC[85]. Of the HBV types described, type A is predominant in the U.S. and the Caribbean, type D in Europe and in the Middle East, and types B and C predominate in Asia. In Africa, types A, C and D are equally prevalent[86]. Likewise, the prevalence of type 1b of HCV in PLC cases ranges from 50% in England and Germany to 70-90% in Italy and Spain[83].

This geographic variability among variants of HBV could partially explain the heterogeneity in RRs of PLC observed in studies of HBV carriers. This hypothesis should be investigated taking into account effects of study design.

The heterogeneity of HBV and HCV variants warrants research describing how the natural history of these variants may affect viral spreading and carcinogenesis. In previous cohort studies of HBV carriers high variability between RRs of PLC has been observed. This heterogeneity could be explained by geographical variability in those variants.

In addition, research on HBV and HCV variants could evaluate their potential impact in modifying responses to the HBV vaccines currently employed.

1.5.2. Aflatoxins

PLC has been related to aflatoxin exposures in human diets in countries where fungal infestation of crops and animal feed are common. Studies that used aflatoxin/albumin adducts, aflatoxin M1 in urine, aflatoxin-N7-Guanine adducts in urine or p53 specific mutations (G to T transversions at codon 249), tend to indicate that individuals who are carriers of persistent HBV infection and who are exposed to aflatoxins in their diets have an increased risk of progression to PLC as compared to non-aflatoxin exposed HBV carriers. A similar interaction with chronic HCV has not been documented[87]. The evidence is, however, limited to a few studies and not fully consistent[74]. The impact of aflatoxin exposure in the absence of viral infections has been difficult to document.

1.5.3. Alcohol and tobacco

Chronic alcohol abuse and alcoholic cirrhosis have long been recognized as a cause of PLC. However, it is not certain whether alcohol is a true carcinogen or if it acts as a co-factor in the presence of coexistent infection with HBV and/or HCV. Several epidemiological studies among alcoholics have described a high prevalence of HBV markers (16-70%) and of HCV markers (10-20%) as compared to a background prevalence of close to 5% and less than 1% respectively. These prevalences are even higher in PLC patients who are also alcoholics (27% to 81% of HBV markers and 50-77% of HCV markers) suggesting a complex interaction between alcohol and viral infections in the etiology of PLC[88]. Studies conducted in Northern Italy and Greece estimated that the attributable fraction (%) for high levels of alcohol consumption, once adjusted for HBV and HCV status, were 45% in Italy[89] and 15% in Greece[90].

The association between cigarette smoking and PLC has been suggested from the results of some epidemiological studies, notably in Japan. The evidence is, however, not consistent and residual effects of non-detected HBV or HCV cannot be safely ruled out in the majority of the epidemiologic studies. Recent data from China and Taiwan have also reported an association of cigarette smoking with PLC independent of HBV status[91,92].

The sex specific impact of alcohol, tobacco and oral contraceptive use could explain the sex ratio pattern in PLC as well as the steady increase in incidence by age observed in the populations where these exposures are common.

1.5.4. Hormonal factors

The occurrence of benign liver adenomas and occasional PLC among women who were long term oral contraceptive users has been documented[93-100]. Several case-control studies conducted in developed countries where substantial numbers of women have used oral contraceptives for extended periods of time, have found RRs between 1.6 and 5.5 among ever oral contraceptives users and a relationship with duration of use was also observed in some studies[93]. In one of the few studies conducted in the black population in South Africa, a country with a high prevalence of HBV, no association between oral contraceptives and PLC in women was found[94]. Analysis of the mortality trends for PLC in young women in the United Kingdom, U.S., Japan and Sweden, provide no support for a measurable effect of oral contraceptives on mortality due to PLC[95-97]. A multicenter study conducted with 317 cases of PLC among women under 65 years and 1060 controls, found an association between this tumor and duration of oral contraceptive use in the small subgroup of PLC cases without liver cirrhosis and with negative serology for HBV and HCV[98]. Further studies are needed to clarify and quantify the role of oral contraceptives in PLC, an issue of considerable public health interest.

Other hormonal factors have been explored to explain the increased male susceptibility worldwide. A recent study in the U.S. found a protection linked to parity, late menopause, use of hormonal replacement therapy and early age at menarche. Except for the latter, all factors were independent of HBV status[99]. Likewise, androgens have been postulated as a risk factor for PLC[100].

1.5.5. Other and emerging risk factors

Other factors that may modulate the long term impact of HBV or HCV persistent infections include dietary factors[101-105], some chemicals (such as arsenic[106]) and some hereditary conditions (haemochromatosis[107,108] and Wilson's disease[109,110]). Growing interest is currently focusing on the associations of PLC with Diabetes Mellitus[111-114] and obesity[115]. Non-alcoholic fatty liver disease is being proposed as a risk factor for PLC[116]. Because of the considerable prevalence of some of these conditions in western populations, it is of importance to conduct research to properly describe the nature of the associations observed.

1.6. Prevention of liver cancer

By the year 2003, WHO estimates of the worldwide percentage of the target population vaccinated against Hepatitis B was 42%[10]. A major input from donor agencies has made vaccination campaigns possible in the last decade and such efforts should be encouraged and supported. HBV vaccination trials initiated in the 80s have already shown the ability of HBV vaccines to prevent the chronic carriage of HBsAg[117,118] and to reduce the development of liver cancer either when vaccination takes place at birth in Taiwan[119], or in adults among HBsAg negative men in Korea[120].

Screening of blood products for HCV markers in the countries that have introduced such programs, has substantially reduced the rate of post transfusion HCV infections[121]. Awareness of the negative impact of unprotected sex, substantial alcohol consumption and smoking has probably had some impact in risk behavior modification. However, their potential impact on the incidence of PLC remains to be documented. Finally the claim has been made that aflatoxin reduction has reduced the incidence of PLC in Singapore and Shanghai although this has not been adequately substantiated.

1.7. Summary

PLC remains a major health problem with great geographical variability. Men are consistently more affected than women and survival is poor worldwide. Increasing trends in incidence in some developed countries, including the U.S., suggest an underlying cohort effect linked to HCV and HBV exposure. Efforts to reduce the PLC burden in most developing countries should give priority to HBV vaccination campaigns and to the prevention of HBV and HCV contamination. This implies reinforcing control of blood and derivatives, as well as the use of sterile medical equipment. HBV chronic carries may further benefit from reductions in the aflatoxin exposure in their diets. If achieved, aflatoxin reduction may also offer some protection to HCV carriers. In low risk populations, alcohol consumption may account for the majority of the PLC cases that do not show viral markers.

1.8. Primary liver cancer in Spain

In Spain, it has been estimated that 4,366 new cases of PLC occurred during 2002, which corresponded to 3,027 in men and 1,339 in women[20]. Based on estimates of PLC in 38 European countries, these cases were 8.6% and 7.4% of the new European PLC cases estimated in that year for men and women, respectively[20]. In Europe, Spain ranks seventh in terms of PLC incidence (AAIR: 9.2 per 100,000) in men and eleventh (AAIR: 2.9 per 100,000) in women[20].

The number of deaths from PLC in Spain during 2002 were 4,481, which corresponded to 2,898 in men and 1,583 in women[27]. In 2002, Spain ranks seventh among European countries in terms of number of PLC deceased subjects (AAMR for men: 8.4 per 100,000; AAMR for women: 3.3 per 100,000)[27].

The mortality to incidence ratio is greater than one for both men and women, which is related with the problem of liver metastases misclassified as PLC in mortality data or perhaps due to infraestimation of liver cancer incidence. It also reflects the low one-year survival after diagnosis for this tumor[27].

Trends in liver cancer incidence in Spain have only been evaluated in two registries: Zaragoza and Tarragona. In Zaragoza, a decrease of PLC incidence[26] during 1978-92 was detected. However, in Tarragona no change in PLC incidence was observed during the period 1980-97[50].

Trends in PLC mortality in Spain were first evaluated during the period 1975-87 analyzing data from the Spanish population older than 35 years. In this study a sharp decrease in PLC mortality rates was detected in both sexes[122]. Time trends in PLC mortality were also analyzed during the period 1970-96 in 20 European countries, detecting downward trends in Spain compared to those of the other European

countries included in that study[123]. These downward trends were explained by better diagnosis and assignment of metastatic tumors to the site of their primary origin[122,123] and also to modification of the ICD for the eigth to the ninth revision during the study period. One study carried out in Catalonia reported that PLC mortality remained stable during the period 1980-97[50]. In the same study, an increase of mortality for cholangiocarcinoma was observed for all age groups and for both sexes[50].

Chapter 2

2. Hypotheses and objectives

2.1. Meta-analysis of cohort studies of risk of liver cancer death among HBV carriers

Hypotheses about heterogeneity of RRs of PLC:

1)   The risk of developing or dying from PLC in South-Eastern Asia is higher than that of developed countries: The high variability in RRs of PLC observed in cohort studies among HBV carriers could be partially explained by geographic variability, probably as a consequence of differences in risk factors among countries.

2)   Study design could influence the estimation of RRs: In each cohort study, the RRs have been calculated comparing the PLC mortality rates of the cohort with those of a reference population or group. Heterogeneity among those comparison groups could partially explain the variability of RRs.

Objectives:

1)   To propose and assess a statistical procedure for the combined analysis of the RR of death due to PLC among male cohorts of HBV carriers conducted worldwide before 2006.

2)   To determine and explain the influence of geographic variability and study design in the RR of death due to PLC among HBV male carriers.

2.2. Geographic distribution of primary liver cancer in Europe in 2002

Hypothesis about geographic variability:

The risk of incidence and mortality from PLC in Southern European countries is the highest in Europe: HBV and HCV prevalences in European countries could partially explain these differences among European countries.

Objectives:

1)   To propose and assess a statistical procedure for model selection for the mapping of PLC risk, taking into account the effect of geographic area and HBV and HCV prevalences.

2)   To estimate the effect of HBV and HCV seroprevalences on PLC mortality and incidence by geographical area.

2.3. Time trends in liver disease in Spain during the period 1983-97

Hypothesis about incidence and mortality from liver disease in Spain:

An increasing incidence and mortality by PLC in France and Italy has been observed during the 90s attributed to a cohort effect related with a previously HCV exposure 30-50 years ago, whereas cirrhosis mortality decreased during that period. In Spain, where the pattern of HBV and HCV prevalences is similar to that of Italy and France, a similar increase in PLC incidence during that period would be expected. On the other hand, a decrease of cirrhosis mortality is expected, as it has been observed in most developed countries.

Objectives:

1)   To propose and assess a statistical procedure for an age-period-cohort analysis of time trends of liver disease in Spain

2)   To determine time trends in incidence and mortality for liver tumours (PLC and cholangiocarcinoma) and liver cirrhosis mortality in Spain during the period 1983-97.

Chapter 3

3. Meta-analysis of cohort studies of risk of liver cancer death among HBV carriers

3.1. Background

The role of chronic HBV infection in the aetiology of PLC has been long established[66]. The link between HBV and PLC appear to have been found for the first time during the 60s, after the discovery of the HBsAg[124]. Since then, several cohort studies have been carried out among HBsAg-positive subjects. The monograph about viral hepatitis and PLC edited by the International Agency for Research on Cancer (IARC) collected the results of 15 cohort studies of HBV carriers published between 1970 and 1992[3]. All these studies were based on comparisons all-cause and liver-related mortality among healthy HBsAg-positive subjects compared to HBsAg-negative subjects or the general population. The mortality comparisons between HBsAg-positive and HBsAg-negative subjects is the basis for the estimation of the risk of death from PLC among HBV infected subjects.

Most of these studies were conducted in Asian countries and some few in Western countries, showing RRs of death from PLC ranging from 5.3 to 148. The RRs reported from Asian studies were higher than those of Western countries. Even among relatively similar western countries risks were found, some differences have been reported. For example, two cohort studies of HBV carriers have been conducted in the Mediterranean area: one in Catalonia (Spain) and another in Italy[24]. Although these two studies were carried out in similar populations (Southern European), differences of results were found. In the Spanish study, the cohort constituted 2,206 HBsAg-positive subjects (1,575 men, 631 women), and 15,504 HBsAg-negative subjects (8,783 men, 6,721 women) voluntary blood donors selected from four hospitals in the Barcelona area (Hospital Universitari de Bellvitge, Hospital de la Vall d'Hebron, Hospital Clínic i Provincial and Hospital de la Creu Roja). Mortality in both groups was detected by record linkage with the Catalan Mortality Registry. The risk of death from PLC was 14-fold higher in HBsAg-positive men than in those who were HBsAg-negative. For women, this risk was 7-fold higher although it was not found to be

statistically significant[21]. However, the study conducted in Italy, with a mean follow-up of 30 years, did not detect an excess risk of death due to PLC among HBV carriers[24].

The heterogeneity found in the estimated risk of death from PLC associated to HBV in these cohort studies is difficult to explain, but could be explained partly due to:

a) Selection of comparison group: If the HBsAg-positive subjects are workers or blood donors (WBD) and the comparison group constitutes the general population (GP), the study could underestimate risk of death as a result of selection bias. This selection bias is known as the "healthy donor effect" or "healthy worker effect"[125].

b) Size of the cohort: The small number of subjects included in some longitudinal studies could diminish the statistical power to detect an excess of PLC risk.

c) Cofactors not taken into acount at the time of the design of study: Some of these results could be potentially affected by cofactors such as alcohol, tobacco consumption, HCV infection and aflatoxins.

d) Since information about these cofactors are not usually available, at least at the individual level, one approach to assess their impact is to assume that they vary among geographical areas and, thus, geographical area can be a good subrogate variable.

For those reasons, it is necessary to investigate the factors that contribute to the heterogeneity of risks of death by PLC observed in different studies.

3.2. Hypotheses

1) There could be an effect of geographic area in the estimation of RR of HBV-associated death from PLC: some of these studies were carried out in Taiwan and China, two high risk areas for PLC, whereas the remaining studies were carried out in developed countries with intermediate-low risk for PLC.

2) There could be an underestimation of RR of PLC mortality linked to HBV in those studies for which the comparison group comprises the general population while the cases have been ascertained from blood donors.

3.3. Sources of data

Studies considered in the meta-analysis were cohort studies which reported the risk of death due to PLC among male HBV carriers. The initial pool of eligible studies was extracted from the IARC Monograph on viral hepatitis and liver cancer[3] and included studies published through the period 1970-92. A computerized search of the Medline, Cancerlit and Pubmed databases was also performed jointly with a review of literature listed in relevant papers. The search strategy included related terms such as liver tumors, HBV infection, HBsAg, liver cancer, primary liver cancer, hepatocellular carcinoma, and cohort, prospective or longitudinal studies. A total of 23 studies with potential interest were extracted with this search strategy.

Inclusion criteria

a) Mortality studies: studies should report RR of death due to PLC.
b) Cohorts constituted by male HBsAg-positive subjects who do not present liver disease at the beginning of the study.
c) Cohorts which had an exhaustive follow-up performed by record linkage with the mortality registry of their country, medical examination or/and revision of medical records.

Studies excluded

A total of 12 studies were excluded:

Four studies were excluded because they reported incidence jointly with deceased cases. Of these, one was conducted in Taiwan[92], two in Japan[126;127], and one in Alaska[128], with relative risks of developing PLC from 7 to 148.

One study from Japan was also excluded because it was carried out in a cohort which included only women[64]. An study carried out in Hawaii[129] was excluded because the

risk of death from PLC was not reported. For the same reason, a study carried out in Italy[24] was not included. Finally, we excluded two prospective nested case-control studies conducted in different areas of China[70,71] because they included some subjects with chronic liver disease as cases.

Updated Studies

We found 14 studies eligible for the analysis. Of these we did not include in the analysis one study conducted in Japan by Sakuma et al. in 1982[62] because its cohort members were included in a later study published in 1988[61]. Two other cohort studies presented updated data on their original cohorts: 1) the study of Beasley et al in 1981[66] was updated in 1991 by Beasley and Hwang[130] and 2) the study of Hall et al. in 1985[22] was updated in 2003 by Crook et al[23].

Studies selected in the meta-analysis

Therefore 11 studies met the inclusion criteria and were based on unique patients. These studies are described in *Table 3.1* which shows the geographic area where the study was carried out, the cohort size, years of follow-up (mean), type of comparison group, the number of PLC deaths observed, the RR of PLC and if the study accounted for the presence of other cofactors. HBsAg-positive subjects were blood donors in six studies[21,23,58-60,65] and workers in the remaining studies. Eligible studies were carried out in several geographical areas: two in the U.S.[58,59], one in England and Wales (United Kingdom, U.K.)[22,23], four in Japan[60-62,65,126], three in Taiwan and China[66,68,74] and one in Spain[21].

Two studies carried out in the Southeast Asia[68,74] used comparison groups comprising general population tested for HBsAg. Cofactors as alcohol consumption or tobacco smoking were also reported in some of these studies[60,68,74]. The standardized mortality

ratio was reported in seven of these studies[21,23,58-60,65,126] as an estimator of the relative risk, while the others reported a relative risk derived from a Cox model.

Table 3.1. Cohort studies of HBV carriers during the period 1981-2006.

Author	Geographic Area	Cohort (N)	Follow-up (years)	Control Group	O	RR	Cofactors
Prince et al, 1982[58]	U.S.	5,353	4.3	GP	3	9.70	Not Reported
Oshima et al, 1984[60]	Japan	8,646	6.2	GP	20	7.80	Aflatoxins & Tobacco
Ijima et al,1984[126]	Japan	495	5.5	GP	8	10.40	Not Reported
Dodd et al, 1987[59]	U.S.	10,654	3.6	WBD	6	26.80	Not Reported
[a]Sakuma et al,1988[61]	Japan	513	7.3	WBD	9	21.00	Not Reported
Tokudome et al, 1988[65]	Japan	2,595	5.9	GP	15	7.30	Not Reported
[b]Beasley et al, 1991[130]	Taiwan	3,454	11.0	WBD	184	103.00	Not Reported
Yang et al,2002[68]	Taiwan	2,361	13.0	GP*	82	17.50	Aflatoxins & Tobacco
Evans et al,2002[74]	China	8,795	8.0	GP*	643	18.80	Aflatoxins & Tobacco
[c]Crook et al, 2003[23]	U.K.	2,681	22.4	GP	20	26.31	Not Reported
Ribes et al, 2006[21]	Spain	1,575	20.5	WBD	14	14.14	Not Reported

a: update from the initial study of Sakuma et al in 1982[62] which reported a RR of 28.3 with a mean follow-up of 5 years; b: update from the initial study of Beasley et al. in 1981[66] which reported a RR of 223 with a mean follow-up of 3.5 years; c: update from the initial study of Hall et al in 1985[22] which reported a RR of 42 with a mean follow-up of 7.6 years; (N): number of men in the cohort; O: Observed deaths in the cohort by PLC; RR: Relative Risk; GP: General Population; GP*: General Population tested for HBsAg; WBD: Workers or Blood Donors; Cofactors: cofactors reported in the study.

3.4. Statistical methods

Meta-analysis can be defined as a quantitative analysis of the results of a variety of studies with the aim of an integrated representation of them[131]. It could be viewed as a systematic analysis which allows the researchers to evaluate the consistence of the results, leading them to propose explanations in case of heterogeneity. The procedure consists in: i) reviewing the available bibliography on a potentially causal relation and ii) applying statistical methods to integrate results, estimating a combined measure of effect and its statistical significance. The aims of these procedures are to increase the precision of the effect measure under study and to explore existence of and reasons for any heterogeneity of results observed in the studies[132].

These results should refer to one effect to one type of effct measure. Among others, these effect measures could be the odds ratio, the incidence rate of a determined disease, the relative risk or the risk difference between groups. In order to model each one of these effect measures we could presume homogeneity among them. If so, we use a fixed effects model, for which it is supposed that the variability observed is due to entirely intra-study variability, or the variability within each study[133,132]. If we suppose that the effect measure is non-homogeneous among studies, we will use a random effects model, which allows for the extra-variability due to a combination of the intra and inter-study variability[132,133].

The effect measure of interest in our analysis is the Standardized Mortality Ratio (SMR), which is an estimator of the RR of death[134]. The SMR is defined as the ratio between observed deaths in the cohort from the cause of interest (O) and the expected deaths (E) the cohort from that cause, where the latter is based on the mortality of a reference population[134]. In those studies where a relative risk (RR) has been reported rather than an SMR, the number of expected deaths has been estimated as the ratio of O and RR[134].

For an observed SMR it has been assumed that O follows a Poisson distribution with rate $\lambda = RR \cdot E$, where the SMR is the maximum likelihood estimator for the RR. The variance of the crude SMR may be approximated through the logarithm of the SMR[134],

$$\text{Var}(\log(\text{SMR})) = \frac{1}{O} \qquad (1),$$

and 95% confidence intervals for the log(SMR) can be obtained[134] as

$$\log(\text{SMR}) \pm 1.96\sqrt{\frac{1}{O}} \qquad (2).$$

A combined RR of the studies can be estimated by means of a weighted average of SMRs, giving each study a weight proportional to its precision[134]. The precision is given as the inverse of the variance of the measure of interest[131]. However, the modelling of events by means of a Poisson distribution could entail overdispersion, also known as extra-Poisson variability, which is present if $Var(O) > E(O)$ [135]. Overdispersion could indicate that there is clustering of data, which often can be corrected through modelling with appropiate explanatory variables or using mixtures of probability distributions[136].

The meta-analysis in this chapter focuses on these methods, comparing the Frequentist approach for mixtures of probability distributions with the Bayesian one. Tthe extra-Poisson variability has been modelled through explanatory covariates once identified the sources of heterogeneity.

3.4.1. Frequentist approach to meta-analysis based on mixtures of probability distributions

We have performed this part of the analysis with a non-parametric mixture approach. It has been assumed that the number of observed deaths in one study, O, can be modelled as a linear combination of k different Poisson distributions. We will refer to each one of these distributions as components of the mixing distribution. Each one of these components has a weight p_j on the mixture and mean μ_j. The subsets of parameters in the mixture of distributions are $P = [p_1,...,p_k]$ with $\sum_{j=1}^{k} p_j = 1$, $M = [\mu_1,...,\mu_k]$, and E as the expected number of cases of the study. Then, the mixing distribution G for O is

$$G(O \mid M, P, E) = \sum_{j=1}^{k} p_j \cdot f(O \mid \mu_j, E) \qquad (3),$$

where f is a Poisson distribution with parameter $\lambda = \mu_j \cdot E$ and μ_j estimates the RR in each component of the mixture.

The expected value for a random variable with mixing distribution (3) can be estimated assuming that P, M, E are constant values[136],

$$E(O \mid M, P, E) = \sum_{j=1}^{k} \sum_{i=1}^{\infty} O_i \cdot f(O_i \mid \mu_j, E) \cdot p_j = E \cdot \sum_{j=1}^{k} p_j \cdot \mu_j \qquad (4).$$

From (4), the expected value for SMR is

$$E(SMR) = \sum_{j=1}^{k} p_j \cdot \mu_j = RR_{pooled} \qquad (5),$$

being a pooled RR. With the same assumptions one can estimate the variance of the

SMR, $Var(SMR) = \dfrac{RR_{pooled}}{E} + \sum\limits_{j=1}^{K} p_j \cdot (\mu_j)^2 - (RR_{pooled})^2$. To find maximum

likelihood estimates of P, M and k, Böhning proposed a non-parametric maximum likelihood estimation (NPMLE) implemented in the software C.A.MAN (Computer Assisted Mixtures Analysis)[136]. The components of the mixture could be also interpreted as subpopulations or clusters, and $k > 1$ means heterogeneity[136]. If we denote $\hat{P}, \hat{M}$ and $\hat{k}$ as the NPMLE of the mixing distribution, then the RR_{Pooled} can be determined with these constant parameter estimates.

It is necessary to establish to which of the components each study belongs in order to investigate the studies which are sources of heterogeneity. Let us consider a latent random vector Z_i of length k, $Z_i = [Z_{i1}, ..., Z_{ik}]$ for each study, consisting of 0s except for one 1 at a certain position, say the jth, which indicates that the study belongs to the jth component. Applying Bayes theorem and using the estimated mixing distribution of (3) as a prior distribution, the probability that each study belongs to certain component is

$$P(Z_{ij} = 1 \mid O_i, M, P, E_i) = \dfrac{f(O_i \mid \hat{\mu}_j, E_i) \cdot \hat{p}_j}{\sum\limits_{s=1}^{k} f(O_i \mid \hat{\mu}_s, E_i) \cdot \hat{p}_s} \qquad (6).$$

The ith study is then assigned to the population to which it has the highest posterior probability of belonging.

In order to determine confidence intervals for the RR_{Pooled} a question that arises is which distribution do we assume for the RR_{Pooled}. Bootstrap methods over the NPMLE obtained with C.A.MAN have been suggested[136] in order to obtain confidence intervals for (5). In addition, if we want to consider the parameters P, M and k as

random variables, we can estimate their expected values and variances. Moreover, the expected value for O becomes

$$E(O \mid E) = \int \left[\sum_{i=1}^{\infty} O_i \cdot G(O_i \mid M, P, E) \right] \pi(P, M) dP dM \quad (7),$$

Where $\pi(\cdot)$ is a probability density function and $\pi(P, M) = \pi(P \mid M) \cdot \pi(M)$. Specific numerical methods are required to solve these integrals. We have proposed an alternative methodology based on the Bayesian approach which allows for the estimation of the parameters and their variances a posteriori.

3.4.2. Bayesian approach to meta-analysis based on mixtures of probability distributions

3.4.2.1. Bayesian inference

Let us suppose that our parameter of interest is θ, the RR_{Pooled}. The prior knowledge about this parameter, in terms of probability statements, is specified by means of a probability density function $\pi(\theta)$. The information extracted from the data is summarized by the likelihood function which we denote by $\pi(Data \mid \theta)$. In order to make probability statements about θ once observed data, a model for the joint probability distribution of parameter and data is necessary, being the joint probability distribution $\pi(Data, \theta) = \pi(Data \mid \theta)\pi(\theta)$. Simply conditioning on the known data, using the basic property of conditional probability known as Bayes' rule, yields the posterior probability $\pi(\theta \mid Data) = \dfrac{\pi(Data, \theta)}{\pi(Data)} = \dfrac{\pi(Data \mid \theta)\pi(\theta)}{\pi(Data)}$, where $\pi(Data) = \int \pi(Data \mid \theta)\pi(\theta)d\theta$. The factor $\pi(Data)$ does not depend on θ and, with fixed data, can thus be considered a constant value, yielding an unnormalized posterior density, which is

$$\pi(\theta \,|\, Data) \propto \pi(Data \,|\, \theta)\pi(\theta) \qquad\qquad (8).$$

Hence, the posterior median, mean, variance and other statistics can be obtained when (8) has been estimated. From each statistic of interest we can estimate its credibility interval (CRI) which is analogous to the confidence intervals of the Frequentist estimation methods. The CRI is the interval where θ is located with a certain probability. As an example, the posterior 95% CRI for θ should be calculated solving $\int_{a}^{b} \pi(\theta \,|\, Data)d\theta = 0.95$, a and b being the lowest and upper values of the CRI[137].

If we obtain new data, the new *prior* distribution for the θ parameter would be the last *posterior,* so $\pi(\theta \,|\, Data) \rightarrow \pi_{new}(\theta)$. This is the way that information is updated in the Bayesian framework [137,138].

In the case of several parameters for the model, as in the regression framework, the Bayes theorem can be applied in the same way as in (8), giving a joint distribution for all parameters. Let us suppose a set of *m* parameters of interest, $\Theta = \{\theta_1,..,\theta_m\}$, then (8) becomes

$$\pi(\Theta \,|\, Data) \propto \pi(Data \,|\, \Theta)\pi(\Theta) \qquad\qquad (9).$$

To obtain the *posterior* marginal distribution for any of the *m* parameters we should integrate (9) with respect to the remaining *m-1* parameters keeping in mind that each one has a prior distribution. To solve these integrals, specific numerical methods are required. Among the different alternatives are Markov Chain Monte Carlo Methods (MCMC)[139]. These are simulation methods based on drawing values of Θ from approximate distributions and correcting these samples to better approximate the

target posterior distribution (9). The samples are drawn sequentially, with the distribution of the sampled draws depending on the last value drawn; hence, the draws from a Markov Chain[137]. In a Markov Chain simulation, several independent sequences of simulation draws are created; each sequence Θ^t with $t = \{1,2,...\}$ is produced by starting at some point Θ^0 and then, for each t, drawing feasible values for Θ^t from a transition distribution, $T_t(\Theta^t \mid \Theta^{t-1})$ that only depends on the previous draw, Θ^{t-1}. The transition probability distributions must be constructed so that the Markov Chain converges to a unique stationary distribution, that is, the posterior distribution (9). The Gibbs sampler, Metropolis and Metropolis-Hastings are algorithms developed to perform Markov Chain simulation [138-139] in order to generate values from a random variable with a certain probability distribution.

In the last decade several statistical packages have been developed to facilitate these simulation methods. In this study we performed the analyses using the Bayesian inference package WinBUGS 1.4 [140] which allowed us to perform Bayesian inference using Gibbs Sampling. This package can be run directly from within R[141] using functions of the library R2WinBUGS[137]. Initially, for each parameter of interest noninformative prior distributions have been used, which have large prior variance, and, therefore, low prior precision.

By means of these MCMC methods a simulated *posterior* distribution is obtained. The point estimate used most frequently is the mode or the median[137]. In this study the empirical CRI have been set to 95% and the point estimate used is the median value.

The Bayesian inference in this study has been performed by means of the following steps:

a) Determining the number of components and estimating the theoretical RRs for each one;

b) Determining the sources of heterogeneity by means of the posterior classification of the studies; and

c) Covariate modeling of the RRs observed in each study taking into account the sources of heterogeneity.

Steps a) and b) are described in Section 3.4.3.1, whereas section 3.4.3.2 describes step c).

3.4.2.2. Mixture Model based on Dirichlet Process

The Dirichlet Prior Process (DPP) is an approach used to make inferences about the underlying location for a certain observation (study) in a mixture of probability distributions such as (3)[142-145]. Let us consider that O_i are the observed number of deaths by PLC of the ith study which is a random variable that follows a mixture distribution such as (3). Also let all the parameters μ_j have the same prior distribution H_0. This mixture model can be expressed hierarchically as[144,145]

$$
\begin{aligned}
&O_i \mid Z_i, M, E_i \sim Poisson(\mu_{Z_i} \cdot E_i) \\
&Z_i \mid P \sim Multinomial(P) \\
&\mu_j \sim H_0 \\
&P \sim Dirichlet(\delta, A_k)
\end{aligned}
$$

(10),

where Z_i are the latent variables or labels that assign $\dfrac{O_i}{E_i}$ to a parameter value μ_{Z_i},

$P = [p_1, ..., p_k]$ with $\sum_{j=1}^{k} p_j = 1$ and $M = [\mu_1, ..., \mu_k]$. We assumed that the latent variables follow a discrete (Multinomial) distribution with $j = \{1, ..., K\}$ possible values, each one of this values with probability p_j. The parameters of each component μ_j can be drawn from H_o beforehand, and then a distribution would be on the probability of selection of these parameters, the set M. The weights P follow

a prior distribution which is Dirichlet with parameters (δ, A_k), $A_k = [\alpha_1, ... \alpha_k]$ being a prior base measure, such as $\sum_{j=1}^{k} \alpha_j = 1$, and δ being the spread of the base measure[145]. The key problem is how to assume a prior distribution for (δ, A_k). If we consider P in the limit $k \to \infty$, then the Dirichlet Distribution becomes a Dirichlet Process $DP(M, \delta, A_k)$, which is an extension of the Dirichlet distribution to continous spaces[145]. For each indicator Z_i, drawn conditioned on all previous (i-1) indicators from the Multinomial distribution, there is a corresponding μ_j that is drawn from H_0. In the limit, $k \to \infty$, the labels lose their meaning as the space of possible values becomes continuous. We can discard the use of labels in the model and let the parameters be drawn from a Dirichlet Process with base measure A_∞. Hence the DP model is

$$
\begin{aligned}
O_i \mid M, E_i &\sim Poisson(\mu_i \cdot E_i) \\
\mu_i \mid H &\sim H(M) \qquad\qquad (11), \\
H(M) &\sim DP(M, \delta, A_\infty)
\end{aligned}
$$

where $H(M)$ is a mixing distribution drawn from the DP. Sethuraman[145] showed that the mixing distribution $H(M)$ can be constructed by means of P in the following procedure also known as the "constructive definition" of the Dirichlet process[142,144,145] or "stick-breaking"[144] method. Let $r_1, r_2, ..., r_k$ be a sequence of $Beta(1, \delta)$ random variables, and let us define $\phi_1 = r_1, \phi_2 = r_2(1 - r_1), ..., \phi_k = r_k(1 - r_k)$. Informally, this construction can be thought of as a stick-breaking procedure, where at each stage we independently, and randomly, break what is left of a stick of unit length and assign the length of this break to the current ϕ_j value[144]. Finally, we should define $p_j = \dfrac{\phi_j}{\sum_{t=1}^{k} \phi_t}$ to ensure that

$\sum_{j=1}^{k} p_j = 1$ so that the "construction" will end. Note that due to the "constructive definition" of the Dirichlet process from $Beta(1,\delta)$, models (10)-(11) depend only on δ. With these criteria, the number of components that emerges from a particular data set depends on δ and are random[142-144]. To assess the impact on the number of components we performed this analysis with three different schemes for δ. In the first scheme we assumed that $\delta \sim Gamma(1,1)$, whereas we set $\delta = 1$ and $\delta = 5$ in the remaining two schemes, respectively. It should be noted that the emerging number of components from a particular data set depends on the prior value assumed for δ. In order to perform a 'constructive definition' of the Dirichlet process, the maximum number of components was initially set to 11, the number of studies, and at each iteration of the algorithm the number of nonempty components could be calculated. We repeated this simulation again reducing the maximum number of components until we found their optimal number, because it should take into account that the "stick-breaking" method computes the RR_{Pooled} on the basis of the maximum number of components. We performed each analysis with two different prior distributions for μ_j, being $\mu_j \sim Gamma(0.01, 0.001)$ and $\log(\mu_j) \sim N(0, \tau)$, where τ is the precision (inverse of the variance) for the normal distribution used, for which we set $\tau = 0.001$. Posterior classification of the studies has been performed by means of (6) once components of the mixture have been estimated through the described procedure. Identifiability of μ_j has been imposed by means of an ordering constraint[146] $\mu_1 < \mu_2 < .. < \mu_k$, where each μ_j is sampled from a left censored distribution on the μ_{j-1} value such that $\mu_j \sim Gamma(0.01, 0.001) | (\mu_{j-1}, \infty)$ and $\log(\mu_j) \sim N(0, \tau) | (\log(\mu_{j-1}), \infty)$.

3.4.2.3. Modelling heterogeneity using covariates

Once identified how the studies were grouped, a generalized linear model (GLM) was used to explain the effect on the RR of the possible sources of heterogeneity. With the assumption of O_i following a Poisson distribution, the GLM is defined as

$$\log(O_i) = \log(E_i) + \alpha_1 Y_{1i} + \alpha_2 Y_{2i} + \beta_1 X_{1i} + \beta_2 X_{2i} \qquad (12),$$

with X and Y as dichotomous variables. For the ith study the variable Y refers to geographical areas with low risk of death by PLC, while X refers to those areas with high risk of death by PLC. The variable Y_{1i} and X_{1i} indicates whether the comparison group mortality rates were extracted from the GP, whereas Y_{2i} and X_{2i} indicate if the comparison group was selected from WBD. For each study, one of the four variables $(Y_{1i}, X_{1i}, Y_{2i}, X_{2i})$ has value 1, whereas the remaining three have value 0. In that sense, the parameters $\alpha_1, \alpha_2, \beta_1, \beta_2$ are mean log RRs.

For the low risk areas of death by PLC we have estimated the ratio of RRs between studies which used WBD comparison groups versus those which used GP as comparison group. This ratio is defined as

$$Q_1 = \frac{e^{\alpha_2}}{e^{\alpha_1}} \qquad (13).$$

In the same line, but for the high risk areas of death by PLC, the rate ratio

$$Q_2 = \frac{e^{\beta_2}}{e^{\beta_1}} \qquad (14),$$

applies as (13).

For the studies which used GP as comparison group we have also estimated the ratio of RRs between studies conducted in high risk geographical areas versus those carried out in geographical areas with low risk of death by PLC. This rate ratio is defined as

$$Q_3 = \frac{e^{\beta_1}}{e^{\alpha_1}} \qquad (15),$$

whereas for studies which used WBD as comparison groups we have estimated the ratio of RRs with

$$Q_4 = \frac{e^{\beta_2}}{e^{\alpha_2}} \qquad (16).$$

Table 3.2 depicts the interpretation of the model parameters estimated with (12).

Table 3.2. Effects of comparison group and geographical area in the estimation of risk of death by PLC considering the parameters of the covariates modeling.

| | Comparison Group | | |
| | Blood Donors/workers RR | Ratio² | General Population RR |
Area			
High Risk	e^{β_2}	Q₂	e^{β_1}
Ratio¹	Q₄		Q₃
Low Risk	e^{α_2}	Q₁	e^{α_1}

High risk: High risk area of death by PLC; **Low Risk:** Low risk area of death by PLC; **Ratio¹:** RR of death by PLC in a High risk area versus RR in a Low risk area. **Ratio²:** RR of death by PLC for studies which used Workers or Blood Donors as comparisonl group versus RR in studies which used General Population as comparison group.

Models (10) and (12) have been implemented in WinBUGS, and code for these methods can be found on Appendix A.1. For each model it has been run 3 dispersed chains with 60,000 iterations, discarding the first 10,000 burning samples. It has been used the sample traces plots and the Gelman and Rubin convergence diagnostics[147,148], in order to check for convergence of the chains. After convergence has been assessed, the empirical 95% credibility interval has been obtained for each parameter jointly with its median value.

3.5. Results

Table 3.3 shows the SMR for PLC jointly with the RR_{Pooled} obtained through the Frequentist approach with the studies included in the meta-analysis. The highest SMR was reported in the study carried out in Taiwan in 1991 (SMR=103.0; Confidence Interval (CI): 88.72-119.10) whereas the lowest SMR was found in a study conducted in Japan in 1988 (SMR=7.3, CI: 4.1-12.7).

Table 3.3. Studies included in the meta-analysis.

Author	Geographical Area	Comparison Group	O	E	SMR	95% CI
Prince et al., 1982	U.S.	GP	3	0.31	9.70	(1.95-28.36)
Ijima et al.,1984	Japan	GP	8	0.77	10.40	(4.48 - 20.49)
Oshima et al., 1984	Japan	GP	20	2.56	7.80	(4.76 - 12.05)
Dodd et al, 1987	U.S.	WBD	6	0.22	26.80	(9.95-56.36)
Sakuma et al.,1988	Japan	WBD	9	0.43	21.00	(9.57 - 39.83)
Tokudome et al.,1988	Japan	GP	15	2.06	7.30	(4.09 - 12.07)
Beasley et al., 1991	Taiwan	WBD	184	1.79	103.00	(88.72- 119.10)
Evans et al., 2002	China	GP	643	34.20	18.80	(15.70-22.50)
Yang et al.,2002	Taiwan	GP*	82	4.68	17.50	(13.93-24.75)
Crook et al., 2003	U.K.	GP*	20	0.76	26.31	(16.06-40.54)
Ribes et al., 2006	Spain	WBD	14	0.99	14.14	(8.84-31.74)
*RR_Pooled**					23.2	--

WBD: Workers or Blood Donors; **GP:** General Population; **GP*:** General Population tested for HBsAg; **O:** observed number of deaths by PLC; **E:** expected number of deaths by PLC; **95% CI:** 95% confidence interval; *RR_Pooled*: Relative Risk pooled estimated with the Frequentist approach.

Identification of subpopulations and parameters: Frequentist approach

This heterogeneity observed was explored through the mixture analysis. Three possible subpopulations with estimated mean RRs $\hat{M} = \{\hat{\mu}_1 = 8.0, \hat{\mu}_2 = 18.8, \hat{\mu}_3 = 103.0\}$ and weights $\hat{P} = \{\hat{p}_1 = 0.32, \hat{p}_2 = 0.59, \hat{p}_3 = 0.09\}$ were determined through C.A.MAN. With these estimate of the parameters it was obtained a RR_{Pooled} of 23.2 (Table 3.1). *Figure 3.1* shows the graphical representation of the log-scale SMRs, the possibly partitions of the log-scale SMRs into 3 components and the log-RR_{Pooled}.

Figure 3.1. Summary of the log(SMR) of cohort studies included in the pooled analysis.

95% Confidence Intervals for the log(SMR)

Study

Beasley,1981
Prince et al., 1982
Ijima et al.,1984
Oshima et al., 1984
Dodd et al., 1987
Sakuma et al.,1988
Tokudome et al,1988
Evans et al.,2002
Yang et al.,2002
Crook et al., 2003
Ribes et al.,2006
Pooled*

0 1 2 3 4 5 6

*Calculated by means of C.A.MAN although no confidence intervals were available.

SMR (log scale)

92
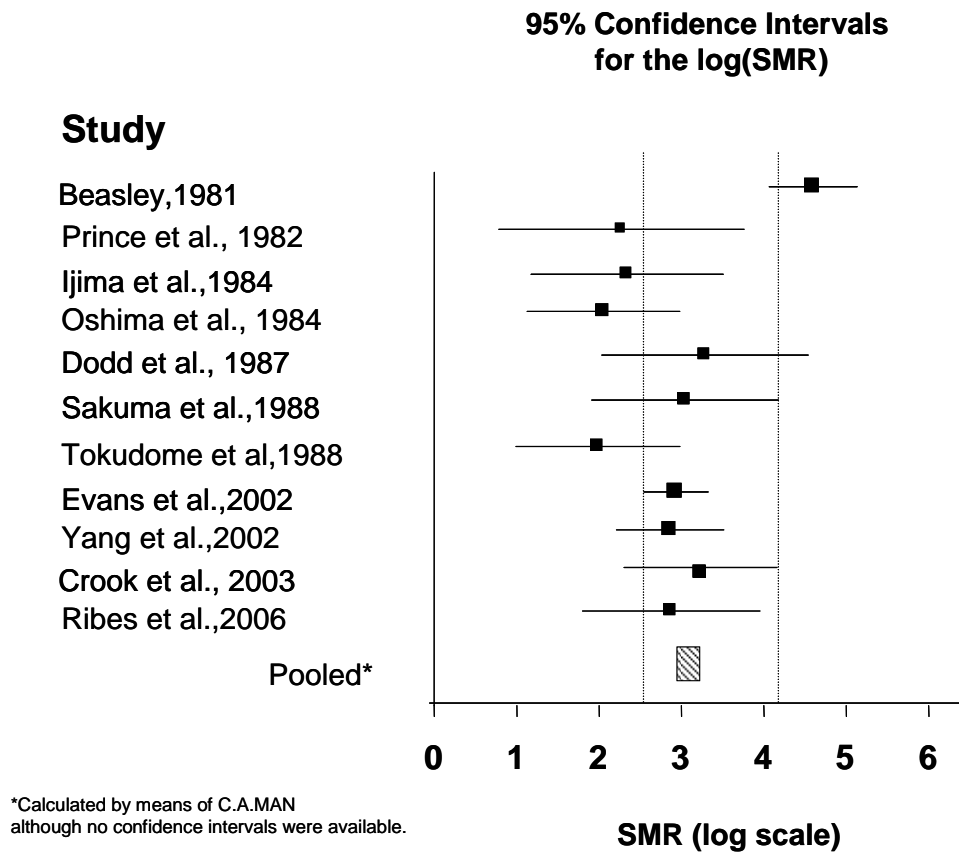
Identification of subpopulations and parameters: Bayesian approach

Table 3.4 shows results of the different simulation schemes using $\mu_j \sim Gamma(0.01, 0.001)$. This table reports the posterior distribution of δ, the number of nonempty components found, and the RR_{Pooled} obtained in each scheme. We first proceeded to assume that the maximum number of components of the mixture of Poisson distributions were 11 (Max=11). The parameter δ showed high variability when it was assumed to be random variable (median=5.7, 95% CRI: 0.4 - 77.1). There was found a median of 3 nonempty components with an upper value for the 95% CRI of 4 (95% CRI: 3 - 4) with an RR_{Pooled} greater than 30 in each scheme. However, the median RR_{Pooled} was calculated using the 11 components at each iteration. In the second step we assumed Max=4 and we found that the median number of nonempty components was also 3 (95% CRI: 3 - 3). Although the 95% CRI of the RR_{Pooled} and δ were narrower than the previous scheme, we performed the analysis with Max=3. In this last case, we found similar values for the RR_{Pooled} in each scheme, whereas the parameter δ (median=1.5) showed the narrowest 95% credibility intervals (95% CRI: 0.2 - 4.2) compared with the previous schemes.

Table 3.4. Results of the "stick-breaking" method using gamma prior distributions for μ_j with different maximum number of components: Median values and 95% credibility intervals.

Max	δ prior	δ posterior	Number of non-empty Components	RR_{Pooled}
11	Random	5.7 (0.4 – 77.1)	3 (3 - 4)	30.8 (25.7 - 78.2)
	Constant	1	3 (3 - 4)	30.3 (24.5 – 75.6)
	Constant	5	3 (3 - 4)	30.0 (24.7 - 79.2)
4	Random	5.1 (0.3 – 20.1)	3 (3 - 3)	29.1 (20.1 - 47.2)
	Constant	1	3 (3 - 3)	27.1 (21.5 - 48.1)
	Constant	5	3 (3 - 3)	28.8 (21.6 - 48.3)
3	Random	1.5 (0.2-4.2)	3 (3 - 3)	23.5 (14.9-44.5)
	Constant	1	3 (3 - 3)	22.4 (14.1-43.5)
	Constant	5	3 (3 - 3)	24.1 (16.3-47.1)

Max: maximum number of components of the mixture distribution; **δ prior:** prior assumption for the δ parameter on the "stick-breaking" method; **Random:** Gamma prior distribution for δ; **Constant:** δ parameter assumed to be constant; **δ posterior:** posterior distribution for δ (when it was considered as random variable) or constant value; **Number of nonempty components:** number of nonempty components of the mixing distribution; **RR_{Pooled}:** Estimator of the pooled Relative Risk.

Table 3.5 shows the comparison between results obtained with the Frequentist approach with those obtained with the Bayesian ones. We compared the results of the Bayesian approach with both Gamma and Normal prior distributions for μ_j and $\log(\mu_j)$, respectively, and with Max=3. Point estimate for the RR_{Pooled} obtained with the Frequentist approach (RR_{Pooled} =23.2) was almost identical to the median RR_{Pooled} values obtained with the Bayesian models (RR_{Pooled} with prior Normal for μ_j: 23.9 and RR_{Pooled} with prior Gamma for μ_j: 23.5). The same conclusion was extracted in the comparison of the Frequentist approach with the Bayesian one for the parameters μ_j and p_j obtained with different prior assumptions. Both RR_{Pooled} extracted from the Bayesian approach were almost identical in median value and 95% CRI. It was confirmed that the Bayesian method was not sensitive to the prior distribution used

for μ_j. In addition, both Bayesian models showed similar probabilities of classification a posteriori. The median value of the RR_{Pooled} calculated with the assumption of Gamma prior distribution for μ_j was closer to the point estimate of the RR_{Pooled} calculated with the NPMLE (RR_{Pooled} =23.2) than that estimated with the assumption of Normal prior distribution for μ_j. For that reason we chose the RR_{Pooled} calculated with the assumption of Gamma prior distribution for μ_j to continue with the analysis.

Table 3.5. Estimations of the parameters of the mixture of probability distributions: comparison of the Frequentist approach versus the Bayesian approach.

Method[*]	RR_{Pooled}		
Frequentist	23.2		
Prior Normal[1]	23.9 (15.1-45.3)		
Prior Gamma[2]	23.5 (14.9-44.5)		
	μ_1	μ_2	μ_3
Frequentist	8	18.8	103
Prior Normal[a]	7.8 (5.7-10.8)	18.7 (17.4-20.1)	102.6 (88.4-118.0)
Prior Gamma[b]	7.9 (5.7-10.8)	18.8 (17.3-20.8)	103.0 (88.2-119.3)
	p_1	p_2	p_3
Frequentist	0.32	0.59	0.09
Prior Normal[a]	0.29 (0.10-0.60)	0.61 (0.31-0.87)	0.10 (0.01-0.31)
Prior Gamma[b]	0.30 (0.11-0.61)	0.60 (0.30-0.89)	0.10 (0.01-0.31)

Method[*]: parameters estimated through Bayesian models are reported with their median value and 95 % credibility interval; **a:** parameters estimated with the Bayesian model with Normal prior distribution for $\log(\mu_{ij})$; **b:** parameters estimated with the Bayesian model with Gamma prior distribution for μ_j; ***RR_{Pooled}:*** Pooled estimator of the Relative Risk; μ_j: Estimation of the Relative Risk in the jth component of the mixing distribution; p_j: weight of the jth component of the mixing distribution;

In that sense, we have considered that the RR_{Pooled} due to PLC among HBV male carriers was 23.5 (95% CRI: 14.9 - 44.5) estimated from a mixture of three Poisson distributions. The first component of this mixture of distributions showed a RR of 7.9 (95% CRI: 5.7 - 10.8) with weight 0.30 (95% CRI: 0.11 – 0.61), the second component showed a RR of 18.8 (95% CRI: 17.3 – 20.8) with weight 0.60 (95% CRI: 0.30 - 0.89), whereas the third component presented a RR of 103 (95% CRI: 88.2 – 119.3) with weight 0.10 (95% CRI: 0.01 – 0.30).

Classification of the studies a posteriori and identification of sources of heterogeneity

Table 3.6 shows the a posteriori classification of the studies into the components detected in the mixture of probability distributions. Four studies constituted the first component, six the second and one the third.

All the studies classified into the first component used general population as comparison group and were carried out in developed countries (3 in Japan and 1 in the U.S.).

In the second component there has been classified the studies for which subjects of the comparison group have been tested for HBsAg. The comparison group was constituted by blood donors in three studies carried out in developed countries (U.S., Japan and Spain). In this component, two studies were conducted in China and Taiwan. Comparison group of these studies were GP HBsAg-negative (subjects tested at the beginning of the study)[68,74]. The study of U.K. was also classified in this component. This study used a comparison group which was not tested for HBsAg but it should be noted that HBsAg seroprevalence in the population of U.K. is one of the lowest among European countries (about 0.1%)[23]. The risk estimated in this study is the second highest and it has been classified into the second component. Finally, the study carried out in Taiwan with a comparison group constituted by general population HBsAg-negative reported an SMR of 103, comprised the third component.

Based on this classification we found that was reasonable to explain heterogeneity in terms of i) comparison group used in the study (GP or WBD) and ii) geographical area. We defined two geographical areas of risk by PLC labeled as low and high risk. Low risk area, in terms of PLC mortality, included studies conducted in Europe, U.S. and Japan, whereas studies carried out in China and Taiwan were included in high risk areas for PLC.

Table 3.6. Posterior classification of the studies into the components detected in the mixture analysis.

Author	Geographical Area	Comparison Group	Component number	Probability of component
Prince, 1982	U.S.	GP	1	0.52
Ijima,1984	Japan	GP	1	0.65
Oshima, 1984	Japan	GP	1	0.99
Dodd, 1987	U.S.	WBD	2	0.91
Sakuma,1988	Japan	WBD	2	0.93
Tokudome,1988	Japan	GP	1	0.99
Beasley,1991	Taiwan	WBD	3	1
Evans,2002	China	GP	2	0.99
Yang,2002	Taiwan	GP*	2	0.99
Crook, 2003	UK	GP*	2	0.96
Ribes,2006	Spain	WBD	2	0.95

WBD: workers or blood donors; *GP:* general population; *GP*:* General Population tested for HBsAg; *Component number:* Component of the mixture with maximum posterior probability of classification for the ith study ; *Probability of component:* probability of c*omponent number* for the ith study.

Covariates modeling of the SMRs

Estimates of the parameters of the model are shown in *Table 3.7.* It was observed that studies carried out in countries with low risk for PLC which had GP as comparison group showed an RR of death by PLC of 10.2 (95 % CRI: 7.9 – 12.8). This RR was 1.8-

fold higher (95% CRI: 1.4 – 2.3) for those studies carried out in high risk countries for PLC mortality, which showed a RR of death by PLC of 18.6 (95 % CRI: 17.3 – 20.1).

Table 3.7. Effects of comparison group and geographical area in the estimation of risk of death by PLC considering the parameters estimated by the covariates modeling.

| | Comparison Group | | |
Geographical Area	Blood donors /workers RR	Ratio[2]	General Population RR
High risk	103.0 (88.6 – 118.3)	5.5 (4.6 – 6.5)	18.6 (17.3 – 20.1)
Ratio[1]	5.3 (3.4 – 7.9)		1.8 (1.4 – 2.3)
Low risk	20.1 (13.5 – 21.6)	1.9 (1.2 – 3.1)	10.2 (7.9 – 12.8)

High Risk for PLC: High risk area of death by PLC (China and Taiwan); **Low Risk for PLC:** Low risk area for of death by PLC (Europe, U.S. and Japan). **Ratio[1]:** RR of death by PLC in a High risk area versus RR in a Low risk area. **Ratio[2]:** RR of death by PLC for studies which used Workers or Blood Donors as comparison group versus RR in studies which used general population.

If the comparison group was constituted by WBD, the RR was 20.1 (95% CRI: 13.5 – 21.6) for those studies carried out in low risk area of death by PLC, whereas the RR was 5.3-fold higher (95% CRI: 3.4 – 7.9) if the study was carried out in a high risk area of death by PLC (RR=103.0, 95% CRI: 88.6 – 118.3).

In order to assess the "healthy donor effect" we have estimated the Ratio of RR of death by PLC for studies which used WBD as comparison group versus that RR in studies which used general population as comparison group. This ratio was 1.9 (95% CRI: 1.2 – 3.1) in low risk areas for risk of death by PLC whereas for areas with high risk for PLC it was 5.5 (95% CRI: 4.6 – 6.5).

3.6. Discussion

3.6.1. Statistical analysis

This analysis illustrates a way to explore heterogeneity in a pooled analysis by means of the mixture of probability distributions approach based on Bayesian methods. These methods allowed for the assessment of heterogeneity between studies without relying on the assumption of a Normal distribution, as in classical meta-analysis. Using a nonparametric approach, we have identified the number of components of the mixture and we estimated a RR_{Pooled} of death due to PLC associated with HBV. Having identified the mixing distribution, each study has been classified into one of the components assuming that geographical area and comparison group were the most important sources of heterogeneity. On this basis and by means of a generalized linear model, RRs accounting for sources of heterogeneity accross studies have been reported.

Classical approaches to meta-analysis assume Normal distributions for the measure of interest extracted from each study. However, in this study our measure of interest was the SMR, which is more likely to be related with the Poisson distribution than the Normal distribution, as described in the methods section.

We should note that a key aspect of the meta-analysis is the use of mixtures of probability distributions. This method has allowed us i) to estimate a RR_{Pooled}, and ii) to explore sources of heterogeneity by means of the a posteriori classification using the Bayes rule of classification a posteriori (see equation 6). We performed this analysis using both Frequentist and Bayesian approaches. With the Frequentist approach we were not able to obtain a standard error for the RR_{Pooled}, although we did not attempt some alternative solutions such as the use of bootstrap methods. It would be interesting to implement this resampling method in order to compare the results extracted using bootstrapping with those obtained with the Bayesian approach.

We have shown the efficiency of the Bayesian approach to estimate the RR_{Pooled} using the "stick-breaking" method. In the Bayesian approach to mixtures of distributions, the "stick-breaking" method allows for empty components[142]. This is a difference with parametric mixture models estimated with DPP, for which we should know the number of components a priori. The main limitation of the "stick-breaking" method is the prior distribution assumed for the δ parameter, which is the spread of the base measure[145]. It has been described that some numerical problems arise when it is assigned a prior distribution for this parameter. In particular, if small values for that parameter are drawn then the cumulation of products $\phi_1 = r_1, \phi_2 = r_2(1 - r_1), ..., \phi_k = r_k(1 - r_k)$ obtained through sampling on a $Beta(1, \delta)$ may generate small numbers[142]. In order to avoid this problem we considered $\delta \sim Gamma(1,1)$. Ishwaran and James[144] discussed other sampling strategies for r_j such as taking $Beta(a_j, b_j)$ with $a_j = 1 - \gamma$ and $b_j = j \cdot \gamma$ with $0 \leq \gamma \leq 1$. We tried different prior distributions for δ such as $\delta \sim Gamma(0.001, 0.001)$ in a previous analysis. We found that convergence was not achieved (data not shown) for this parameter with this prior distribution using three dispersed chains. We also observed the same convergence problems when we tried $\delta \sim Uniform(0, L)$ for several values of L, such as 1, 5 and 10. We also considered $\delta = 1$ and $\delta = 5$ as an alternative to set δ as a random variable, which lead to almost identical results to those obtained with δ random. We have not chosen large values for δ because it would lead to find more components that may result an overfitting[142]. We observed that it was enough to select a few trial values of δ to assess their impact on the number of components[142]. We finally recommend to consider δ as a random variable due to its possibly posterior variability, as we observed in our results (Table 3.5, median of $\delta = 1.5$ with 95% CRI:0.2-4.2). In addition, we have found that sensivity in posterior estimates of $P = [p_1, ..., p_k]$ and $M = [\mu_1, ..., \mu_k]$ was not affected by variations in δ when the maximum number of components was set to three.

The posterior distributions of the unknown parameters of the mixture of Poisson distributions can be estimated also with reversible jump algorithms[149,146], which also allow to vary the dimension of the mixture[149,146]. In this last approach, the prior distribution for the number of components is a discrete $Uniform(1,...,K)$, assuming also Dirichlet distribution for the weights of the mixture, and Gamma or Uniform prior distributions for the $M = [\mu_1,...,\mu_k]$ parameters[146]. This last assumption entails a difficulty related with find ninonformative prior distributions for the Gamma parameters. It has also been described that changing the values of this prior distribution lead to different posterior distributions for K, P and M. However, in our analysis we have used prior Normal and prior Gamma for the $M = [\mu_1,...,\mu_k]$ parameters, and both lead to similar posterior distributions for those parameters. This could be due to the "stick-breaking" method of our analysis, for which the key point is the prior distribution of the δ parameter, as we noted previously.

Sources of heterogeneity were explored through the rule of posterior classification of the studies. This procedure has some resemblance with graphical explorations of heterogeneity in meta-analysis[133]. By means of the classification a posteriori of the studies we assessed that geographical area where the study was carried out and comparison group selected were the main sources of heterogeneity. The Bayesian modeling allowed to obtain credibility intervals for the RRs between areas and comparison groups.

In summary, we recommend to use the "stick-breaking" method in order to obtain a pooled measure of interest in meta-analysis because it allows also to identify the sources of heterogeneity. In a future work we should compare results of the Frequentist approach based on bootstrap with those obtained with the Bayesian approach to determine which are the differences in both methods in terms of variance of the parameter estimates.

3.6.2. Epidemiological discussion

The pooled RR of death by PLC estimated in our study was 23.5 (95% CRI: 14.9 - 44.5) in male HBV carriers. Variability in RR of death was substantially explained by geographical area and comparison group selected in the study.

Longitudinal studies carried out in high PLC risk areas showed higher risk of death associated with HBV infection than those carried out in low risk areas, independently of the comparison group used. The PLC mortality among HBV male carriers in high risk areas was 1.8-fold higher (95% CRI: 1.4 – 2.3) than that in low risk areas. However, if the comparison group was WBD, the mortality risk from PLC among HBV carriers was 5.3-fold higher (95% CRI: 3.4 – 7.9) in high risk areas than in low risk areas.

This excess PLC risk detected in some geographical areas may be aprtly attributable to differences in age at HBV infection and various environmental factors. The route of transmission of HBV varies accross geographical areas depending on the prevalence of HVB carriers in the general population. In Southeast Asia, HBV transmission mainly occurs by perinatal infection through the mother-to-child transmission and in the first years of life. In contrast, in developed areas, HBV acquisition tends to occur in adolescence or adulthood through sexual contact, blood transfusions or other invasive procedures under non-sterile conditions[74]. the longer duration of HBV infection could partly explain the excess risk of death by PLC among HBV male carriers reported in studies of Southeast Asia[73]. On the other hand, the exposure to aflatoxins in food in Southeast Asia is also common [150]. Several epidemiologic studies suggested that aflatoxin exposure among HBV carriers leads to an increased risk of progression to PLC when compared with unexposed HBV carriers [87,150-152].

The estimated RR of PLC death linked to HBV infection in the male cohorts from Japan, where PLC is mostly associated with HCV infection, ranges from 7 to 30. A

cohort study of HBsAg-positive Japanese females also confirmed the association between HBV and PLC (RR: 5.6, p<0.05) in women[65]. The RRs obtained in Japanese studies are similar to those reported in the North-American (RR: 9.70 to 26.80) and European studies (RR: 14.14 to 26.31)[21-23,58,59]. Cohort studies conducted in the U.K.[23] and U.S.[58,59] showed that migrants had higher RRs of death by PLC than the host population. The U.K. study included 141 (11% of the cohort) foreign born subjects. The RR of death by PLC among male HBsAg-positive migrants (SMR=51.91, p<0.001) was 2-fold higher than the risk observed among men born in U.K. (SMR=22.54, p<0.001)[23]. One of the U.S. studies observed 6 deaths by PLC, of which 4 were black men, 1 was of Asian origin and 1 was white[59]. These results are consistent with those found in other migrant studies. Several studies reported 1.3 to 10.9-fold excess mortality by PLC among Asian migrants to U.S., Australia and Europe[2].

The studies conducted in the U.K. and Spain reported similar RRs of death by PLC (26.31 in U.K. and 14.14 in Spain)[21,23]. In contrast with these results, in a recent study carried out in Italy which used a cohort of 296 blood donors, it was found that HBsAg-positivity was not associated with a higher mortality due to liver disease after 30 years of follow-up[24]. In that study, however, PLC incidence among HBsAg-positive subjects was 33.80 per 10^5 person-years (calculated from the data reported in the study) similar to that obtained in the cohort from the UK (33.1 per 10^5 person-years)[153] and in the Spanish cohort (34.1 per 10^5 person-years)[21]. However, the Italian study had the highest incidence rate in HBsAg-negative subjects (28.9 per 10^5 person-years, based on 1 subject with a daily alcohol intake greater than 60 g/day on 120 controls) among the studies conducted worldwide (range: 3.6-10.4), being very similar to the PLC incidence of their HBsAg-positive subjects[21]. The similarity of the Italian PLC incidence rate with those found in other western cohort studies of HBV carriers suggests that the lack of excess PLC risk associated with HBV could be due to the limited number of subjects which minimizes its statistical power, and increases susceptibility to error and bias.

We also demonstrated the underestimation of RRs in those studies which used the general population as a comparison group. This underestimation was observed among studies carried out in all countries with high and low PLC risk. In low risk areas for PLC, the mortality risk by PLC in studies which used workers or blood donors as the comparison group was 1.9-fold higher (95% CRI: 1.2 – 3.1) than in studies which used general population. However, in studies carried out in a high risk areas, the mortality risk from PLC was 5.3-fold higher (95% CRI: 3.4 – 7.9) in studies which used workers or blood donors as comparison group than that in studies which used general population. This is the first time that the "healthy donor effect" has been quantified in longitudinal studies. Voluntary blood donors and workers are generally healthier than the general population[154,155] and this result proves that a selection bias could be introduced when mortality in workers/blood donors is compared with rates of general population.

It si necessary to take into account some limitations of this analysis:

First, we have estimated the pooled RRs without taking into account the years of follow-up in each study. Two studies conducted in U.S.[58,59] had a very short follow-up (range of follow-up 3.6 to 4.3 years). However, all other studies included in the meta-analysis had a follow-up greater than 20 years (the cohorts from the U.K.[153] and Spain[21] with 22.4 and 20.5 years , respectively). The initial study in the U.K. cohort (published in 1985) reported a risk of death by PLC of 42-fold after 7.3 years of follow-up[22]. An update to this study (published in 2003) reported a risk of death by PLC of 26.3-fold after 30 years of follow-up[153]. This decrease in the RRs of death from PLC was also observed in the Taiwan cohort study which used WBD as comparison group. The initial study (published in 1981), which had 3.5 years of follow-up, reported a RR of death by PLC of 223[66], whereas in 1991, after 11 years of follow-up, this RR dropped to 103[130]. These results suggest that the pooled RR of our analysis could be overestimated.

A second limitation of our analysis is that we have estimated a pooled RR of death by PLC with few studies conducted in high PLC risk areas, given the lack of studies among HBV carriers in Africa.

In summary, we have found that 1) cohort studies conducted in geographical areas with high risk of death by PLC showed higher RRs of death by PLC than those conducted in low risk geographical areas independently of the comparison group selected, and 2) there is a "healthy worker/donor effect" that leads to a underestimation of RRs of death by PLC in those studies which selected general population as comparison group.

Chapter 4

4. Geographic distribution of primary liver cancer in Europe in 2002

4.1. Background

It has been estimated that in 2002[20] 53,618 new cases of PLC were diagnosed, which represents about 2% of the total new cancers diagnosed. This PLC incidence corresponded to 35,303 (65.8%) cases in men and 18,315 (34.2%) in women. The estimated number of PLC deaths during the same period were 57,486 during the same period, the incidence/mortality ratio being close to 1[20]. The greater number of deaths than incident cases in the same year could be explained because mortality data is based on death certificates which may be more prone to misclassification while incidence data is based on cancer registries with more quality controls on the data.

Three areas in Europe have been distinguished according to PLC incidence defined as high, intermediate and low risk areas[25]. The highest PLC incidence rates have been observed in Southern Europe (Men: 11.6 per 10^5 person-years; Women: 4.0 per 10^5 person-years)[20]. Greece and Italy showed the highest incidence (Greece: 12.9 per 10^5 person-years in men and 4.9 per 10^5 person-years in women; Italy: 15.9 per 10^5 person-years in men and 5.1 per 10^5 person-years in women) in this area.[20] Eastern and Western European countries have intermediate incidence rates (Men: 5.8 per 10^5 person-years; Women: 2.1 per 10^5 person-years)[20], France being the country with highest PLC incidence (Men: 10.5 per 10^5 person-years; Women: 2.2 per 10^5 person-years) in these areas. Northern European countries have the lowest European PLC incidence rates (Men: 3.4 per 10^5 person-years; Women: 1.7 per 10^5 person-years)[20]. Mortality pattern is very similar to that of incidence rates. In the comparison of rates between men and women, there is an excess of PLC incidence and mortality rates in men (ratio men to women close to 2)[20].

It has been estimated that, in Europe, near to 80% of PLC cases are related to HBV or HCV infections[83,156]. The estimated correlation of the PLC incidence rates between men and women in 38 European countries was 0.83 (p<0.001) [25], suggesting that risk

factors for PLC are similar in both sexes. *Figure 4.1* shows HBsAg-positive and anti-HCV positive prevalences in Europe[157]. Correlations between HBV and HCV prevalences and PLC incidence rates in men of the European Union, Norway, Iceland and Switzerland showed a significant association between both risk factors and PLC (Correlation Coefficient between PLC and HBV: 0.74, $p<0.01$, and between PLC and HCV: 0.63, $p<0.01$)[25].

In this chapter, the objective was to evaluate whether PLC risk in European regions was modified due to a joint exposure to HCV and HBV. This aim has been investigated through a spatial analysis of PLC incidence and mortality rates in Europe.

Figure 4.1. HBsAg-positive and anti-HCV prevalences in Europe.
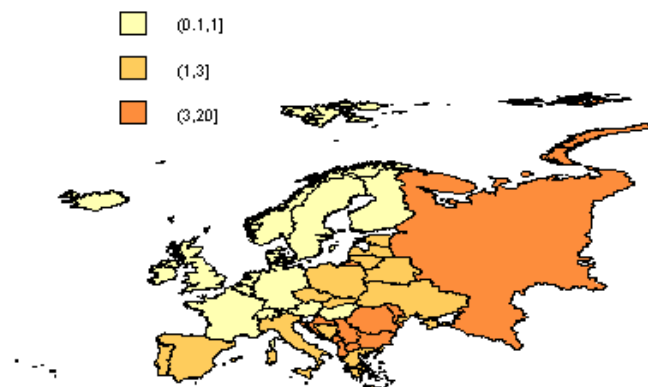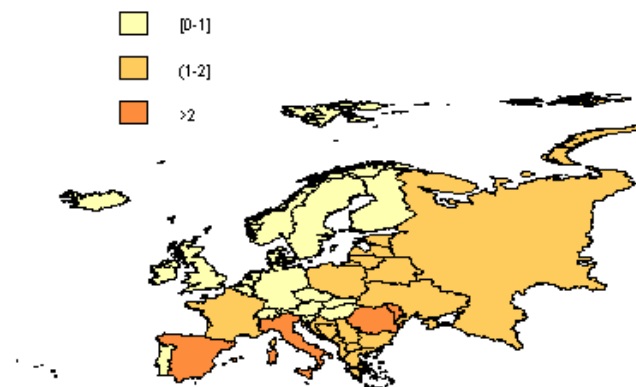
HBsAg-positive prevalence reproduced from Ribes et al, 2002[157].
Anti-HCV prevalence updated from data reported in Shepard et al, 2005[15]

4.2. Hypothesis

The variability of PLC incidence and mortality rates in Europe can be explained by heterogeneity of HBV and HCV prevalences among European countries.

4.3. Sources of data

Incidence and mortality data have been obtained through a combination of the reports of the population-based cancer registries collected in Globocan 2002[20] and the World Health Organization (WHO) mortality databank[27]. A total of 38 countries have been selected, and these are: Albania, Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Belarus, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Macedonia, Malta, Moldova, Serbia and Montenegro, The Netherlands, Norway, Poland, Portugal, Romania, Russian Federation, Slovakia, Slovenia, Spain, Sweden, Switzerland, United Kingdom and Ukraine. Age distribution in each country has been obtained by the WHO mortality databank[27]. Code selected for PLC based on the 10th revision of ICD was C22[53].

The estimated number of incident and death cases from PLC by age and sex were available for each country for the year 2002. PLC incidence rates were based on estimates of the neighbour countries for those which did not have National cancer registries or did not have region-based cancer registry[20]. Nation-wide cancer registry data was available for the period of analysis in Croatia, Czech Republic, Denmark, Estonia, Finland, Iceland, Ireland, Malta, The Netherlands, Norway, Slovenia, Sweden, United Kingdom and Ukrania. PLC incidence in Austria, Belgium, France, Italy, Spain, Poland and Switzerland was estimated with data from region-based

cancer registries in each one of these countries and their national mortality data. In those countries which mortality data was not available during 2002 as Bosnia and Herzegovina (last year available: 1991), Denmark (last year available: 2001) and Belgium (last year available: 1997), data were estimated on the basis of mortality rates of their neighbour countries.[27]

Incidence and mortality cases were grouped into five age-groups: 0-14, 15-44, 45-54, 55-64 and older than 64[20]. HBV and HCV seroprevalences were obtained from several WHO reports.[10-12,158] Updates of both HBV and HCV seroprevalences have been extracted from reviews published on global epidemiology of HBV[5] and HCV infection[15].

4.4. Statistical methods

The Standardized Incidence Ratio (SIR)[134] has been used to compare PLC incidence cases in each country with those expected using the European population, whereas the SMR has been considered for PLC mortality. SIR and SMR values have been calculated separately for men and women. These values were obtained by means of the following steps: first it was calculated the whole European sex-age-specific PLC incidence and PLC mortality rates; second, PLC expected number of incident and death cases have been estimated in each country by multiplying its population distribution (persons at risk to develop or to die from PLC) by the whole European PLC incidence and PLC mortality rates; and third, we have calculated the ratio between observed and expected number of PLC cases in each country.

SIR and SMR have been subsequently smoothed (model fitted values) accounting for effect of HBV and HCV prevalences. Random effects models based on Bayesian inference have been used to map the smoothed SIRs and SMRs. These methods have been described below on the basis of incidence data, although those are identical for mortality data.

4.4.1. Disease mapping and the Poisson model

Let O_i be the number of observed PLC incident cases and E_i the number of expected PLC incident cases in the ith country. The O_i are considered random variables whereas the E_i are considered fixed values which depend on n_i, the number of individuals at risk in the ith country. The estimation of the RRs has been performed through internal standardization[134].

The expected number of cases was calculated by

$$E_i = \sum_{j=1}^{5} n_{ij} r_j \qquad (16),$$

where n_{ij} is the number of person-years at risk of the ith country in their j^{th} age group, respectively, and r_j is the PLC incidence rate in the jth age group of the European population. This incidence rate is obtained through

$$r_j = \frac{\sum_{i=1}^{38} O_{ij}}{\sum_{i=1}^{38} n_{ij}} \qquad (17),$$

where O_{ij} is the observed number of incident cases of the jth age group in the ith country. Through (16)-(17) the relation $\sum_{i=1}^{38} E_i = \sum_{i=1}^{38} O_i$ holds.

For the ith country

$$SIR_i = \frac{O_i}{E_i} \qquad (18),$$

is the Standardized Incidence Ratio (SIR), which is an estimator of the RR of incidence.

The O_i are usually modelled by the Poisson distribution,

$$O_i \mid \lambda_i, E_i \sim \text{Poisson}(E_i \lambda_i) \qquad (19),$$

being λ_i the true relative risk of the disease in the ith country. In this model the SIR_i is the maximum likelihood estimator of λ_i [134]. Next section describes methods to

smooth the SIR values by modelling the implicit extra-Poisson variability through a regression procedure.

4.4.2. Modelling and mapping the relative risks

In order to estimate and to generate a map of the underlying relative risk surface $\{\lambda_i, i = 1,...38\}$ it is necessary to naturally think of a random effects model for the λ_i, assuming that all the true risks come from a common underlying distribution. The random effects can be estimated in terms of the Poisson likelihood through hierarchical Bayesian modelling. Thus, inference about $\{\lambda_1,..., \lambda_{38}\}$ has been based on the resulting posterior distribution.

As example, if $\lambda_i \sim Gamma(a,b)$, the posterior distribution of the estimate of the relative risk is $\lambda_i \mid O_i, E_i \sim Gamma(a + O_i, b + E_i)$. But this procedure does not account for the effect of available covariates, in this case the HBV and HCV population prevalences.

Let us consider $e^{\gamma_i} = \lambda_i$ where (19) is transformed in

$$O_i \mid E_i, \gamma_i \sim Poisson(E_i e^{\gamma_i})$$

$$\gamma_i = \beta \cdot X_i + h_i$$

(20),

where X_i is an explanatory spatial covariate with parameter coefficient β, and h_i captures the country wide heterogeneity. The β and h_i parameters have prior distributions

$$h_i \sim N(0, \tau_h)$$
$$\tau_h = \sigma_h^{-2}$$
$$\sigma_h \sim Uniform(0,100) \qquad (21),$$
$$\beta \sim N(0, \tau_\beta)$$
$$\tau_\beta = 0.0001$$

where τ parameters are the precision (inverse of the variance) terms. We assumed a non informative prior distribution Uniform (0, 100) for the standard deviations of the h_i (due to that perform better than gamma prior distributions in this analysis, see Sensibility Analysis on Appendix A.2), being h_i the random effects spatially unstructured. For the regression coefficients, β, we assumed prior low precision set to 0.0001[147]. In equation (21) the extra-Poisson variability is captured by h_i in the log-relative risks, which vary over the whole study area.

4.4.3. Model definition and procedure for data analysis

The procedure of analysis has been based on the selection of a regression model which allow us to generate a map of the model fitted relative risks. RRs have been stratified into four groups (strata) due to the data available: PLC incidence for men, PLC incidence for women, PLC mortality for men and PLC mortality for women. The observed and expected numbers of PLC deaths have been obtained (crude SIRs and SMRs) for each strata and country. It has not been possible to distinguish HBV and HCV prevalences according sex and age, because these data were available only at population level.

Most of models included an effect of geographic area defined on the basis of the European areas described in the Globocan 2002 software: Eastern, Northern, Southern and Western (see *Figure 4.2*)[20]

Figure 4.2. European geographical areas defined in Globocan 2002[159].
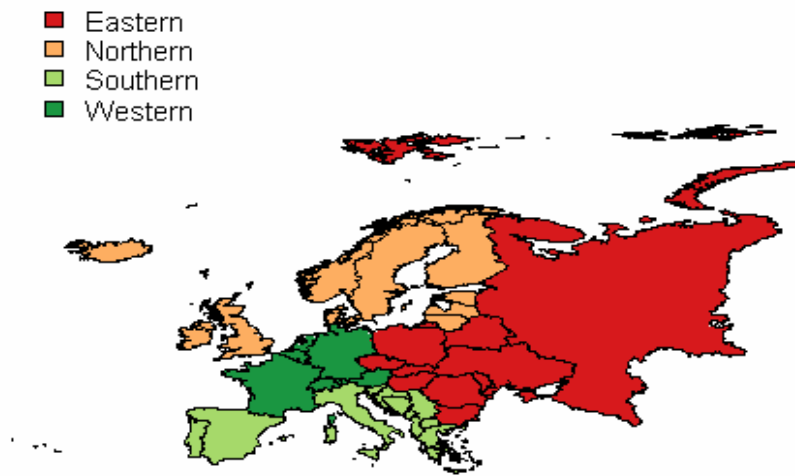
4.4.3.1. Model choice and the Deviance Information Criterion

Six different models have been considered in the regression analysis to select the model which explains variability of PLC accounting for HBV and HCV infections. HBV and HCV prevalences have been transformed into categorical covariates. HBV infections values were stratified into two categories: 0% to 2% and >2%. The cutting value of 2% was selected due to that was the median value for the HBV infection in those 38 European countries. Values for HCV categories were defined as: 0% to 1%, >1% to 2% and >2%. These cutting values were selected on the basis of those suggested in the most recent report on the estimation of HCV prevalences[15]. Models proposed based on the equation (21) were:

$$\text{M1) } \gamma_i = \mu + h_i,$$

$$\text{M2) } \gamma_i = \alpha_i + h_i,$$

$$\text{M3) } \gamma_i = \alpha_i + \beta_1 + h_i,$$

$$\text{M4) } \gamma_i = \alpha_i + \beta_2 + \beta_3 + h_i,$$

$$\text{M5) } \gamma_i = \alpha_i + \beta_1 + \beta_2 + \beta_3 + h_i,$$

$$\text{M6) } \gamma_i = \alpha_i + \beta_1 + \beta_2 + \beta_3 + \beta_{1,2} + \beta_{1,3} + h_i,$$

being μ a global mean, α_i a four level factor variable with $\alpha_i \in [\alpha_j \mid j = \{1,2,3,4\}]$ and related to which area (Eastern, Northern, Southern and Western) belongs the ith country, β_1 the effect of HBV prevalence > 2%, β_2 the effect of HCV prevalence ranging from >1% to 2%, β_3 the effect of HCV prevalence >2%, $\beta_{1,2}$ the effect of HBV prevalence >2% and HCV prevalence ranging from >1% to 2%, and $\beta_{1,3}$ the effect of HBV and HCV prevalences > 2%.

Note that M1 is the only model that does not take into account the effect of geographical area (model without random intercept α_i). By means of e^{α_j} we have estimated the median RR of incidence (or mortality) for those countries with HBV prevalence $\leq$ 2% and HCV prevalence $\leq$ 1% in the jth geographical area. In that sense, e^{β_1} is the increase (in multiplicative scale) of the median RR (IMRR) for those countries with HBV prevalence > 2%, whereas e^{β_2} and e^{β_3} is the IMRR for those countries with HCV prevalence > 1% and HCV prevalence > 2%, respectively. The interaction effects in terms of IMRR between HBV and HCV infections could be calculated by $\mathrm{e}^{\beta_{1,2}}$ and $\mathrm{e}^{\beta_{1,3}}$.

The prior distributions chosen for α_j and μ parameters are normal with zero mean and large prior variance (as equation 21 describes).

A tool for the model choice that recently has gained popularity has been the Deviance Information Criterion (*DIC*)[160]. *DIC* was proposed as a generalization of the Akaike Information Criteria based on the posterior distribution of the deviance statistic[147,148] which is

$$D(\theta) = -2\log L(O \mid \theta) + \log B(O) \qquad (22),$$

where θ is the vector of unknown parameters, O is the vector of observed values (incident or death cases), $L(O \mid \theta)$ is the likelihood function and *B(O)* some

standardizing function of the data alone. From (22) it has been suggested to summarize the fit of a model by the posterior expectation of the deviance $E(D)$ and the complexity of the model by pD, the "effective number of parameters" or the number of "unconstrained parameters" in the model. The DIC and pD are defined[160] as

$$DIC = E(D) + pD$$
$$pD = E(D) - D(\bar{\theta})$$

(23).

Lower values of DIC indicate better fit of the model, because low values of $E(D)$ indicate good fit and low pD values indicate that a more parsimonious model might be considered[160]. These values are easily estimated via MCMC methods and computed automatically in WinBUGS[148] via *bugs* function of library R2WinBUGS[137]. It should be noted that R2WinBUGS library approaches the estimate of the pD with the proposal of Gelman et al[137], which defines $\hat{pD} = \frac{1}{2}Var(D(y,\theta) \mid y)$. This last definition slightly differs from (23), the initial proposal of Spiegelhalter et al.[160] which was based in an asymptotic χ^2 distribution.

For each model it has been run 3 chains with 60,000 iterations, discarding the first 10,000 burning samples. Samples from every 10th iteration has been stored (being 10 the value for the thin[139,140,142,143,147,148] parameter in the run of the MCMC) in order to reduce autocorrelation in the sample and to reduce Monte Carlo errors[147,148]. Sample traces plots and the Gelman and Rubin convergence diagnosis[147,148] have been used to check for convergence of the chains. After convergence the 95% credibility interval has been obtained for each parameter jointly with its median value. Statistical significance for the model parameters has been established on the basis of the 95% credibility interval of these parameters. If that interval does not include the value 0 (or the exponential scale of this parameter does no include the value 1), we have then assumed that the parameter is statitically significant.

Convergence of the chains has been verified with the Scale Reduction Factor (SRF)[137,147]. In this line, using m overdispersed chains and running them for $2N$ iterations, convergence of the parameters could be monitorized through

$$\sqrt{\hat{SRF}} = \sqrt{\left(\frac{N-1}{N} + \frac{m+1}{mN}\frac{B}{W}\right)\frac{df}{df-2}} \,,$$ where B/N is the variance between the means

from the m parallel chains, W is the average of the m within-chain variances, and df is the degrees of freedom of an approximating $t-Student's$ density to the posterior distribution. The SRF should approach 1 as $N \rightarrow \infty$, but if the potential SRF is high, then there would be necessary further simulations to improve inference about the target distribution for the parameter of interest[137,147].

Section A.3 of the Appendix shows the WinBUGS code for the final model considered in this study.

Two PLC risk maps have been done for each combination gender-incidence/mortality PLC rates: one map with the fitted RRs and another with $P(RR_i > 1 \,|\, Data)$. Inspection of the first map has allowed us to identify those areas with similar relative risks, whereas inspection of the second map has allowed us to identify those areas with high and low probability of PLC risk.

4.5. Results

a) Descriptive analysis

Table 4.1 describes correlations among variables of the study in each European area. Eastern, Northern and Western European countries showed positive correlation between PLC SIRs and PLC SMRs and HBV and HCV prevalences for both sexes.

Table 4.1. Correlations between HCV and HBV infections, SIR and SMR of PLC by European area and sex.

	European Area			
	Eastern	Northern	Southern	Western
Men				
SIR (mean, SD)	0.91 (0.49)	0.45 (0.16)	1.18 (0.58)	0.88 (0.41)
SMR (mean, SD)	0.88 (0.38)	0.48 (0.17)	1.14 (0.35)	0.97 (0.48)
Correlations between:				
HBV and SIR	0.34	0.38	-0.28	0.16
HBV and SMR	0.56	0.32	-0.29	0.24
HCV and SIR	0.62	0.41	0.14	0.53
HCV and SMR	0.52	0.31	0.34	0.71
Women				
SIR (mean, SD)	0.99 (0.45)	0.56 (0.26)	1.18 (0.52)	0.64 (0.36)
SMR (mean, SD)	0.96 (0.41)	0.61 (0.25)	1.29 (0.43)	0.75 (0.22)
Correlations between:				
HBV and SIR	0.81	0.36	-0.03	0.02
HBV and SMR	0.44	0.04	-0.26	0.33
HCV and SIR	0.56	0.41	0.41	-0.12
HCV and SMR	0.41	0.06	0.31	0.72
Viral infections				
HBV (mean, SD)	3.39 (2.51)	0.88 (1.14)	5.14 (5.10)	0.37 (0.14)
HCV (mean, SD)	1.56 (1.41)	0.45 (0.75)	1.24 (0.68)	0.51 (0.43)
Correlations between:				
HBV and HCV	0.64	0.98	0.12	0.45

mean: mean value; **SD:** standard deviation; **HBV:** hepatitis B Virus prevalence; **HCV:** hepatitis C Virus prevalence; **SIR:** Standardized Incidence Ratio; **SMR:** Standardized Mortality Ratio.

In Southern European countries, a negative correlation between PLC RRs and HBV prevalence was detected jointly the lowest correlation between HBV and HCV infections (0.12%). A negative correlation (-0.12%) between SIR of PLC and HCV prevalence has been observed in women of Western area.

The highest mean PLC SIR and mean PLC SMR values were found in Southern European countries (rank: 1.14 to 1.29), followed by Eastern and Western European countries (rank: 0.64 to 0.99), whereas the lowest mean RRs were found in Northern European countries (rank: 0.45 to 0.61).

b) Model Selection

In order to assess the impact of the inclusion of parameters in the model we focused the model selection criterion on the reduction of the a *posteriori* standard error of the spatially unstructured random effects ($\tau_h^{-\frac{1}{2}}$) jointly with the *DIC* and *pD* criterion. *Table 4.2* shows the model selection procedure for all combinations of gender-incidence/mortality.

The pattern of model selection in all these combinations was identical. Model M1 showed the highest standard errors of the spatially unstructured random effects and the highest *DIC* values. The inclusion of the geographical area covariates (M2-M6) reduced these standard errors and the *DIC* value. It was also observed a *pD* lower than the total parameter counts (all of them include at least 38 random effects, one for each country) in models M2-M6.

Table 4.2. Model selection procedure.

			Men			Women
			PLC Incidence			
Model	DIC	pD	$((\tau_h)^{-\frac{1}{2}}\|y)$ (95%CRI)	DIC	pD	$((\tau_h)^{-\frac{1}{2}}\|y)$ (95%CRI)
M1	450.3	40.4	0.47 (0.38-0.59)	429.5	41.3	0.46 (0.36-0.59)
M2	399.4	37.7	0.35 (0.22-0.38)	396.8	40.7	0.37 (0.29-0.39)
M3	391.3	39.2	0.28 (0.22-0.38)	370.6	40.3	0.29 (0.28-0.38)
M4	391.1	37.6	0.27 (0.18-0.31)	372.1	40.5	0.29 (0.18-0.35)
M5	**369.2**	**36.1**	**0.17 (0.11-0.23)**	**354.9**	**39.1**	**0.18 (0.12-0.25)**
M6	370.1	36.3	0.16 (0.10-0.22)	355.3	39.0	0.17 (0.10-0.24)
			PLC Mortality			
M1	441.1	40.3	0.41 (0.32-0.51)	426.2	37.9	0.45 (0.28-0.48)
M2	397.4	37.1	0.34 (0.19-0.34)	388.9	34.7	0.38 (0.14-0.24)
M3	385.1	38.1	0.29 (0.19-0.32)	370.7	37.6	0.31 (0.14-0.27)
M4	384.2	36.9	0.28 (0.16-0.28)	371.8	35.4	0.28 (0.13-0.27)
M5	**362.9**	**34.8**	**0.19 (0.15-0.26)**	**354.1**	**33.8**	**0.17 (0.12-0.22)**
M6	363.1	34.3	0.19 (0.15-0.26)	357.4	33.5	0.16 (0.09-0.23)

DIC: Deviance Information Criterion; **pD:** effective number of parameters; $(\tau^{-\frac{1}{2}}|y)$: *posteriori* standard error of the spatially unstructured random effects (*y* refers to observed data); **95% CRI:** 95% credibility interval.

The last model, M6 (model whith the interaction of HBV and HCV prevalences), showed the lowest standard errors of the spatially unstructured random effects. However, these interaction terms had wide credibility intervals. For that reason, M6 was not the selected model due to M5 (model without interaction parameter) also showed almost identical *DIC* and *pD* value.

Table 4.3 shows the median PLC RRs of each geographical area and the IMRR for European countries with HBV prevalence > 2% and HCV prevalence > 1%. For men, Northern countries showed median PLC RRs lower than 1 (median RR Incidence: 0.48, 95% CRI: 0.37 to 0.61; median RR Mortality: 0.51, 95% CRI: 0.41 to 0.60) and with probability of RR>1 lower than 0.001. Median PLC RRs for the remaining areas included the value 1 on the 95% CRI. Countries with HBV prevalence > 2% and those with HCV prevalence ranging from 1 to 2% did not show an increase of PLC RRs in front of those countries with lower prevalences. However, countries with HCV

prevalence > 2% showed an increase of the median RR (IMRR Incidence: 1.78, 95% CRI: 1.15 to 2.73; IMRR Mortality: 1.48, 95% CRI: 1.14 to 1.93). Similar results were observed for women, with the exception of Western area which showed median PLC RRs lower than 1 (median RR Incidence: 0.65, 95% CRI: 0.49 to 0.85; median RR Mortality: 0.71, 95% CRI: 0.56 to 0.89).

Table 4.3. Model M5: Median PLC relative risks and increase (in multiplicative scale) of the median relative risk due to the effects of geographical area and HBV and HCV prevalences.

	Men	Women
	PLC Incidence	
European Area	Median RR (95% CRI)	Median RR (95% CRI)
Eastern	0.81 (0.63 – 1.07)	0.91 (0.67 – 1.21)
Norhtern	0.48 (0.37 – 0.61)	0.55 (0.42 – 0.71)
Southern	0.94 (0.71 – 1.21)	0.95 (0.75 – 1.31)
Western	0.75 (0.57 – 1.02)	0.65 (0.49 – 0.85)
Prevalence	IMRR (95% CRI)	IMRR (95% CRI)
HBV ≤ 2%	Reference	Reference
HBV > 2%	1.02 (0.81 – 1.32)	1.15 (0.89 – 1.57)
HCV ≤ 1%	Reference	Reference
HCV 1 - 2%	1.01 (0.78 – 1.28)	1.05 (0.81 – 1.42)
HCV > 2%	1.78 (1.15 – 2.73)	1.36 (1.09 – 2.25)
	PLC Mortality	
European Area	Median RR (95% CRI)	Median RR (95% CRI)
Eastern	0.83 (0.65 – 1.04)	0.94 (0.75 – 1.17)
Norhtern	0.51 (0.41 – 0.60)	0.65 (0.53 – 0.79)
Southern	0.96 (0.76 – 1.19)	1.09 (0.87 – 1.36)
Western	0.82 (0.67 – 1.03)	0.71 (0.56 – 0.89)
Prevalence	IMRR (95% CRI)	IMRR (95% CRI)
HBV ≤ 2%	Reference	Reference
HBV > 2%	1.05 (0.85 – 1.34)	1.03 (0.84 – 1.31)
HCV ≤ 1%	Reference	Reference
HCV 1 - 2%	1.06 (0.86 – 1.36)	1.06 (0.86 – 1.33)
HCV > 2%	1.48 (1.14 – 1.93)	1.28 (1.05 – 1.75)

Median RR: PLC Relative Risk for countries with HBV prevalence < 2% and HCV prevalence < 1%; **IMRR:** Increase (in multiplicative scale) of the Median Relative Risk; **HBV:** Hepatitis B Virus; **HCV:** Hepatitis C Virus.

Based on M5 model, *Table 4.4* shows the *a posteriori* RRs of incidence and mortality by PLC for both sexes in Eastern and Northern European countries, jointly with the *a posteriori* probability of RR greater than 1. Median RRs for European Areas have been obtained from the RRs of those countries with HBV ≤ 2% and HCV ≤ 1%.

Table 4.4. Eastern and Northern European countries: *a posteriori* Relative Risks of incidence and mortality by PLC for both sexes and their *a posteriori* probability of Relative Risk greater than 1.

	PLC Incidence				PLC Mortality			
	Men		Women		Men		Women	
Eastern Europe	RR_I	$P(RR_I >1)$	RR_I	$P(RR_I >1)$	RR_D	$P(RR_D >1)$	RR_D	$P(RR_D >1)$
Belarus	0.73	0.02	0.69	0.01	0.66	0.02	0.62	0.01
Bulgaria	0.73	0.05	0.97	0.32	1.07	0.94	1.13	0.98
Czech Republic	1.08	0.96	1.05	0.81	1.15	0.99	1.26	0.99
Hungary	1.45	0.99	1.61	0.99	1.16	0.98	1.31	0.95
Moldova	1.69	0.98	1.49	0.98	1.27	0.95	1.15	0.93
Poland	0.54	0.01	0.95	0.17	0.64	0.02	1.12	0.95
Romania	1.48	0.99	1.51	0.99	1.31	0.99	1.36	0.99
Russian Federation	0.68	0.02	0.92	0.05	0.86	0.08	1.04	0.98
Slovakia	0.99	0.44	1.14	0.88	1.01	0.53	1.06	0.75
Ukrania	0.49	0.01	0.62	0.03	0.42	0.01	0.53	0.02
Northern Europe	RR_I	$P(RR_I >1)$	RR_I	$P(RR_I >1)$	RR_D	$P(RR_D >1)$	RR_D	$P(RR_D >1)$
Denmark	0.63	0.02	0.85	0.11	0.52	0.01	0.79	0.02
Estonia	0.61	0.02	0.65	0.03	0.64	0.02	0.73	0.04
Finland	0.51	0.01	0.55	0.01	0.62	0.01	0.99	0.43
Iceland	0.36	0.03	0.36	0.03	0.46	0.05	0.54	0.07
Ireland	0.36	0.01	0.36	0.01	0.46	0.02	0.54	0.02
Latvia	0.59	0.01	0.75	0.02	0.65	0.01	0.68	0.02
Lithuania	0.49	0.01	0.69	0.03	0.57	0.02	0.62	0.03
Norway	0.37	0.02	0.43	0.03	0.32	0.01	0.47	0.03
Sweden	0.55	0.01	0.75	0.06	0.62	0.03	0.82	0.08
United Kingdom	0.57	0.01	0.76	0.02	0.65	0.01	0.63	0.02

RR_I: Relative Risk of PLC Incidence; **RR_D:** Relative Risk of PLC mortality; **P:** *a posteriori* probability of RR > 1.

Among Eastern European countries, Czech Republic, Hungary, Moldova, Romania and Slovakia showed the highest RRs of PLC incidence for both sexes with high posterior probability of RR greater than 1. In men, mortality data in this area showed

a similar pattern. In women, all countries showed RRs greater than one with the exception of Belarus (RR death by PLC: 0.62) and Ukrania (RR death by PLC: 0.53).

The lowest RRs of PLC incidence among Northern European countries were found in Iceland and Ireland with identical RR (RR: 0.36) for both sexes. The highest RRs of PLC incidence in men were found in Denmark (RR: 0.63), Estonia (RR: 0.61), United Kingdom (RR: 0.57) and Sweden (RR: 0.55). PLC incidence in women showed a similar pattern but adding Latvia (RR: 0.75) to those countries with high PLC risk. Finland was the country with highest RR of death by PLC in women (RR:0.99) with a posterior probability of 0.43 of RR greater than 1.

Among Southern European countries (*Table 4.5*), Italy and Greece showed the highest RRs of PLC incidence (RR men Italy: 2.37; RR Women Italy: 2.09; RR men Greece: 1.91; RR Women Greece: 1.81) and PLC mortality (RR men Italy: 1.87; RR Women Italy: 1.78; RR men Greece: 1.67; RR Women Greece: 1.76), these RRs being the highest for both sexes among all Europe. In the same area, Spain showed high RRs of PLC incidence (RR: 1.36) and PLC mortality (RR: 1.25) in men, followed by Croatia (RR incidence: 1.07; RR death: 1.37) and Macedonia (RR incidence: 0.96; RR death: 1.06). Malta, Portugal and Slovenia showed PLC RRs lower than 1 in women, with also low probability of RR greater than 1.

Table 4.5. Southern and Western European countries: *a posteriori* Relative Risks of incidence and mortality by PLC for both sexes and their *a posteriori* probability of Relative Risks greater than 1.

| | PLC Incidence | | | | PLC Mortality | | | |
| | Men | | Women | | Men | | Women | |
Southern Europe	RR_I	$P(RR_I >1)$	RR_I	$P(RR_I >1)$	RR_D	$P(RR_D >1)$	RR_D	$P(RR_D >1)$
Albania	0.79	0.01	1.14	0.82	0.97	0.37	1.33	0.99
Bosnia & Herzegovina	0.84	0.05	1.14	0.89	0.99	0.46	1.25	0.98
Croatia	1.07	0.86	1.37	0.99	1.08	0.89	1.13	0.93
Greece	1.91	0.99	1.87	0.99	1.67	0.99	1.76	0.99
Italy	2.37	0.99	2.09	0.99	1.87	0.98	1.78	0.96
Macedonia	0.96	0.34	1.12	0.79	1.06	0.71	1.31	0.99
Malta	0.48	0.01	0.49	0.03	0.53	0.01	0.47	0.01
Portugal	0.82	0.12	0.68	0.03	0.81	0.09	0.67	0.05
Slovenia	0.76	0.01	0.61	0.01	0.94	0.23	0.81	0.09
Spain	1.36	0.99	1.12	0.95	1.25	0.96	1.16	0.98
Serbia and Montenegro	0.75	0.34	1.15	0.99	0.97	0.21	1.38	0.99
Western Europe	RR_I	$P(RR_I >1)$	RR_I	$P(RR_I >1)$	RR_D	$P(RR_D >1)$	RR_D	$P(RR_D >1)$
Austria	1.17	0.98	1.12	0.96	1.06	0.95	0.91	0.04
Belgium	0.63	0.02	0.81	0.05	0.68	0.03	0.81	0.09
France	1.55	0.99	0.98	0.21	1.69	0.99	0.96	0.18
Germany	0.63	0.01	0.66	0.06	0.73	0.08	0.79	0.09
Luxembourg	0.64	0.01	0.61	0.01	0.75	0.03	0.73	0.02
The Netherlands	0.27	0.02	0.25	0.01	0.37	0.03	0.46	0.04
Switzerland	0.85	0.25	0.65	0.01	0.87	0.15	0.59	0.01

RR_I: Relative Risk of PLC Incidence; **RR_D:** Relative Risk of PLC Mortality; **P:** *a posteriori* probability of RR >1.

Among Western European countries (see *Table 4.5*), the highest RR of PLC incidence and PLC mortality in men was found in France (RR incidence: 1.55; RR death: 1.69), this country being the third in the ranking of European countries with high incidence and mortality by PLC in men. However, the RR of incidence and death by PLC for women in this country showed low posterior probability to be greater than 1. Austria also had high RRs of incidence (RR: 1.17) and death (RR: 1.06) by PLC in men and the highest RR of PLC incidence in women in Western Europe (RR: 1.12) with high posterior probability of RR of PLC incidence greater than 1 (0.99). The lowest RRs of PLC incidence in men (RR: 0.27) and women (RR: 0.25) among all European countries

were observed in The Netherlands. RR of death by PLC in this country was also the lowest among Western countries in men (RR: 0.37) and women (RR: 0.46).

Figure 4.3. Distributions of the log-SIR and log-SMR values and their posterior model fitted log relative risks.
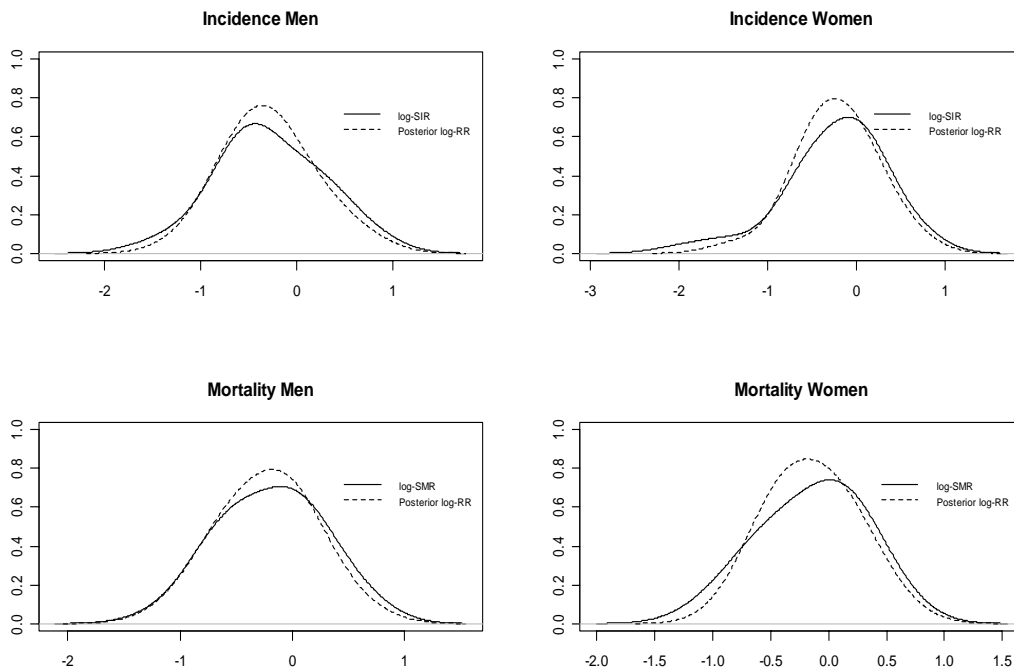
Figure 4.3 depicts the effect of the smoothing of the PLC SIR and PLC SMR values, comparing the distribution (density estimates) of log-SIR and log-SMR values for both sexes jointly with their corresponding posterior log-RRs. Note the robustness of the log-SIR and log-SMR values, which is reflected in the similarity of the crude probability density and posterior (model fitted log-RRs) density of these values.

Figure 4.4 shows the projected RRs of incidence and death by PLC for both sexes. RRs showed a gradient north-south for all strata of PLC data and for both sexes. Southern

Europe, some countries of Eastern Europe and Austria and France had the highest RRs of incidence and mortality by PLC.

Figure 4.5 shows the map with the *a posteriori* probability of RR of PLC data in order to assess the magnitude of the fitted PLC RRs. For both, incidence and mortality data, countries with posterior probability greater than 50% for PLC RR greater than 1 were located in Southern European and some Eastern European countries.

Figure 4.4. Maps of the a *posteriori* (model adjusted) RRs of incidence and death by PLC for both sexes.
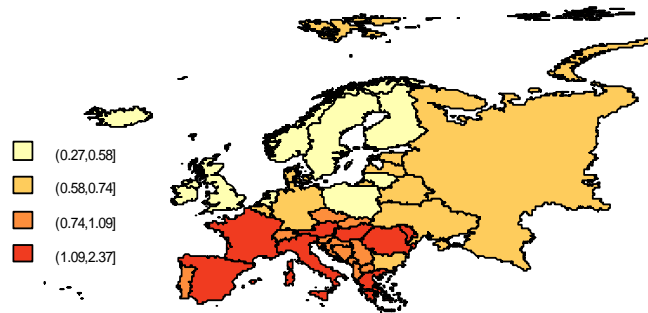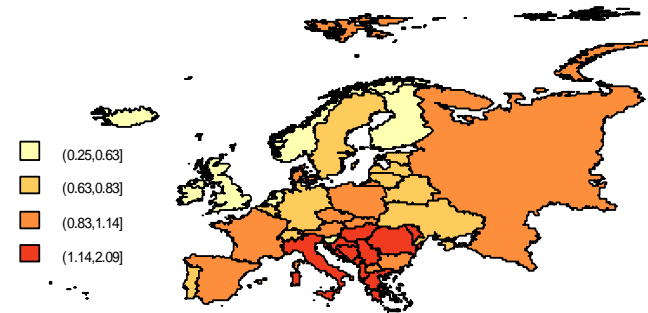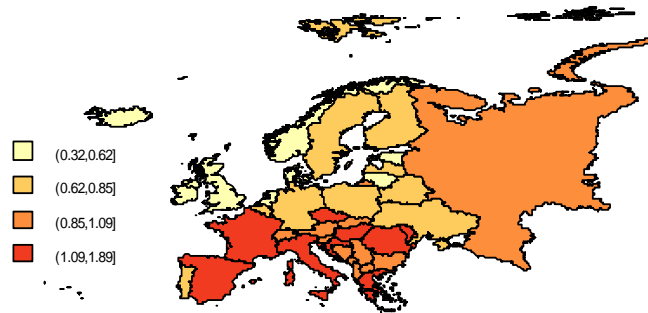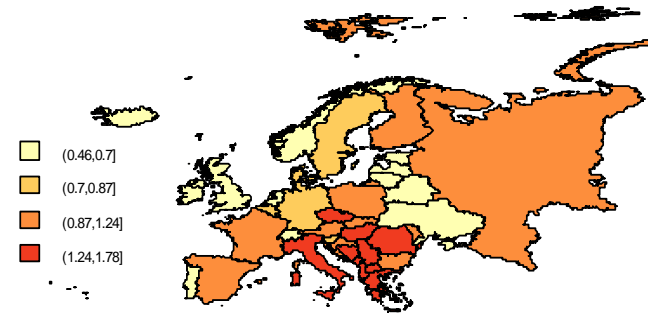
Figure 4.5. Maps of the *a posteriori* probabilities of RR>1 of incidence and death by PLC for both sexes.

(a) Incidence in Men: P(RR>1)
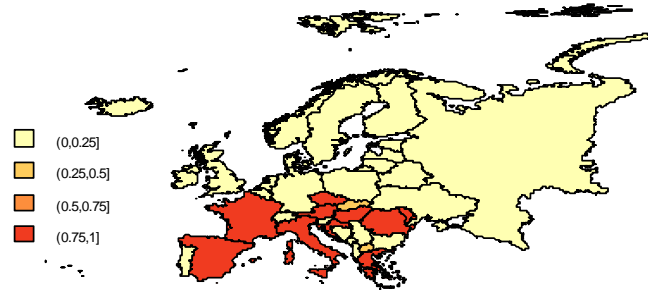
(b) Incidence in Women: P(RR>1)

(0,0.25]
(0.25,0.5]
(0.5,0.75]
(0.75,1]
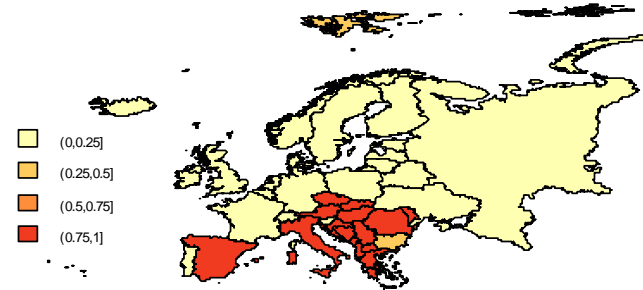
(c) Mortality in Men: P(RR>1)
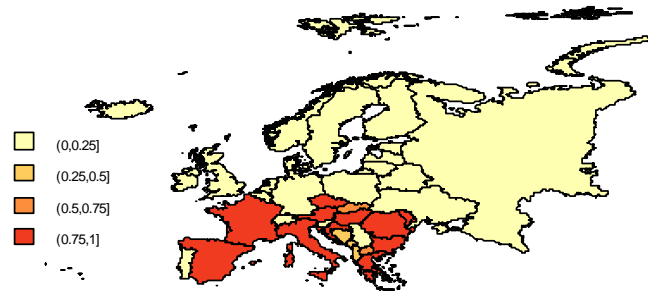
(d) Mortality in Women: P(RR>1)
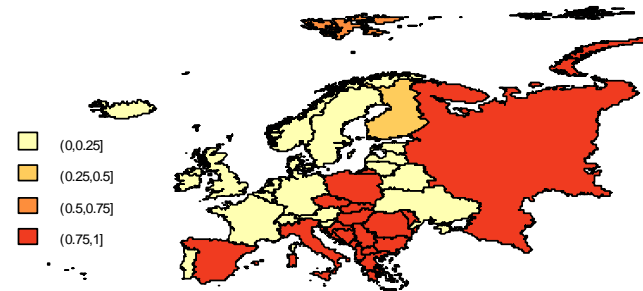
4.6. Discussion

4.6.1. Statistical methodology

This descriptive analysis has been related to a disease map reconstruction on the basis of an ecological analysis[161], because the aim has been to describe the relationship between geographical variation of PLC risk and explanatory covariates. We have performed an analysis of the spatial variation of PLC taking into account HCV and HBV exposures in a country-area level. The model considered to explain variability of PLC incidence and mortality rates included unstructured random effects in order to deal with extra-Poisson variability and to map the underlying relative risk surface[147].

The hierarchical Bayesian model proposed for this study has been based on a Poisson-lognormal model with unstructured random effects. Magnitude of these random effects is controlled through their precision parameter, usually modelled with a noninformative prior gamma distribution[147]. It has been tried different prior distributions for the prior precision of the random effects as well as an Uniform(0,100) prior distribution for their prior standard error. In the sensibility analysis, it has been observed that modelling the prior distribution of the standard deviation of those random effects, the SRF for the model parameters converged better than modelling the precision of random effects through different Gamma prior distributions (see Appendix A.3). In that sense, we have modelled the precision of these random effects by means of their prior standard deviation assuming a Uniform(0,100) prior distribution. It has been suggested that uniform prior distributions for standard deviation of parameters of hierarchical models are expected to perform well unless the number of levels of the variable is approximately below 5[162]. The use of prior gamma distributions have been based on previous works on spatial statistics[163] and those which initially refer to disease mapping and Bayesian methods[147,164,165]. In these works,

135

the key assumption was to select an equal prior emphasis on unstructured and structured heterogeneity (clustering).

In our model we also considered an "European area" covariate. Thus, countries included in the same area should have similar risks. With this approach we included the effect of "distance" between countries in terms of geographical area. Alternative ways to attempt with the notion of "distance" is to assume a structure for that effect, modelled by a multivariate distribution. An example of this last could be the Besag-York-Mollié[166] model, as a generalization of the Poisson-Normal-CAR (Conditional Autoregressive) model[167]. This model assumes that model parameter values in the neighbouring countries are similar. We believe that this assumption has more sense in the small-area analysis than in our country-area analysis due to greater similarity among districts or towns than among countries. In addition, this "European area" covariate had the role of random intercept in models M2-M6 with four levels. If this covariate included more than 4 levels, then it might be appropiate to model its prior variance through a simulation study.

In this analysis, it has been assumed that observed number of cases was Poisson distributed. In all countries the observed number of cases were greater than 50, and a normal distribution for that variable could also be assumed[134,168]. It has not been determined if this last issue could affect the final results, although the Poisson distribution for the number of cases seems to be the most reliable.

It has been observed the robustness of the SIR and SMR values by PLC due to those were calculated with large populations, although the fitted RRs have shown still large variability. However, one advantage of this country-area modelling of PLC risks was to determine the *a posteriori* probability of the high-risk and low-risk areas for PLC incidence and mortality. These probabilities showed also heterogeneity among RRs due to the number of cases included in each country. As an example, we have detected that Belgium and The Netherlands (Western European countries) have the

same posterior probability of RR of death by PLC (0.03) but different PLC RRs values (RR Belgium: 0.68; The Netherlands: 0.37).

Three dispersed chains have been considered in this study with a burnin of 10,000 samples for each model and gathering the next 50,000 samples. Autocorrelation of parameters in the samples was initially found in M3-M6 until at least lag 5. That was the reason why we chose to keep the sample values from every 10th iteration. We have observed that the inclusion of the area covariate in M2 provided a reduction in *DIC*, *pD* and *a posteriori* standard error of unstructured random effects, $((\tau_h)^{-\frac{1}{2}})$, compared with those of M1. That reduction was clearly observed when adding both HBV and HCV covariates.

A relevant limitation in our study was the non-availability of HBV and HCV prevalences stratified by age and sex for each country. It has long been understood that the standardization of the dependent variable in a regression framework without accounting for age, sex or some other set of confounders could be problematic. It has been demonstrated that state-level associations between mortality rates and several socioeconomic variables may be extremely sensitive to different age-adjustment methods[169].

Finally, we suggest two alternative analysis in order to account for extra Poisson variability: in the first one we could assume that observed cases follow a mixture of probability distributions[164,170] and in the second one, we could perform a joint analysis of incidence and mortality by PLC in both sexes by means of their joint modelling. The basis of the methodology for the first one has been developed previously in Chapter 3, whereas multivariate methodology for spatial data has been developed recently based on multivariate Poisson models[171,172].

4.6.2. Ecological and epidemiological analysis

In Europe, high risk areas for PLC are located in Southern and Eastern European countries and HCV prevalence could explain part of the observed variability on PLC risk. These results are in agreement with the previous published studies[2,7,25]. Italy and Greece showed the highest incidence and mortality from PLC with also high HBV and HCV prevalences.

The relation between HCV exposure and the significant increase on PLC incidence in Italy over de last decades has been related with the large cohort of subjects infected with HCV via the iatrogenic route before the 1960s[173,174]. In Greece, HBV represents the major aetiological factor of PLC[175]. However, in this country it has been reported a decrease of HBV in the role of PLC aetiology and a significant increase of HCV[176]. This issue has been evaluated in a simulation study which pointed out the expected increase in HCV-related morbidity in the next 20-30 years[177,178]. In Spain, where it has not been reported at present any increase of incidence and mortality rates from PLC, HCV infection and alcohol have been suggested as the main aetiological agents for PLC[179,180]. The predominant role of HCV and alcohol consumption in the PLC risk has been also described in several countries, as in France[42,181], Austria[182], Germany[183] and Sweden[184].

In Europe, the lowest RRs for PLC were found in The Netherlands, a Western European country. Although this country has low HBV and HCV prevalences, there has been detected an increase of PLC mortality during the last decade[185].

In our study, some former communist European countries showed low risk of PLC, whereas HBV and HCV prevalences were the highest among those European countries included in our study. However in these countries, high PLC risk were confirmed in Romania, Croatia, Czech Republic, Hungary, Moldova, Slovakia, Poland, Bulgaria and Hungary. In some of them it has been reported the lack of safety in

transfusion therapy[186-189] which entails a risk of transmission of hepatitis. Some studies have also detected that HBV and HCV co-infection are relatively frequent in some Eastern countries[190,191], respect to the coinfection prevalence observed in the rest of European countries. In Albania, where prevalence of HBV is high (more than 7%)[158], there was found those RRs greater than 1 for both incidence and mortality in women. In this line, high RRs of PLC in women have been also observed in Bosnia and Herzegovina, Macedonia and Serbia and Montenegro, but in this countries data was based on estimates.

We have observed that countries with HCV prevalence greater than 2% showed an IMRR of PLC incidence and mortality in men (IMRR incidence: 1.78, 95% CRI: 1.15 – 2.73; IMRR mortality: 1.48, 95% CRI: 1.14 – 1.93) and in women (IMRR incidence: 1.36, 95% CRI: 1.09 – 2.25; IMRR mortality: 1.28, 95% CRI: 1.05 – 1.75). These results are in concordance that HCV infection will play the most important role nowadays and in the years to come[7]. This is sustained by the decreasing role of HBV, specially after the universal HBV vaccination programs initiated for most countries[10], with the exception of some countries of Eastern Europe.

We should note substantial limitations in this study: i) the association between PLC and HBV and HCV exposures are at a country level due to data for individual level (age-sex distribution) was not available (possible ecological bias), ii) HVB and HCV prevalences and PLC incidence and mortality rates were estimations for some countries, mainly in Eastern Europe. These estimations have been obtained on the basis of the incidence and mortality rates of the neighbour countries, iii) a future analysis should include the effect of alcohol consumption, another risk factor for PLC which has not been taken into account in our analysis. A recent analysis, performed in the context of the World Health Organization's Global Burden of Disease 2000 project, ascribes 32% and 25% of worldwide cirrhosis and hepatocellular carcinoma cases, respectively, to alcohol[192]. All these limitations should be taken into account in

the interpretation of results that only could suggest a possible pattern of the PLC burden in Europe.

In conclusion, this disease mapping study could depict the pattern of PLC distribution in Europe at the beginning of 21st century where the high risk areas are Southern and Eastern Europe. There may be an underestimation of PLC incidence and mortality rates in some Eastern European countries due to the low PLC RRs detected in contrast with their high HBV and HCV seroprevalences observed. It is necessary the implementation of population-based cancer registries in these countries, as well as to carry out studies to determine viral infections prevalence across Europe by sex and age-groups in order to predict the future impact of this infection on the temporal trend of PLC.

Chapter 5

5. Time trends in liver disease in Spain during the period 1983-97

5.1 Background

In the period 1978-92, PLC incidence rates have shown a decline in some high-risk populations of Asia, such as Chinese populations of Singapore and Shanghai, as well as India[26]. However, PLC incidence has been increasing in some developed countries, as in Australia, U.K., U.S., Canada, Japan, France and Italy[26,42,47,49,78,193,194]. An increase between 83% and 108% in PLC incidence has been reported in these last two European countries[26,123,195]. The increase on PLC incidence rates in Southern European countries, which have a mean HCV prevalence greater than the European mean[15], could be explained by HCV infection[26]. However, there has not been detected an increase on incidence and mortality of PLC rates in the same period in Spain, also a Southern European country, although its HCV pattern (0.5% to 2.5%) is similar to that of Italy (HCV prevalence: 2.2%)[15,196-198].

PLC time trends were evaluated in Catalonia for the period 1980-97[50]. In that study, it was found an increasing trend in cholangiocarcinoma mortality rates for both sexes, whereas hepatocellular carcinoma incidence and mortality rates remained stable. On the other hand, a significant increase of liver cirrhosis mortality was detected in 25-35 year old males, while cirrhosis mortality rates fell for both sexes in the other age groups [50].

Coding recommendations of PLC and liver metastasis changed slightly from the 8th to the 9th revision of the ICD. These were implemented after 1965 and 1975, respectively[51,52], whereas the 10th revision was introduced in 1992[53]. The impact of coding practices in mortality statistics is relevant in PLC, although it has probably been overestimated in relation to that impact of improving diagnostic technology such as ultrasound and Computed Tomography, as well as on the screening practices among cirrhotic patients and chronic carriers of HBV or HCV infections. This temporal variations in the implementation of the diagnostic techniques and the

changes in the codification criteria of PLC makes difficult to interpret the time trends of PLC. In Spain, about 90% of the PLC cases have an underlying cirrhosis[199], being this an additional source of error due to the fact that some PLC deaths can be attributed to cirrhosis in mortality statistics[200]. Therefore, comparison of time trends among countries should be done cautiously[2].

The study presented in this chapter has been designed in order to estimate the time trends of mortality and incidence rates due to chronic liver disease in Spain during the period 1983-97. This study will allow us to determine: 1) incidence and mortality trends by liver tumors and 2) mortality trends of liver cirrhosis in a long period (15 years) using the same version of ICD (9th). Liver cancer data will be analyzed separately by: hepatocellular carcinoma, cholangiocarcinoma and liver cancer "unknown" if primary or metastatic.

5.2 Hypotheses

1) In Spain, based on the pattern of PLC incidence in its neighbour countries and in the prevalence of HCV, it should be expected an increase of the incidence of PLC during the last two decades.

2) There could be an increase of mortality by liver cirrhosis in men younger than 40 years as it was reported previously in Catalonia.

3) There is expected an increase of cholangiocarcinoma mortality for both sexes in Spain, as well as it has been observed in several industrialized countries.

5.3 Sources of data

Liver cancer incidence and population data were extracted from the Cancer Incidence in Five Continents Database, based on data collected by the IARC[201]. Data have been selected for three 5-year periods and for twelve 5-year age groups. Based on the ICD on its 9th revision, the codes selected for liver cancer were: Global Liver Cancer (155), hepatocellular carcinoma (1550), cholangiocarcinoma (1551) and liver cancer "unknown" if primary or metastatic (1552). Five Spanish cancer registries had data available during the period 1983-97: Tarragona, Granada, Murcia, Navarra and Zaragoza. These cover about 10% of the Spanish population. The annual average of person-years obtained for the last 5-year period (1993-97) was 3.9 million. Data were aggregated and a total of 4,030 incident liver cancer cases (2,596 men and 1,434 women) were diagnosed throughout1983-97.

PLC mortality data was extracted from the National Statistics Institute of Spain, which depends on the Spanish Government. Data for the whole Spanish population was considered. Codes selected from ICD-9 were identical to those selected for incidence, adding liver cirrhosis (ICD-9 code: 571). The annual average of person-years obtained for the last 5-year period (1993-97) was 40.1 million. A total of 61,647 death cases by liver cancer (36,913 men; 24,734 women) and 117,002 death cases by liver cirrhosis (80,608 men; 36,394 women) were observed during the period 1983-97.

For this analysis, cases were selected with ages comprised between 20-79 years, because liver disease is a rare disease for subjects younger than 20 years. Cases were grouped by 5-year age groups, so for each year there were available 12 age-groups.

5.4 Statistical methods

In this section we introduce the methodology used to estimate incidence and mortality trends of liver disease, by means of an age-period-cohort (APC) analysis based on the Bayesian approach. APC analysis allows us to summarize information related to disease rates with the aim of assessing the effect of these three factors on the rates.

Age is a variable which represents the effect related with the duration of the exposition to some risk factors of one disease[202-204]. Period and birth cohort effects seek to explain changes in the rates associated with time[202-204]. Period effect represents change in the rates associated with all age groups simultaneously[202-204], reflecting improvements in treatments or changes in the registration procedure of the disease. However, cohort effect is associated with a change in rates in successive age groups in successive time periods[202,203]. Cohort effects are associated with habits or exposures related with different generations[203,204], such as HCV infection as an example for PLC[2]. These exposures took place in a certain moment of the life of individuals.

The classical approach to descriptive analysis of registry data has been to tabulate incidence (or mortality) data in rectangular (usually 5x5 year) subsets of the Lexis diagram[203]. Recently, Holford has been developed a new approach to perform these analysis combining unequal intervals[205]. In our study, time intervals have been fixed to 5 years. Section A.4 of the Appendix includes the WinBUGS code for the models and methods described below.

5.4.1. Age, period and cohort tabulation

Cancer cases are tabulated by age and date of diagnosis (period) in 5-year intervals. Data was arranged in three periods (P) (1983-87, 1988-92, 1993-97), and twelve age (A) groups (from 20-24 to 75-79). From this data, birth cohorts can be calculated as

$$Cohort = Period - Age \qquad (26).$$

Then we have C = A + P − 1 = 14 birth cohorts with central years 1908, 1913, 1918,..., 1973. The linear relationship (26) leads to the problem of overlapping cohorts, as have previously discussed Clayton and Schifflers[206,207] and Holford[204,208]. Counts of cases and person-years are indexed from i=1,…,A·P, so an individual cohort, c, can be followed depending on period, p, and the number of age groups, A, through c=A+p-1, as *Table 5.1* shows.

Table 5.1. Indexing of cases and person-years, indexed by i, according to age and period for observed data with A=12, P=3 and C=14.

	20-24 (a=1)	25-29 (a=2)	30-34 (a=3)	…	70-74 (a=11)	75-79 (a=12)
1983-87 (p=1)	i=1, c=12	i=4, c=11	i=7, c=10	...	I=31, c=2	i=34, c=1
1988-92 (p=2)	i=2, c=13	i=5, c=12	i=8, c=11	...	I=32, c=3	i=35, c=2
1993-97 (p=3)	i=3, c=14	i=6, c=13	i=9, c=12	...	I=33, c=4	i=36, c=3

Rates can be modelled as function of age, period and cohort, so the dataset should include the variables O_{ap} and Y_{ap}, O beign the observed number of cases, whereas Y is the person-years (population) at risk for the age group a, period p, and, by equation (26), the implicit c cohort. Let O_{ap} be the observed response, $\ln(Y_{ap})$ be the offset, and let the age, period, and cohort be the categorical explanatory variables in a Poisson model. An estimate for the crude rate λ in a certain age-period-cohort combination would be

$$\hat{\lambda}(a, p) = \frac{O_{ap}}{Y_{ap}} \qquad (27).$$

These rates can be modelled with several combinations of the explanatory variables: age effect (AE) model, age-period (AP) model, age-cohort (AC) model and the full APC model. However, a lack of identifiability among parameters arises when APC is the model selected due to the linear relationship between age, period and cohort effects (see equation 26)[204,206,207,209-211]. The use of a reduced age-period (AP) or age-cohort (AC) model whenever possible has been advised because effects are identifiable in these two models[206,207]. The APC model is required only when neither of these models provides a satisfactory fit[206,207]. Details of such models are described below.

5.4.2. Age-period model

The age-period model states that the age-specific rates have the same shape in all periods. This model has one parameter per age class and one per period. Let $\lambda(a, p)$ be the incidence (or mortality) rate for the age group a in the period p. The model for the log-rate is defined as

$$\log(\lambda(a, p)) = \alpha_a + \beta_p \qquad (28),$$

being the α_a the logs of age-specific rates, and the β_p the effects associated with periods. The natural constraint is to fix one period parameter to be 0, $\beta_{p_0} = 0$, so $\log(\lambda(a, p_0)) = \alpha_a$ for the period p_0. Comparing rates in any age class between period p and period p_0 gives $\log(\lambda(a, p) / \lambda(a, p_0)) = \beta_p$, so the β_p are log rate-ratios relative to the reference period p_0. The age-specific rates are cross-sectional rates which refers to the period p_0, reflecting what we expect to see in a population during a short period of time. In addition, period parameters describe how these rates change as a function of time.

5.4.3. Age-cohort model

The age-cohort model is similar to the age-period model. In this model the age-specific rates have the same shape for any cohort. Following the AP model, let $\lambda(a,c)$ be the observed incidence (or mortality) rate for the age group a in the cohort c, and then the model for the log rate is

$$\log(\lambda(a,c)) = \alpha_a + v_c \qquad (29),$$

being the α_a the logs of age-specific rates, and v_c the effects associated with cohorts. If we set one cohort parameter to be 0, $v_{c_0} = 0$, then $\log(\lambda(a,c_0)) = \alpha_a$ for the cohort c_0. As we have seen for periods, comparing rates in any age class between cohort c and cohort c_0 give $\log(\lambda(a,c)/\lambda(a,c_0)) = v_c$, so the v_c are log rate-ratios relative to the reference cohort c_0. It should be noted that the estimates relating to the youngest and oldest cohorts are less precise because those are based only on a few cases.

5.4.4. Age-drift model

If we replace the period parameters in the AP model (28) by a linear trend in log-rates then the log rate-ratio should show a straight line. The Age-drift (AD) model is

$$\log(\lambda(a,p)) = \alpha_a + \beta(p - p_0) \qquad (30).$$

If we assume that $\beta_p = \beta(p - p_0)$, then we can also perform the same change in the AC model (29) for the cohort parameters and then we obtain $v_c = v(c - c_0)$. Point estimations and standard errors of v and β are identical from the likelihood point of view[206]. But there is only one age-drift model, and its interpretation of this is that rates increase exponentially by time (calendar time or cohort) at the factor $e^\beta = e^v$,

which is the annual percent change (APCH) of rates[206,207] for all age classes. Thus the AD model is a model which belongs to the set of both AP and AC models.

5.4.5. Age-period-cohort model

If we add the cohort variable as a factor into an AP model, with 1+(A-1)+(P-1) parameters, we get only C-2 and not C-1 new parameters due to equation (26) and to the common "drift" term of both AP and AC models.

There are different ways to obtain several sets of estimates for those effects. Suppose first that we have fit the following model

$$\log(\lambda(a, p)) = \alpha_a + \beta_p + \nu_c \qquad (31),$$

second, let us define three constant values such as

$$\theta_a + \theta_p + \theta_c = 0 \Leftrightarrow \theta_a = -\theta_p - \theta_c \qquad (32),$$

and third, let us remember the relationship among parameters, that is

$$c = p - a \Leftrightarrow a - p + c = 0 \qquad (33).$$

Therefore, we can add (32) and (33) to (31) and replace the last by

$$\log(\lambda(a, p)) = (\alpha_a + \delta \cdot a - \theta_p - \theta_c) + (\beta_p - \delta \cdot p + \theta_p) + (\nu_c + \delta \cdot c + \theta_c) \quad (34),$$

where δ is a constant term. This parameterization, which sets certain period and cohort effects to 0, involves to choose values for the parameters θ_p, θ_c and δ.

For that reason, we can define infinity of models related with the lack of identifiability of the APC model. In this last model we can only identify mathematically the second order differences associated with each of the effects due to the constants θ_p, θ_c and δ (linear trend) will cancel when we form the second order differences[204,206,207,209]. As example, the second order differences associated with the age effects are $(\alpha_i - \alpha_{i+1}) - (\alpha_{i+1} - \alpha_{i+2}) = \alpha_i - 2\alpha_{i+1} + \alpha_{i+2}$, so they are one difference between two first order differences of parameters. These second-order differences are the curvatures of the parameter estimates, so deviations from linearity are estimable[204,206,207,209-211].

Holford[202] showed that the deviations for age are the residuals from a linear regression of the estimated age effects, and similarly for the period and cohort deviations and curvatures. Holford suggested to first fit a model with any parameterization of the effects, and then regress the age estimates on age, the period estimates on period and cohort estimates on cohort. These deviations are not independent of each other but the curvatures are local curvatures not influenced by the curvatures at intervals far apart, although curvatures for adjacent intervals are correlated as they have estimated effects in common. So the curvatures and deviations are linearly related. Following this approach and performing all these regression analyses, the APC model is defined as

$$\log(\lambda(a, p)) = (\tilde{\alpha}_a + \hat{\mu}_a + \hat{\delta}_a \cdot a) + (\tilde{\beta}_p + \hat{\mu}_p + \hat{\delta}_p \cdot p) + (\tilde{\nu}_c + \hat{\mu}_c + \hat{\delta}_c \cdot c) \ (35),$$

where $\left(\tilde{\alpha}, \tilde{\beta}, \tilde{\nu}\right)$ are the residuals and $\left(\hat{\mu}, \hat{\delta}\right)$ are the regression parameters. The residuals are on average 0 across age, period and cohort, respectively. The Holford's approach and suggestion for parameterization is based on a biological point of view. It assumes that age is the major time scale, period is the secondary and cohort is the residual. The last represents the curvature effect among birth cohorts. This means that

cohort effects should be 0 on average, period effect should be relative risk relative to some reference period and age effect should represent age-specific rates in a reference period after correction of cohort effects that are 0 on average.

Based on the previous approach and using (35), we can define

$$g(c) = \tilde{v}_c = v_c - (\hat{\mu}_c + \hat{\delta}_c \cdot c) \qquad (36),$$

where $g(c)$ is a function of the cohort effect. The time trends should then be absorbed by a function of the period effect, $h(p)$, as

$$h(p) = \beta_p - \beta_{p_0} + \hat{\delta}_c \cdot (p - p_0) \qquad (37),$$

whereas the remaining function of the age effect, $f(a)$, is

$$f(a) = \alpha_a + \hat{\mu}_c + \hat{\delta}_c \cdot a + \hat{\delta}_c \cdot p_0 + \beta_{p_0} \qquad (38).$$

In this parameterization we can verify that $f(a) + h(p) + g(c) = \alpha_a + \beta_p + v_c$, and this is a way of partitioning the effects in a well defined form between the three factors. The role of period could be equally interchanged in this parameterization, such that we can assume, for example, that period is the residual time scale. This way to overcome the non-identifiability problem is based on one approach to estimable functions[204,206,207,209-211].

There has been proposed other ways to overcoming non-identifiability problem. Among others, some of them involves imposing constraints on the parameters[209,212] whereas some another use individual records of cases[213]. In this work we do not deal

with those, because the interest has been to test whether deviation from linearity is clinically or epidemiologically relevant.

5.4.6. Bayesian approach to age-period-cohort models

In order to smooth effects on each scale on time, Gaussian autoregressive prior models in the forward direction were proposed by Breslow and Clayton[214] and later by Berzuini and Clayton[215] and Bray[216-218]. In these models it was assumed that second differences are independent normal covariates. Trends corresponding to age, period and birth cohort were smoothed using second degree autoregressive smoothing (non-parametric smoothing with autoregressive error component). For age, period and cohort these resulted in linear extrapolations.

Let us suppose $O_{ap} \sim Poisson(\lambda(a,p)Y_{ap})$, where $\lambda(a,p)$ can be estimated with $\log(\lambda(a,p)) = \log(Y_{ap}) + M(a,p,c)$, where $M(a,p,c)$ varies according to the complexity of the model considered: A, AD, AP, AC, APC. Each one of the effects has been modelled with an autoregressive structure as follows:

$$
\begin{aligned}
&\gamma_1 \sim N(0, 10^{-6} \tau_\gamma) \\
&\gamma_2 \mid \gamma_1 \sim N(0, 10^{-6} \tau_\gamma) \\
&\gamma_r \mid \gamma_1, \gamma_2, ..., \gamma_{r-1} \sim N(2\gamma_{r-1} - \gamma_{r-2}, \tau_\gamma) \qquad (39), \\
&\tau_\gamma \sim Gamma(0.01, 0.001) \\
&\sigma_\gamma = \frac{1}{\tau_\gamma}
\end{aligned}
$$

where γ could be α or β or v, r the number of levels of that effect and τ_γ is the precision parameter (inverse of the variance parameter σ_γ^2) with non-informative prior distribution[217]. The drift parameter of the model AD assumes also a normal prior distribution with mean 0 and precision τ_γ, for which a non-informative prior

distribution was assumed, as in (39). In these models we imposed "corner" constraints for period parameters, $\beta_{1983-87} = 0$, and cohort parameters, $\nu_{1933} = 0$, in order to have a reference category for each parameter. Note that successive first order differences between parameters are assumed to be equal in (39). As example, if we consider the first three age parameters, $\{\alpha_1, \alpha_2, \alpha_3\}$, the equality assumption on subsequent first order differences means $(\alpha_3 - \alpha_2) = (\alpha_2 - \alpha_1) \Leftrightarrow \alpha_3 = 2\alpha_2 - \alpha_1 \Leftrightarrow \alpha_2 = \frac{\alpha_3 + \alpha_1}{2}$.

Bashir and Estève[219] modelled the full age-period-cohort model based on (39) but adding a constraint on the second order differences related only with the age parameters. For this constraint it was assumed that one second order difference is estimated as the mean on the previous and subsequent second order differences. So the age effect in this APC parameterization is defined as

$$
\begin{aligned}
\alpha_i \mid \alpha_j, j \neq i &\sim N(\mu_{\alpha_i}, 10^{-6}\tau_\alpha) \\
\mu_{\alpha_1} &= 2\alpha_2 - \alpha_3 \\
\mu_{\alpha_2} &= \frac{2\alpha_1 + 4\alpha_3 - \alpha_4}{5} \\
\mu_{\alpha_i} &= \frac{4\alpha_{i-1} + 4\alpha_{i+1} - \alpha_{i-2} - \alpha_{i+2}}{6}, \quad 3 < i < A - 2 \\
\mu_{\alpha_{A-1}} &= \frac{2\alpha_A + 4\alpha_{A-2} - \alpha_{A-3}}{5} \\
\mu_{\alpha_A} &= 2\alpha_{A-1} - \alpha_{A-2} \\
\tau_\alpha &\sim Gamma(0.01, 0.001) \\
\sigma_\alpha^2 &= \frac{1}{\tau_\alpha}
\end{aligned}
\tag{40}.
$$

This mean that for any given age point (3<i<A-2) the conditional expectation of the age effect is obtained through cubic interpolation from the two points on the other side[219].

We will denote as APC2 model the model proposed by Bashir and Estève, whereas the previous APC parameterization initially proposed and applied by Clayton, Breslow, Berzuini and Bray will be denoted as APC1 model.

One of the applications of both models was projections rather than estimation of the effects. In that models, the use of autoregressive smoothing eliminates the random fluctuation that usually occurs with the uncorrected maximum likelihood estimate of the parameters[219]. It should be noted that APC logistic models have also been proposed[220,221], but we focused the analysis on those based on Poisson regression models.

5.4.7. Procedure of analysis and evaluation of the deviation from linearity in the age, period and cohort effects

The analysis of this part of study has been performed through a graphical representation of trends and a model selection procedure.

5.4.7.1. Graphical representation of trends

There has been performed two graphical representations: a) the age-adjusted rates to the world standard population[134] per 100,000 person-years by calendar year and b) crude rates versus birth cohorts and 5-year periods of diagnosis (incidence) or death (mortality). By means of the first plot it could be detected the increase/decrease of rates throughout the period. We can evaluate the statistical significance of these time trends by means of a log-linear model fitted to the age-adjusted incidence/mortality rates. Let η_i be the age-adjusted mortality/incidence rate for the ith year and let us assume Poisson distribution for those rates. Therefore, we can fit a linear model as

$$\eta_i \sim Poisson(\psi_i)$$
$$\log(\psi_i) = \omega + \rho \cdot i$$
$$\omega \sim N(0,0.01)$$
$$\rho \sim N(0,0.01)$$
$$0 \le i \le 14$$
$$\kappa = (e^\rho - 1) \cdot 100$$

(41),

where ψ_i is the mean of the rates, ω and ρ the intercept and the slope of the log-linear model, respectively, and finally κ the APCH, throughout the 15-year period 1983-97. Note that in (41) the reference year was 1983 (i=0) and APCH could be also calculated with the AD model.

In this first step, the second graphical representation was to represent birth cohorts and periods versus crude rates. That was a way to see whether the major variation in these rates were by cohort or period, and, if this was true, curves were parallel between age groups.

5.4.7.2. Choice of model for the APC analysis

We determined which was the best fitting model, based on the criteria of *DIC* and *pD*[160] that we described in Chapter 4. In this case we started with the AE model, followed by the AD, AP and AC. We also estimated the APC1 and APC2 models in order to evaluate whether the full APC model was required. If the APC model was the chosen one to explain variability of rates, then we determined if there was deviation from linearity in cohort effects by means of a model based on the residual approach proposed by Holford, which we have described in (5.4.5)-(5.4.6). For deviations from linearity of the cohort effect, we have:

$$v_i \sim N(\mu_{v_i}, \tau_v)$$

$$\mu_{v_i} = \mu_c + \delta_c \cdot c_i$$

$$\tilde{v}_i = (v_i - \mu_{v_i}) - v_6$$

$$\mu_c \sim N(0, 0.001) \qquad\qquad (42),$$

$$\delta_c \sim N(0, 0.001)$$

$$1 \leq c_i \leq 14$$

$$\tau_v \sim Gamma(0.01, 0.01)$$

being v_i the point estimates of the 14 cohort effects, and τ_v are their prior precision. Note that residuals, $\tilde{v}_i$, have been estimated using the reference cohort of 1933 (v_6).

Following Holford method, for the period effects we have:

$$\beta_i \sim N(\mu_{\beta_i}, \tau_\beta)$$

$$\tilde{\beta}_i = \mu_{\beta_i} - \mu_{\beta_{p0}} + \delta_c \cdot (p_i - p_0)$$

$$\mu_{\beta_i} \sim N(0, 0.1) \qquad\qquad (43),$$

$$\tau_\beta \sim Gamma(0.01, 0.01)$$

$$1 \leq i \leq 3$$

$$p_0 = 1985$$

where β_i are the point estimates of the 3 period effects and τ_β their prior precision. Note that the central years for the three periods are $p = \{1985, 1990, 1995\}$ and the reference period is 1983-87.

For the age effect we have

$$\alpha_i \sim N(\mu_{\alpha_i}, \tau_\alpha)$$

$$\tilde{\alpha}_i = \mu_{\alpha_i} + \mu_c + \delta_c \cdot a_i + \delta_c \cdot p_0 + \beta_{p_0}$$

$$\mu_{\alpha_i} \sim N(0, 0.1) \qquad\qquad (44),$$

$$\tau_\alpha \sim Gamma(0.01, 0.01)$$

$$1 \leq a_i \leq 12$$

being α_i the point estimates of the 12 age effects and τ_α their prior precision. Note that through (42) to (44) it has been used $Gamma(0.01, 0.01)$ as prior distributions for the precisions of the parameters (see Sensibility Analysis on Appendix A.5.)

As we did for the geographical analysis of Chapter 4, for each model it has been run 3 chains with 60,000 simulations and 10,000 burning samples, storing the samples from every 10th iteration. Statistical significance for the APCH has been established on the basis of the 95% credibility interval of this parameter. If that interval does not include the value 0, we have then assumed that the increase/decrease observed is significant.

5.5. Results

5.5.1. Regular trends: age-adjusted rates by calendar year and period

Figure 5.1 and *Table 5.2* shows mortality and incidence age-adjusted rates for liver disease, by sex, throughout the period 1983-97 in Spain.

Figure 5.1. Trends in age-adjusted mortality and incidence rates of liver disease in Spain during the period 1983-97.
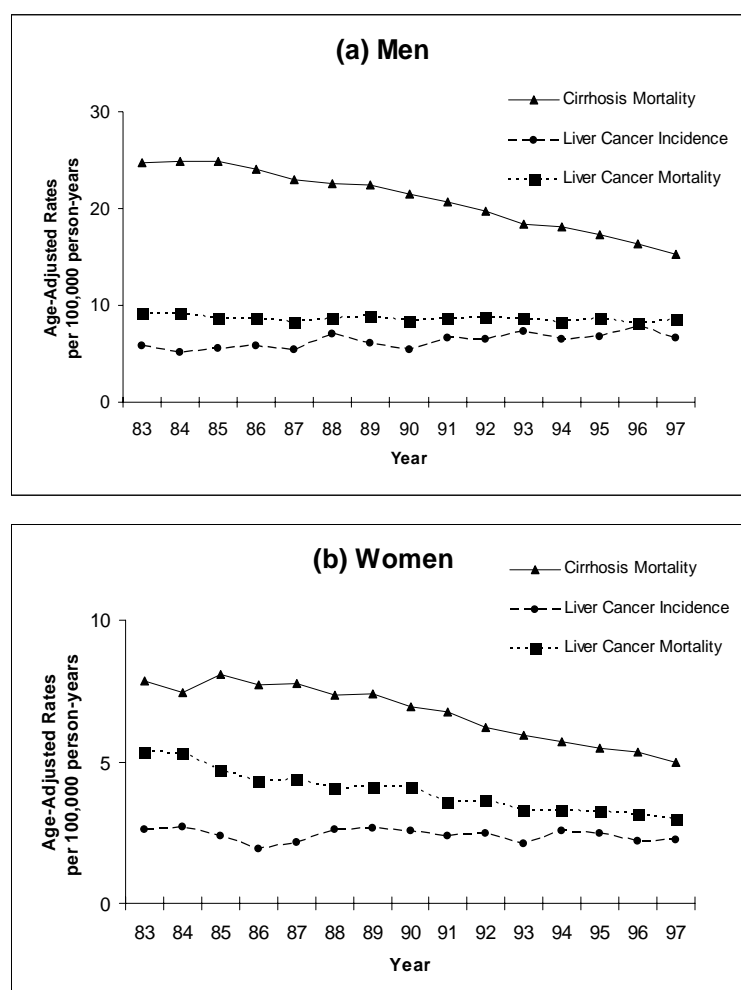
Table 5.2. Annual Percent Change in Mortality and Incidence rates from liver disease in Spain during the period 1983-97.

Men	Period		Period 83-97
	83-84	96-97	APCH (95% CRI)
Liver Cirrhosis Mortality			
N	11,822	8,661	80,608
AAMR	27.3	14.9	-3.1 (-5.1, -1.9)
Liver Cancer Mortality			
N	4,547	5,438	36,913
AAMR	9.6	8.4	-0.5 (-0.9, -0.1)
Liver Cancer Incidence			
N	263	445	2,596
AAIR	5.8	7.1	2.1 (1.3, 7.2)
Women			
Liver Cirrhosis Mortality			
N	4,902	4,008	36,394
AAMR	8.8	4.8	-2.9 (-6.2, -1.3)
Liver Cancer Mortality			
N	3,757	3,071	24,734
AAMR	6.4	2.9	-4.2 (-7.3, -1.9)
Liver Cancer Incidence			
N	188	187	1,434
AAIR	2.7	2.2	-0.5 (-1.3, 0.4)

APCH: Annual Percent Change; **95% CRI**: 95% credibility interval; **AAMR:** Age-Adjusted Mortality Rates per 100,000 person-years; **AAIR:** Age-Adjusted Incidence Rates per 100,000 person-years; **N:** Number of cases.

Table 5.2. shows that the significant fell in Liver Cirrhosis mortality rates in men was similar (APCH:-3.1%) than the observed in women (APCH=-2.9), whereas liver Cancer mortality rates in men decreased slightly (APCH=-0.5%) compared with those of women (APCH=-4.2%). A significant rise in liver cancer incidence was detected in men (APCH=2.1%; 95% CRI: 1.3% to 7.2%), whereas incidence rates in women remained stable (APCH=-0.5, 95% CRI: -1.3% to 0.4%).

Figure 5.2. Trends in age-adjusted incidence rates of Hepatocellular Carcinoma, Cholangiocarcinoma and Liver Cancer Unspecified in Spain during the period 1983-97.
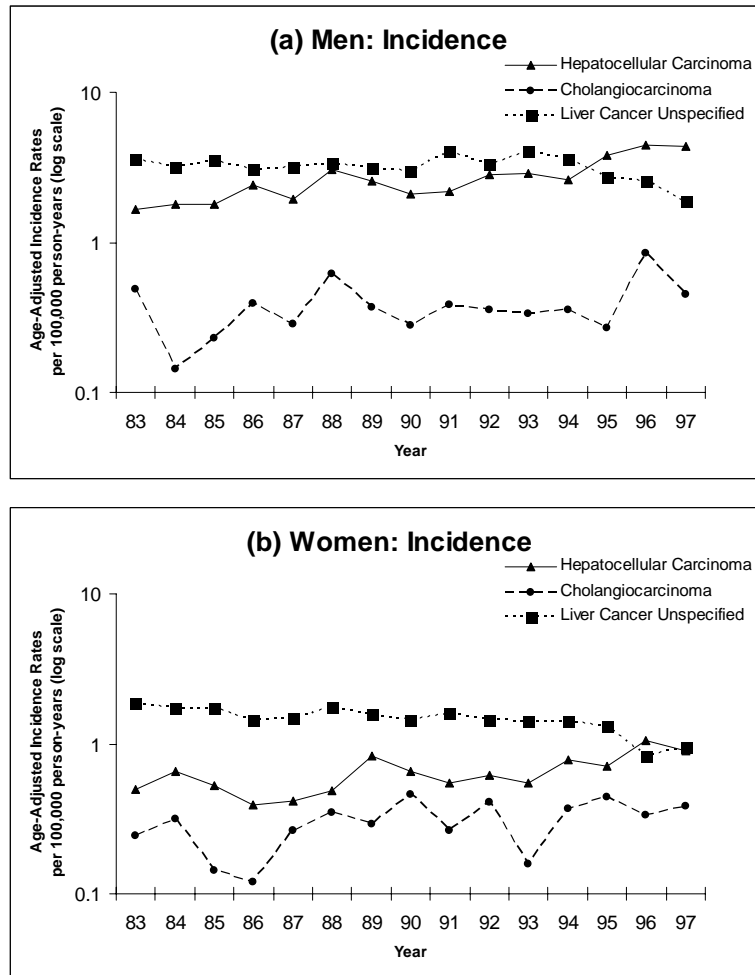
Trends in incidence rates of liver cancer according to histological types are represented in *Figure 5.2* and in *Table 5.3*. In Spain, a significant increase on hepatocellular carcinoma incidence was detected throughout the period of study for both sexes (APCH men=6.6%, 95% CRI: 5.3, 9.2; APCH women=4.5%, 95% CRI: 1.4, 7.3). Although it has been estimated a similar increase in cholangiocarcinoma incidence for both sexes, those were not found to be statistically significant (APCH men=3.3%, 95% CRI: -1.3, 8.3; APCH women=3.7%, 95% CRI: -1.2, 6.1).

Table 5.3. Annual Percent Change in incidence rates from liver cancer according histological types in Spain throughout the period 1983-97.

Men	Period		Period 83-97
	83-84	96-97	APCH (95% CRI)
Hepatocellular Carcinoma			
(N)	90	260	1,088
AAIR	1.7	4.4	6.6 (5.3, 9.2)
Cholangiocarcinoma			
N	15	35	154
AAIR	0.3	0.6	3.3 (-1.3, 8.3)
Liver Cancer Unspecified			
N	158	150	1,354
AAIR	3.4	2.2	-1.2 (-3.4, 2.3)
Women			
Hepatocellular Carcinoma			
N	32	69	321
AAIR	0.5	1.0	4.5 (1.4, 7.3)
Cholangiocarcinoma			
N	19	30	163
AAIR	0.3	0.7	3.7 (-1.2, 6.1)
Liver Cancer Unspecified			
N	137	88	950
AAIR	1.8	0.9	-3.5 (-5.2, -2.1)

APCH: Annual Percent Change; **95% CRI**: 95% credibility interval; **AAIR:** Age-Adjusted Incidence Rates per 100,000 person-years; **N:** Number of cases.

Incidence rates of liver cancer unspecified if primary or metastatic for men also remained stable (APCH=-1.2, 95% CRI: -3.4, 1.2), whereas a significant decrease was detected among women (APCH=-3.5%, 95% CRI: -5.2, -2.1).

Figure 5.3. Trends in age-adjusted mortality rates of Hepatocellular Carcinoma, Cholangiocarcinoma and Liver Cancer Unspecified in Spain during the period 1983-97.
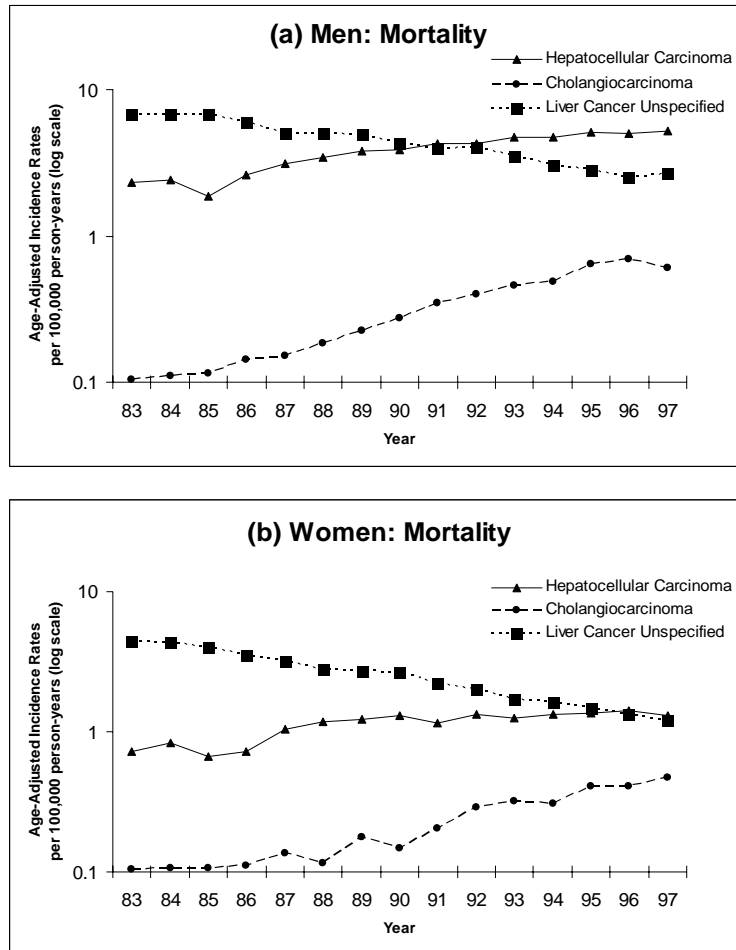
Figure 5.3 and *Table 5.4* show mortality rates according to liver cancer histology. Graphical pattern of these mortality rates is almost identical to that observed for incidence rates in both men and women.

Table 5.4. Annual Percent Change in mortality rates from liver cancer according to histological types in Spain during the period 1983-97.

Men	Period		Period 83-97
	83-84	96-97	APCH (95% CRI)
Hepatocellular Carcinoma			
N	1,158	3,257	16,198
AAMR	2.3	5.2	6.8 (5.8, 8.1)
Cholangiocarcinoma			
N	45	433	1,486
AAMR	0.1	0.7	17.1 (13.5, 21.2)
Liver Cancer Unspecified			
N	3,344	1,748	19,229
AAMR	6.7	2.5	-7.2 (-8.2, -6.2)
Women			
Hepatocellular Carcinoma			
N	552	1,316	6,904
AAMR	0.7	1.3	5.1 (3.5, 6.3)
Cholangiocarcinoma			
N	44	461	1,524
AAMR	0.1	0.7	15.0 (11.5, 19.5)
Liver Cancer Unspecified			
N	3,161	1,294	16,306
AAMR	4.4	1.2	-8.6 (-9.2, -8.1)

APCH: Annual Percent Change; **95% CRI**: 95% credibility interval; **AAMR:** Age-Adjusted Mortality Rates per 100,000 person-years; **N:** Number of cases.

Table 5.4 showed the significant rise in mortality rates for hepatocellular carcinoma (APCH men=6.8%, 95% CRI: 5.8, 8.1; APCH women=5.1%, 95% CRI: 3.5, 6.3) jointly with a significant decrease for mortality coded as liver cancer unspecified (APCH men=-7.2%, 95% CRI: -8.2, -6.2; APCH women=-8.6, 95% CRI: -9.2, -8.1). Note that hepatocellular carcinoma mortality rates increased in a similar APCH as liver cancer unspecified mortality rates decreased. In addition, death rates for cholangiocarcinoma showed a significant rise for both sexes (APCH men=17.1%, 95% CRI: 13.5, 21.2; APCH women=15.0%, 95% CRI: 11.5, 19.5).

5.5.2. Trends by age-cohort, age-period and age-period-cohort

5.5.2.1. Liver Cirrhosis mortality

Figure 5.4 shows crude mortality rates by age groups versus birth cohorts and 5-year periods for liver cirrhosis mortality in men and women. In both sexes, there was observed a decreasing trend in cirrhosis mortality rates for all birth cohorts with the exception of those birth cohorts of men comprised between 1950-1965 (*Figure 5.4-a* and *5.4-b*), for which cirrhosis mortality remained stable (APCH=-6.0, 95% CRI:-21.5 to 13.0). Those cohorts coincide with the age groups comprised between 25-34 years old.

Figure 5.4. Liver Cirrhosis: Crude mortality rates per 10^5 person-years by birth cohort and period of death according to age group.

(a) Cirrhosis Mortality in Men: Age-Cohort

(b) Cirrhosis Mortality in Men: Age-Period

(c) Cirrhosis Mortality in Women: Age-Cohort
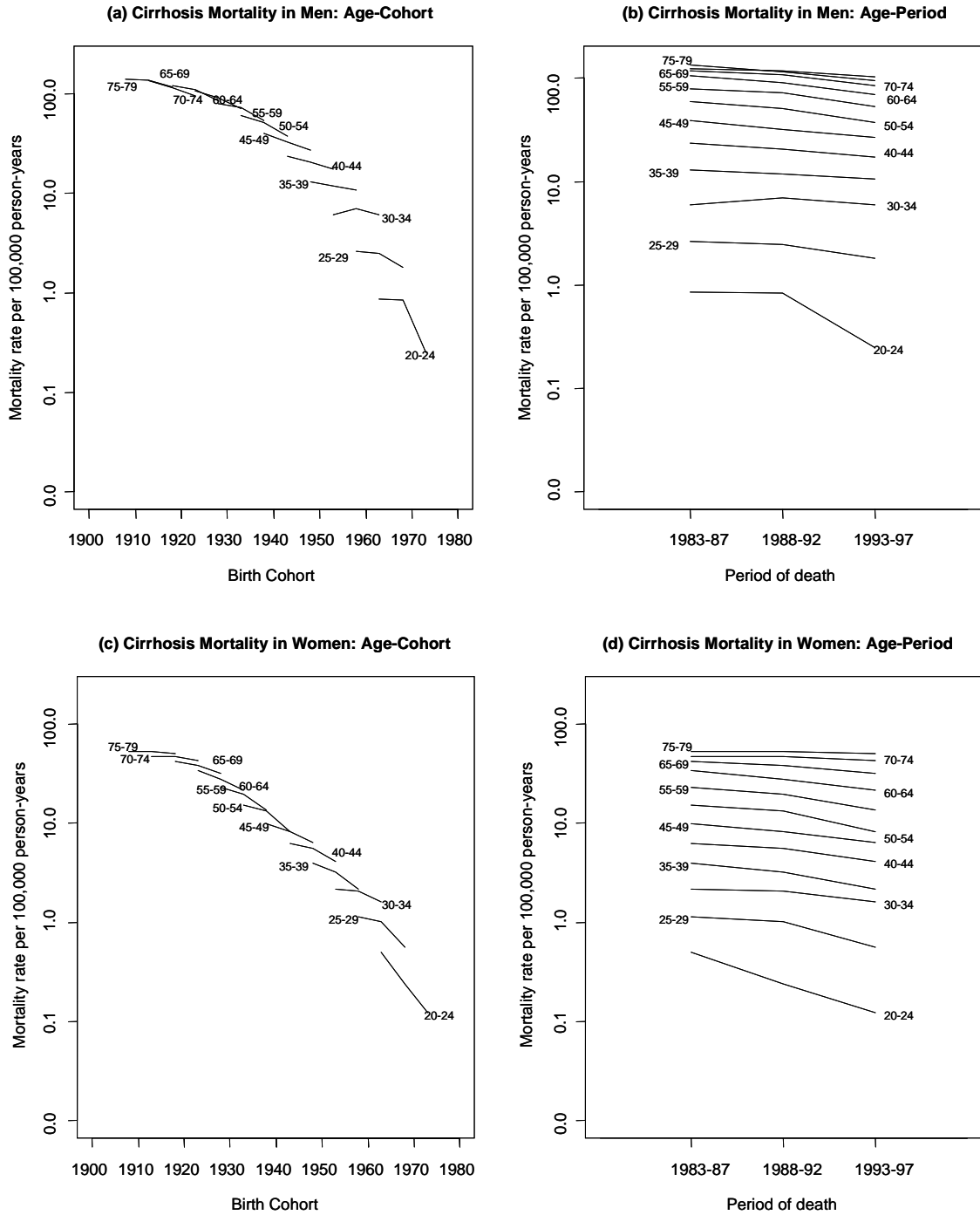
(d) Cirrhosis Mortality in Women: Age-Period

Table 5.5 shows that *DIC* values for the AC model (*DIC* men: 430.1, *DIC* Women: 358.7) were lowest than those of the AP models (*DIC* men: 492.7, *DIC* women:

516.3), reflecting the importance of the cohort effect for this cause of death. However, both APC1 and APC2 models were those which had the lowest *DIC* and *pD* values for both men and women, although differences in *DIC* and *pD* values among both models were minimal. Deviations from linearity for the cohort effect of cirrhosis mortality were estimated on the basis of both APC1 and APC2 models.

Table 5.5. DIC and pD values of the Age-Period-Cohort analysis for liver cirrhosis.

Model	Men		Women	
	DIC	*pD*	*DIC*	*pD*
AE	2008.4	11.8	948.4	12.1
AD	539.3	12.8	533.1	13.3
AP	492.7	15.6	516.3	14.3
AC	430.1	24.6	358.7	25.5
APC1	397.5	27.7	352.4	27.1
APC2	397.7	28.4	355.5	27.7

DIC: Deviance Information Criterion; **pD:** Effective number of parameters; **AE:** Model with Age effect; **AD:** Model with Age and Drift effects; **AP:** Model with Age and Period effects; **AC:** Model with Age and Cohort Effects; **APC1:** Age-Period-Cohort model with autoregressive smoothing for the age, period and cohort parameters; **APC2:** Age-Period-Cohort model as APC1 with constraint on the second order differences related only with the age parameter.

In both sexes, deviations from linearity of the cohort effect obtained with cohort parameter estimates did not differ from APC1 and APC2 model. *Figure 5.5.* represents deviations from linearity of the cohort effect and the exclusive cohort effect obtained with an APC1 model versus birth cohorts. In men (*Figure 5.5-a*), log-rate ratios for subjects born after 1950 showed the curvature in the cohort effect, assessing that mortality rates of cohorts younger than 35 had different risk of death than older cohorts (reference cohort 1933). However, this finding in these younger cohorts was not clearly assessed for women (*Figure 5.5-b*).

Figure 5.5. Logarithm of the rate ratios for deviations from linearity of the cohort effect of liver cirrhosis in men (a) and women (b) (Cohort exclusive effect obtained with the APC1 model).
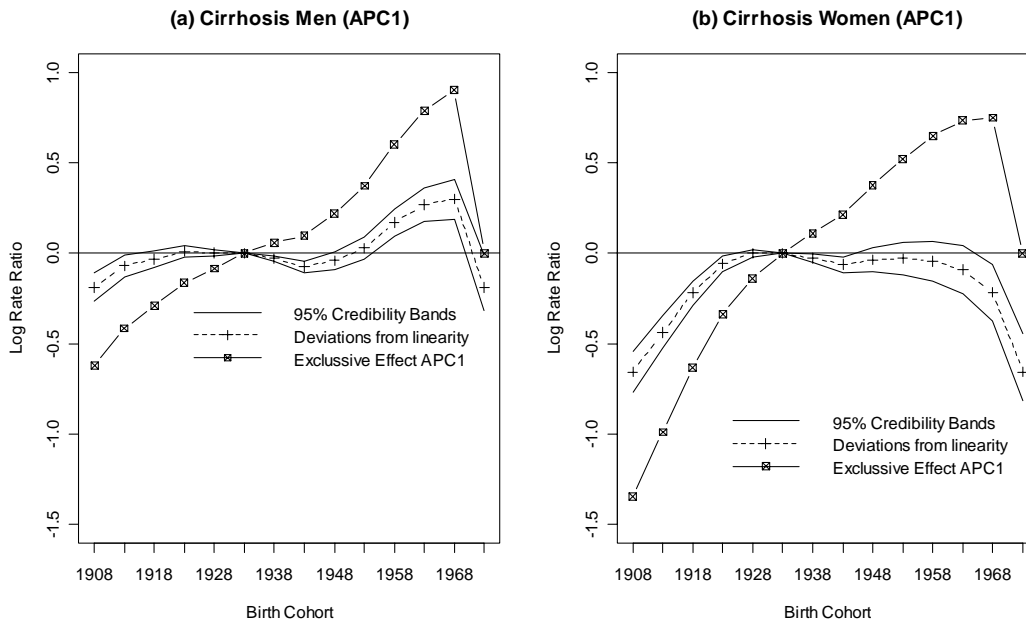
5.5.2.2. Hepatocellular carcinoma

Figure 5.6 represents crude incidence and mortality rates due to hepatocellular carcinoma by age groups versus birth cohorts for men. Rates rose dramatically during the study period for men older than 45 (birth cohorts before 1950) for both incidence and mortality data (*Figure 5.6-a*). Men younger than 30 did not show an increasing incidence of hepatocellular carcinoma throughout the period 1983-97 (*Figure 5.6-b*). However, mortality data showed the increasing rates of hepatocellular carcinoma throughoutthe period 1983-97 in all age-groups with the exception of men younger than 25 years old.

Although mortality and incidence rates due to hepatocellular carcinoma in women were lower than those of men, these showed a similar pattern. Among women, incidence rates (*Figure 5.7-a* and *5.7-b*) showed high variability although their rise

was clearly observed for women born before 1940 (older than 55 years). The rise in death rates for hepatocellular carcinoma in women (*Figure 5.7-c* and *5.7-d*) was observed for those born before 1930 (older than 65 years), whereas mortality rates remained stable or decreased throughoutthe period for women younger than 54.

Figure 5.6. Hepatocellular Carcinoma men: Crude incidence and mortality rates per 10^5 person-years by birth cohort and period of diagnosis or death according to age group.

(a) Hepatocellular Carcinoma Incidence in Men: Age-Cohort

(b) Hepatocellular Carcinoma Incidence in Men: Age-Period

(c) Hepatocellular Carcinoma Mortality in Men: Age-Cohort

(d) Hepatocellular Carcinoma Mortality in Men: Age-Period

Figure 5.7. Hepatocellular carcinoma women: Crude incidence and mortality rates per 10^5 person-years by birth cohort and period of diagnosis or death according to age group.

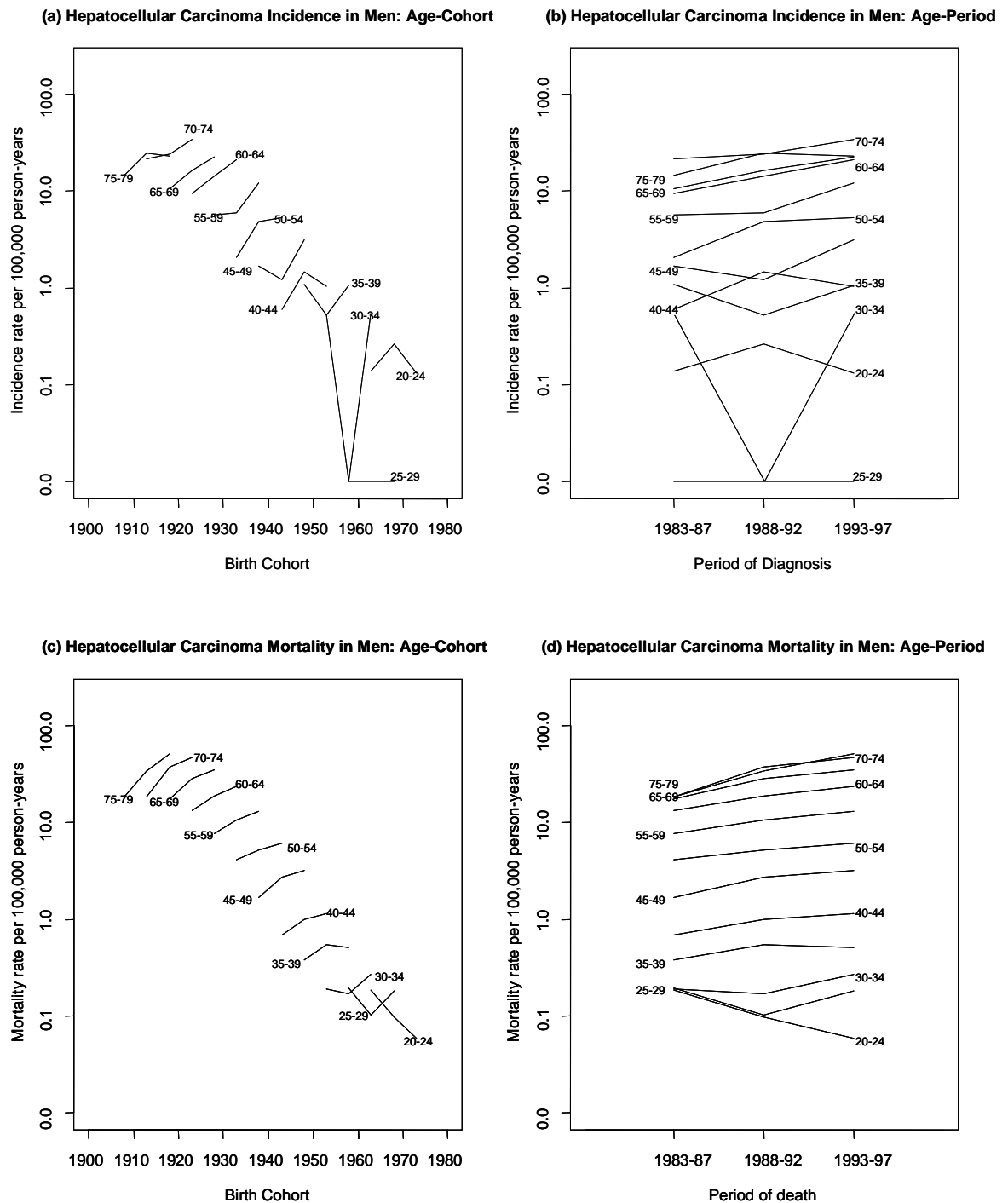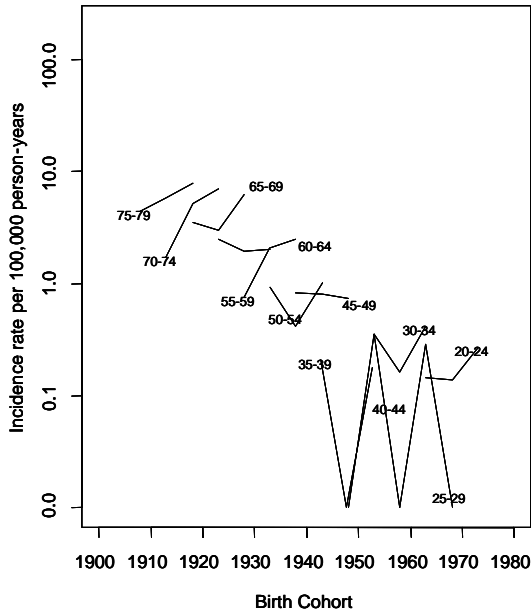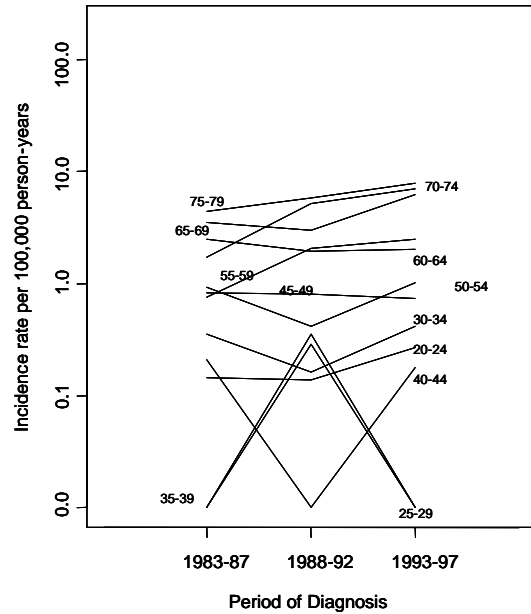**(a) Hepatocellular Carcinoma Incidence in Women: Age-Cohort**

(b) Hepatocellular Carcinoma Incidence in Women: Age-Period

(c) Hepatocellular Carcinoma Mortality in Women: Age-Cohort
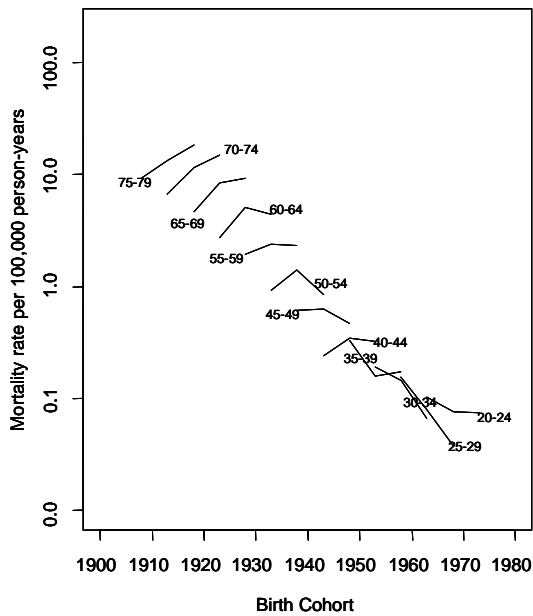
(d) Hepatocellular Carcinoma Mortality in Women: Age-Period
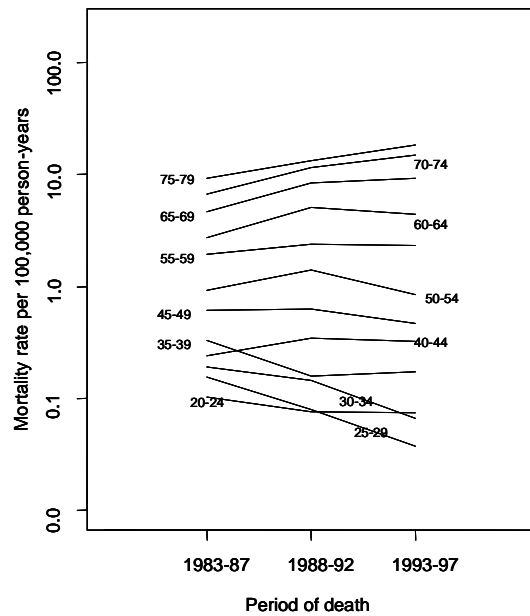
The extent of this cohort effect was only detected for mortality data through the model selection procedure described in *Table 5.6*. Incidence rates of hepatocellular carcinoma in men could be explained by an AD or AP model due to those models showed the lowest *DIC* values (*DIC* AD=200.6, *DIC* AP=207.1). However, the model with the lowest *DIC* value for mortality data was the APC2 model (*DIC*=308.2), which has lower number of effective parameters (*pD*=22.5) than APC1 model (*DIC*=315.3, *pD*=27.2). The same pattern was observed for women in the model selection procedure (Table not shown), being the AP and the APC2 models those which had the lowest *DIC* value for incidence and mortality data, respectively.

Table 5.6. DIC and pD values of the Age-Period-Cohort analysis for hepatocellular carcinoma: men.

	Incidence		Mortality	
Model	*DIC*	*pD*	*DIC*	*pD*
AE	265.7	11.2	919.6	12.2
AD	198.5	13.8	367.4	13.1
AP	200.6	14.3	369.2	14.1
AC	207.1	25.1	319.7	25.4
APC1	231.4	24.5	315.3	27.2
APC2	215.6	19.3	308.2	22.5

DIC: Deviance Information Criterion; *pD:* Effective number of parameters; **AE:** Model with age effect; **AD:** Model with Age and Drift effects; **AP:** Model with Age and Period effects; **AC:** Model with Age and Cohort Effects; **APC1:** Age-Period-Cohort model with autoregressive smoothing for the age, period and cohort parameters; **APC2:** Age-Period-Cohort model as APC1 with constraint on the second order differences related only with the age parameters.

Figure 5.8 presents deviations from linearity for the cohort effect and the exclusive cohort effect obtained with an APC2 model, which has been applied to hepatocellular carcinoma mortality data. In men (*Figure 5.8-a*), deviations from linearity of cohort effect show that risk of death by hepatocellular carcinoma smoothly decreased in

cohorts born after 1940. For the same cohorts in women (*Figure 5.8-b*), a dramatic decrease in mortality rates for this tumor was observed since 1940.

Figure 5.8. Logarithm of the rate ratios for deviations from linearity of the cohort effect of hepatocellular carcinoma mortality in men (a) and women (b) (Cohort exclusive effect obtained with APC2 model).
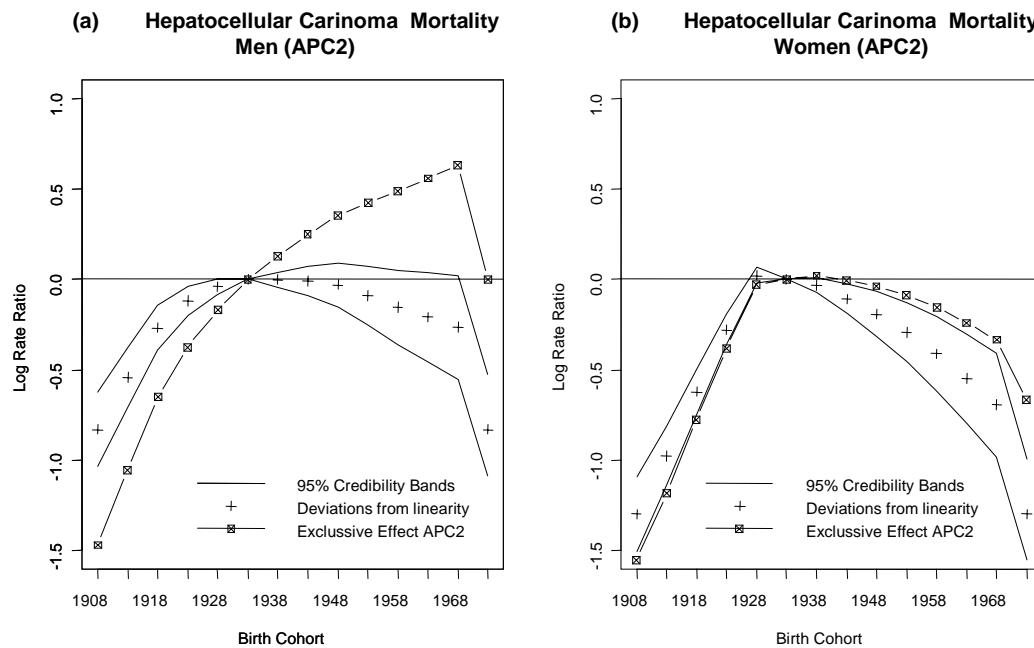
5.5.2.3. Cholangiocarcinoma

Crude incidence and mortality rates for cholangiocarcinoma by birth cohorts and period of diagnosis and death in men are represented in *Figure 5.9*. High variability of rates was observed with both incidence and mortality data, mainly due to the low number of incident and death cases observed throughout the period. Incidence rates rose in men older than 40 years old with the exception of the age group 65-69, for which incidence rates fell (*Figure 5.9-a* and *5.9-b*). However, mortality rates (*Figure 5.9-c* and *5.9-d*) increased in all age-groups with the exception of men 25-29 years old, being the rise more pronounced for those older than 55 years old. *Figure 5.10-a* and *Figure 5.10-b* show the high variability of incidence rates for this cause in

women. As a difference in men, the rise in incidence was more pronounced in women older than 55. However, the rise in mortality rates for this cause (*Figure 5.10-c* and *5.10-d*) was detected in all age groups during the study period.

Figure 5.9. Cholangiocarcinoma men: Crude incidence and mortality rates per 10^5 person-years by birth cohort and period of diagnosis or death according to age group.


(a) Cholangiocarcinoma Incidence in Men: Age-Cohort

(b) Cholangiocarcinoma Incidence in Men: Age-Period

(c) Cholangiocarcinoma Mortality in Men: Age-Cohort

(d) Cholangiocarcinoma Mortality in Men: Age-Period

Figure 5.10. Cholangiocarcinoma women: Crude incidence and mortality rates per 10^5 person-years by birth cohort and period of diagnosis or death according to age group.
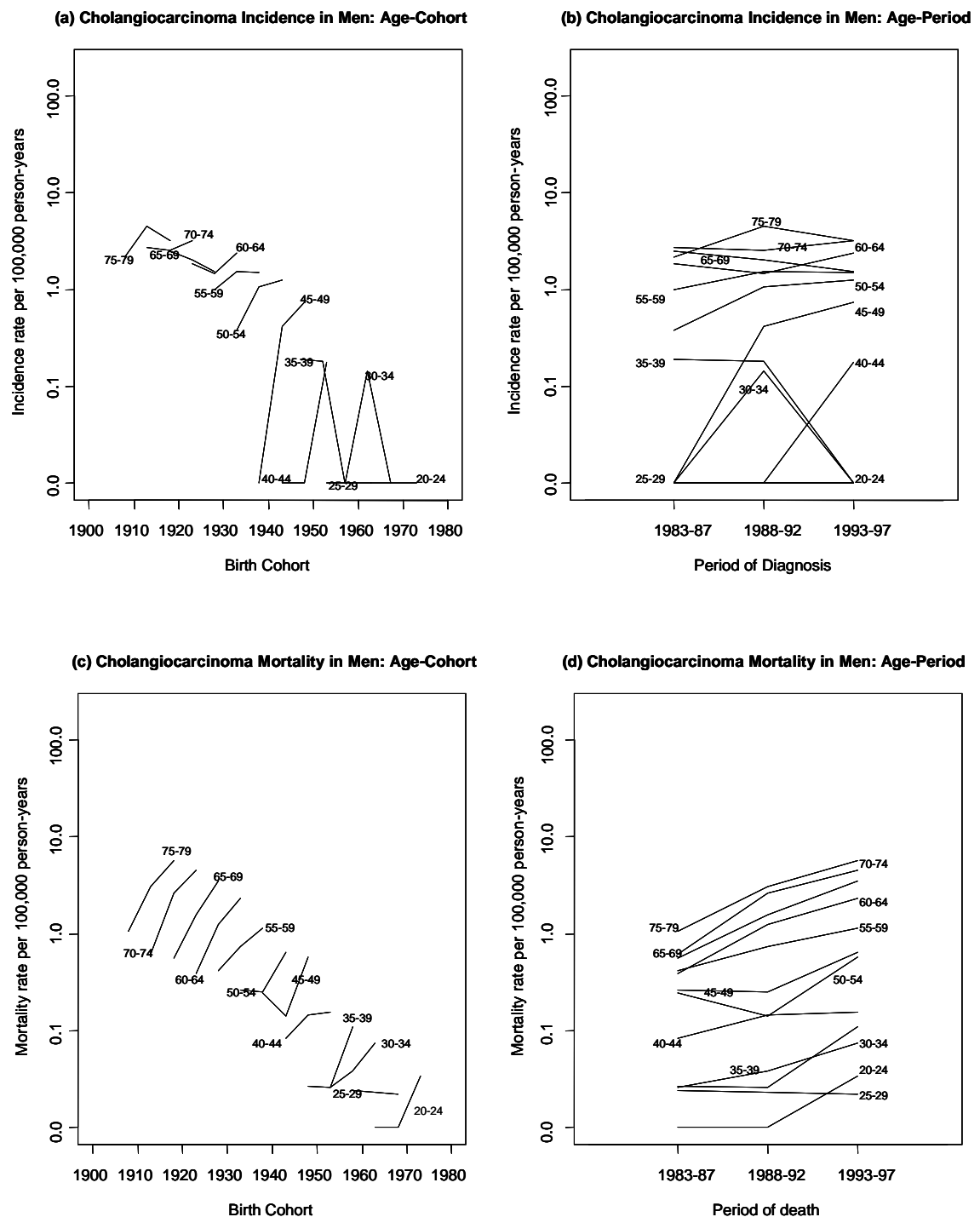
(a) Cholangiocarcinoma Incidence in Women: Age-Cohort

(b) Cholangiocarcinoma Incidence in Women: Age-Period

(a) Cholangiocarcinoma Mortality in Women: Age-Cohort

(b) Cholangiocarcinoma Mortality in Women: Age-Period

177
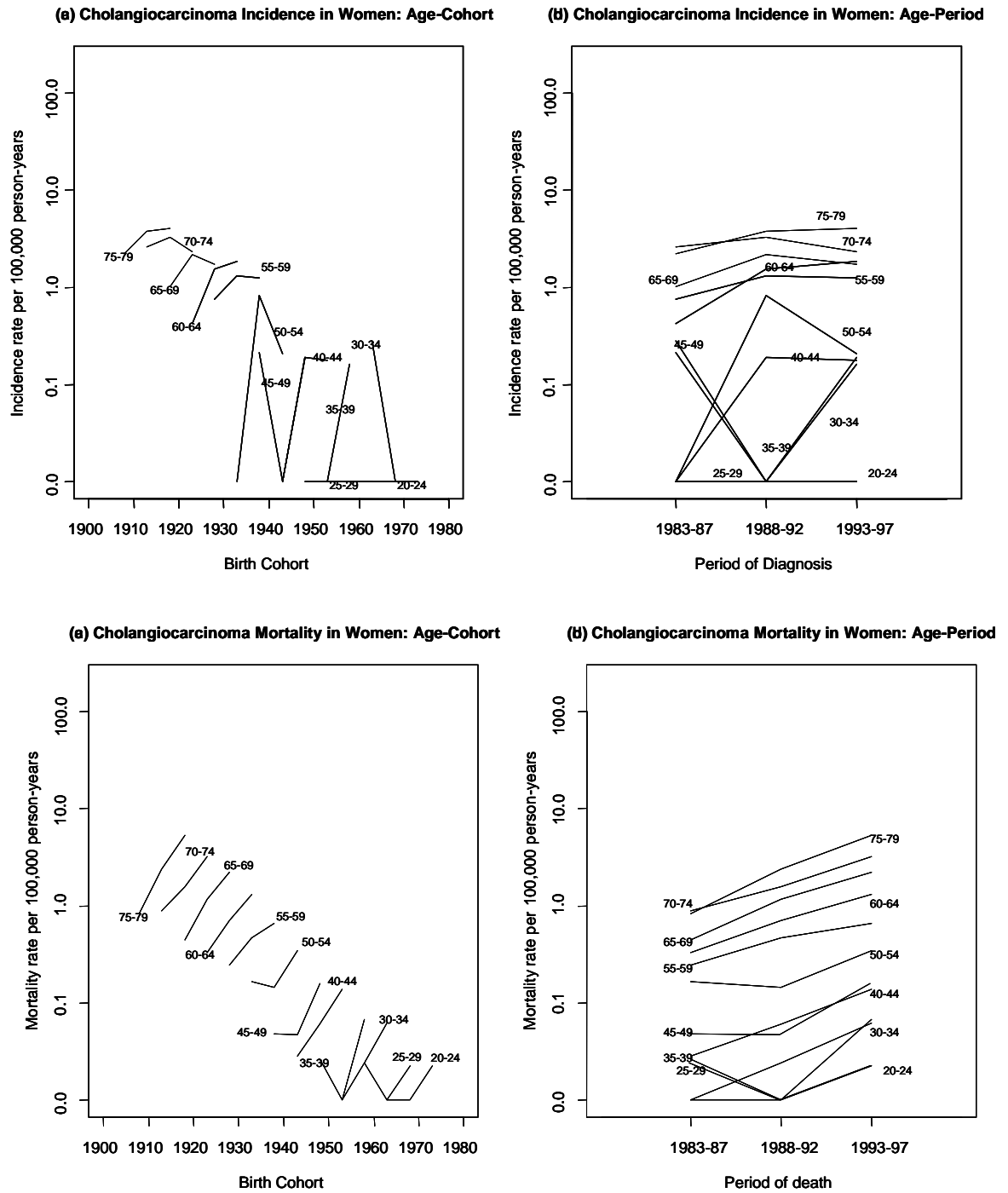
In the model selection procedure, all models showed similar *DIC* values for the cholangiocarcinoma incidence in men (*Table 5.7*). Based on this criteria, both AE (*DIC*=122.5) and AD (*DIC*=122.6) models could explain variability of this incidence tumor in men. Note that the effective number of parameters of these models are similar to the parameters included in the model (AE model 12 parameters and *pD*=11.7, AD model 13 parameters, *pD*=12.3).

Table 5.7. *DIC* and *pD* values of the Age-Period-Cohort analysis for cholangiocarcinoma: men.

	Incidence		Mortality	
Model	*DIC*	*pD*	*DIC*	*pD*
AE	122.5	11.7	574.2	12.7
AD	122.6	12.3	213.1	12.5
AP	123.8	12.4	214.2	14.4
AC	128.5	20.4	224.1	28.5
APC1	123.9	10.6	198.5	7.6
APC2	127.1	14.2	199.5	7.4

DIC: Deviance Information Criterion; *pD:* Effective number of parameters; **AE:** Model with age effect; **AD:** Model with Age and Drift effects; **AP:** Model with Age and Period effects; **AC:** Model with Age and Cohort Effects; **APC1:** Age-Period-Cohort model with autoregressive smoothing for the age, period and cohort parameters; **APC2:** Age-Period-Cohort model as APC1 with constraint on the second order differences related only with the age parameters.

For mortality data, the APC1 or the APC2 models should be the selected ones due to those had the lowest *DIC* values (*DIC* APC1=198.5, *DIC* APC2=199.5). However, these models had much lower number of effective parameters (APC1 *pD*=7.6, APC2 *pD*=7.4) than those included in the model (27 parameters: 12 age parameters, 13 (14 minus reference cohort) cohort parameters and 2 (3 minus reference period) period parameters. This discrepancy could indicate a bad fit of the model, especially if we compare the *pD* values of the AE, AD, AP and AC models which are close to the number of parameters included in the model. For that reason it was considered that

an AD or an AP models were adequate to explain variability of mortality rates as it was concluded for incidence data. The same conclusions in the model selection procedure were applied for men as well as for women, being the AD or AP models the selected ones in terms of *DIC* and *pD* for both incidence and mortality data, respectively.

5.5.3. Summary of time trends analysis

Table 5.8 summarizes the model selection and the period and cohort trends for each liver disease analysed. There has been observed a decreasing mortality trend (APCH about 3%) for liver cirrhosis which could be due to both period and cohort effects in both sexes. In men between 25 to 35 years old, cirrhosis mortality did not show a decreasing trend (cohort effect). An increase in incidence and mortality rates for hepatocellular carcinoma has been observed. In both sexes there was detected a similar magnitude of increase in the incidence and mortality rates (APCH men about 6%; APCH women about 5%). For this tumor, incidence trend was associated with a period effect, whereas mortality trends were explained by both cohort and period effects. Finally, an increase on incidence and mortality by cholangiocarcinoma for both sexes was also observed. Although the incidence trend was not significant, both incidence and mortality trends could be related to a period effect.

Table 5.8. Summary of results for the time trends analysis.

Cause	Model Selected	Period Effect APCH (95% CRI)		Cohort Effect * Deviation from linearity
Mortality				
Liver Cirrhosis	APC1 & APC2	Men:	-3.1 (-5.1, -1.9)	Non decreasing risk of death for birth cohorts after 1950
		Women:	-2.9 (-6.2, -1.3)	Non decreasing risk of death for birth cohorts after 1950
Hepatocellular Carcinoma	APC2	Men:	6.8 (5.8, 8.1)	Smooth decrease for the risk of death for birth cohorts after 1933
		Women:	5.1 (3.5, 6.3)	Dramatic decrease for the risk of death for birth cohorts after 1933
Cholangiocarcinoma	AP	Men:	17.1 (13.5, 21.2)	Not detected
		Women:	15.0 (11.5, 19.5)	Not detected
Liver Cancer Unspecified	AP	Men:	-7.2 (-8.2, -6.2)	Not detected
		Women:	-8.6 (-9.2, -8.1)	Not detected
Incidence				
Hepatocellular Carcinoma	AP	Men:	6.6 (5.3, 9.2)	Not detected
		Women:	4.5 (1.4, 7.3)	Not detected
Cholangiocarcinoma	AP	Men:	3.3 (-1.3, 8.3)	Not detected
		Women:	3.7 (-1.2, 6.1)	Not detected
Liver Cancer Unspecified	AP	Men:	-1.2 (-3.4, 2.3)	Not detected
		Women:	-3.5 (-5.2, -2.1)	Not detected

* Reference cohort : 1933

APCH: Annual Percent Change; **95% CRI:** 95% Credibility Interval; **AP:** Model with Age and Period effects; **APC1:** Age-Period-Cohort model with autoregressive smoothing for the age, period and cohort parameters; **APC2:** Age-Period-Cohort model as APC1 with constraint on the second order differences related only with the age parameters.

5.6. Discussion

5.6.1. Statistical methodology

The statistical methods proposed in this age-period-cohort analysis have allowed us to assess the effect of those parameters in the variability of rates. We have also investigated if there was any deviation from linearity (curvature) of cohort effects based on the methods proposed by Holford[202,204,208,222], which we have adapted to the Bayesian framework. These curvatures have been extracted from the two Bayesian APC models proposed. As a difference from the classical age-period-cohort analysis, the Bayesian models that we have applied assume an autoregressive structure among parameters in order to treat second order differences of parameters as independent normal covariates.

We have followed a selection procedure starting with the AE model, followed by AD, AP and AC models. As we have indicated, an APC model is required only when neither of these models provides a satisfactory fit[206,207]. We have selected the full APC model for liver cirrhosis and hepatocellular carcinoma mortality, whereas the AD or AP models were adequate to explain variability of incidence rates and mortality by cholangiocarcinoma. We should note that an APC model was selected for those causes of death with the largest number of cases.

For liver cirrhosis mortality it has been observed decreasing rates throughout the period of study. However, a deviation from linearity of that decreasing trend in rates was observed for younger cohorts (born after 1950). This was detected with both APC1 and APC2 models, which showed almost identical cohort parameter estimates. For that reason deviations from linearity estimated from both models coincide with shape of curvature.

The APC2 model was chosen for hepatocellular carcinoma mortality. For this cause of death, the APC1 model showed an effective number of parameters value (pD=27.7) almost identical to the sum of age groups, period and cohort parameters included in the model. In addition, pD of APC1 was higher than that obtained with APC2. An explanation for this difference could be the degree of smoothing of the age effect among APC1 and APC2, due to age parameters showed little differences among APC1 and APC2 in terms of median value and variance.

It has been described that constraint on the degree of smoothness appear to be not sensitive to the prior of the roughness parameters (σ_γ) of APC1 and APC2[215,219]. This could be due to the fact that projections based on autoregressive age-period-cohort models are uniquely determined[219]. However, it could be considered other prior distributions on the model parameters or functions of them, such as to assume uniform prior distributions for standard errors of the parameters, as it has been recently described[162].

The AP model has been the chosen one to describe mortality and incidence by cholangiocarcinoma, which accounted for low number of observed cases during the study period. In the model selection procedure for cholangiocarcinoma mortality, the APC1 and APC2 models showed the lowest DIC and pD values, and these models should be the selected ones. But it should be taken into account that low values of pD may indicate a bad fit of the model[137], possibly due to overshinkrage of random effects or collinear fixed effects[147]. Also pD represents a decrease in the deviance due to the inclusion of parameters. In this case, a low pD value represents a low decrease in deviance and therefore, bad fit of the model. For that reason, it was considered the AP model as adequate to describe mortality data. In addition, cholangiocarcinoma incidence showed an excess of zeros in the dataset whereas the models proposed did not take into account this issue. To overcome to this problem a Zero Inflated Poisson model (ZIP) could be an alternative to the models applied in this analysis[142,143,223].

It should be underlined that this decrease in the pD value when using the APC models has been clearly observed in the analysis of incidence and mortality from cholangiocarcinoma, which is the liver disease with the lowest number of cases. As we pointed in methods section of Chapter 4, pD and DIC were calculated by means of the *bugs* function of R2WinBUGS library, which estimates pD as a function of the variance of the posterior average deviance[137]. This is a difference with the initial approach described by Spiegelhalter et al[160] who derived $pD = E(D) - D(\bar{\theta})$. Although both estimatives of pD were derived on an asymptotic χ^2 distribution[137], it remains to investigate if this discrepancy in the estimation of pD could entail a different approximation to the "effective number of parameters" of the model. Despite SRF of all APC parameters reached convergence, it is clear that those APC models does not fit adequately cholangiocarcinoma data. The AP model can be considered adequate to explain variability of cholangiocarcinoma mortality.

Graphical representation of period, cohort and age effects has been performed through the Holford method assuming that cohort was the residual scale of time. The Bayesian implementation of this method has been performed taking into account a sensibility analysis on the prior distributions of the precision parameters (Appendix A.5). It has been observed that posterior precision strongly depends on the prior distribution used. Among several ones, Gamma(0.01,0.01) has been selected as a prior distribution due to its good performance (fast convergence and small standard deviation of the posterior parameter estimates). We believe that this shoudl be an important issue to evaluate in each APC analysis on the Bayesian framework.

In this Bayesian approach to time trends modelling, we suggest to proceed by a previous graphical inspection of rates, and then use DIC and pD procedure to select the adequate model. If the full APC model is considered as adequate, cohort and period effects extracted form an APC1 and APC2 model should not differ. However, if

our interest was to describe the extra Poisson variation, an AC or AP model with an unstructured error term would be adequated[215].

5.6.2. Epidemiological analysis

In Spain during the period 1983-97, there has been observed a statistically significant decrease in liver cirrhosis mortality in both sexes. These results are in concordance with those of previous published studies conducted in some developed countries[224-230] and in Thailand[231]. Some authors have been suggested that the general fall observed on cirrhosis mortality in the last decades could be explained by the advances in the treatment of cirrhosis and its complications[232], such as liver transplantation. In this line, it has been reported that more than 1,000 liver transplantations are being performed per year in Spain[233]. Among European population, the Spanish citizens are those who have the highest probabilities to access to liver transplant if required[234].

In addition, it has been suggested that fell in mortality rates by cirrhosis mortality could be explained by the reduction in alcohol consumption[224]. During 1983-96, in a study conducted in Italy, a country with a similar drinking culture than Spain, it was showed a decrease in the estimated number of deaths due to liver cirrhosis, that was partially explained by the reduction of alcohol consumption[224]. Other studies conducted in the U.S., showed similar results[235]. In Spain, it has been observed a decline in moderate alcohol consumption during the 1987-93 period, which has been associated with a decrease in heavy drinkers (-1.2%)[236,237]. On the contrary, it has been reported in Britain an increase of mortality by liver cirrhosis during the last 50 years, which has been mainly associated with the increase of alcohol consumption[230].

In spite of the decreasing trend in cirrhosis mortality detected in Spain, it should be underline that cirrhosis mortality risk did not decrease in men for those cohorts born between 1950-65. In the same line, in a previous study conducted in Catalonia we found a statistically significant increase of 4.7% in liver cirrhosis mortality among

men with 25 to 35 years old[50]. The slight discrepancy between the studies conducted in Spain and Catalonia could entail that geographical variability of risk of death by liver cirrhosis could exist in Spain. In this line, this variability is reliable and verifiable through the inspection of cirrhosis mortality maps in the analysis of mortality in small areas in Spain during 1985-1997, an study carried out by *Benach et al*[238].

This pattern in liver cirrhosis mortality suggests that younger cohorts could be exposed to some additional risk factors. Some life styles more common between people born in the 60 and 70's[239], as intravenous drug addiction, could explain the exposure to viral infections related to cirrhosis[240,241]. Co-infection with HBV or HCV and human immunodeficiency virus (HIV) is common among intravenous drug users. It has been estimated that in Spain, HIV-HBV co-infection affects more than 5,000 people and HIV-HCV co-infection affects more than 60,000 people[240,241]. These results are compatible with the increase of liver cirrhosis risk among HCV-HIV and HBV-HIV co-infected patients, as it has been described in some studies[242,243].

Although liver cirrhosis incidence data in Spain is not available, we could suppose that this incidence did not diminish during the study period. For that reason, this potential improvement on cirrhosis survival can increase the risk of developing hepatocellular carcinoma in cirrhotic patients. The results presented in this chapter give support an increase of hepatocellular carcinoma incidence and mortality rates as it has been described in other countries[26,46,47,123,185,195,244-247]. Part of this trend is likely to be attributable to the introduction of new diagnostic techniques in the early 1980s, such as ultrasound and computed tomography scanning, which have led to an improvement of diagnosis and increase in the number of hepatic biopsies conducted. This could be an explanation for the period effect observed in the analysis of hepatocellular carcinoma incidence. However, in the analysis of mortality by hepatocellular carcinoma, both period and cohort effects explained the time trends. Two studies carried out in Spain have reported that HCV was the main aetiological agent found among cases diagnosed with hepatocellular carcinoma[179,180]. Population-

based studies which determined HCV prevalence in some Spanish areas showed that HCV infection was most prevalent than HBV infection among general population[197,198]. In one of these studies, which was conducted in Catalonia, it was detected that HCV prevalence increased significantly with age[248]. Under the hypothesis of similar pattern for the whole Spanish population, we can assume an important exposure to HCV in the oldest cohorts, which has been reflected with the increase in hepatocellular carcinoma mortality and incidence at the end of the study period. This fact was also detected in Italy and Japan, suggesting that important exposure to HCV have occurred 30-50 years ago[79,245]. In addition, a significant increase in hepatocellular carcinoma mortality, incidence and hospitalization rates has been observed also in the United States during the period 1993-99 that has been also attributed to HCV exposure[246].

A significant rising in cholangiocarcinoma mortality rates has also been detected in Spain for both sexes. These findings are consistent with those from Japan, Australia, U.K. and U.S.[26,47,49,249]. However, the observed increase in cholangiocarcinoma incidence detected in our study has not been found to be significant. The disagreement between mortality and incidence trends of cholangiocarcinoma could probably be due to the lower number of incident cases (based on five cancer registries which cover 10% of Spanish population) compared to the number of deceased cases (Spanish Mortality Registry).

Period effect explained the increase in cholangiocarcinoma mortality, that could be attributed to improvement in diagnosis from better imaging and diagnostic techniques[47]. This could be also an explanation to the decrease of incidence and mortality rates of liver cancer unknown of our study, which has been also detected in France, Italy, Australia and Japan[47]. Although case ascertainment my not be the sole factor to explain the rise in mortality rates for the increase in cholangiocarcinoma mortality[47]. In addition, the most recently studies suggest that HCV infection, diabetes and obesity probably play a role in cholangiocarcinoma carcinogenesis [250].

In summary, we have observed in Spain an increase of incidence and mortality rates of hepatocellular carcinoma and cholangiocarcinoma. These trends could be partially explained by improvements in diagnostic techniques implemented during the period of study. Exposure to HCV 30 years ago would be also related to hepatocellular carcinoma trends. No obvious changes on risk factors exposures associated to cholangiocarcinoma have been reported in the same period. The non-decreasing risk of death due to liver cirrhosis among 25-35 year-old men, strengths the need to carry out screening studies to detect liver disease among these young populations, in order to prevent a future burden of increase in the incidence of liver cancer in Spain.

Chapter 6

6. General discussion

6.1. Statistical methodology

The statistical methods of this monograph have been based on the Bayesian inference, and three different statistical analyses have been carried out on this context. Inference has been approached by Poisson distribution in each one of these analyses. The modelling of events via Poisson distribution frequently entails extra-Poisson variability, known also as overdispersion, indicating that some unexplained clustering arises in the data. It has been applied different solutions to deal with over-dispersion in each one of the analyses.

In the meta-analysis framework of Chapter 3, extra-Poisson variability has been modelled by means of a Poisson mixture of probability distributions, assuming that data presents unexplained heterogeneity. The "stick –breaking" method has been implemented in order to estimate the parameters of the mixture of Poisson distributions. The RR_{Pooled} has been estimated on the basis of three theoretical subpopulations and their corresponding weights. We should note two remarks concerning with the "stick-breaking" method. First, the maximum number of components of the mixture of probability distributions should be assessed through a sequential process. Second, due to the "constructive" definition of that method, it is necessary to perform sampling from a $Beta(1,\delta)$ distribution. We suggest to consider δ as a random variable due to its possibly posterior variability. By means of the posterior classification of the studies into components of the mixture of distributions, observations entail heterogeneity could be detected.

Random effects models have been used to address extra-Poisson variability in the remaining two analysis of this monograph. In the spatial analysis, RRs have been modelled through a Poisson-lognormal model which includes both HBV and HCV seroprevalences, an "European area" variable and unstructured random effects. The common models used for mapping disease risks take into account both unstructured

and structured ("local" or regional) random effects[166,163]. In addition, Poisson-Normal-CAR (Conditional Autoregressive) models seem more adequate models in the small-area analysis, due to similarities among districts or towns are more credible than similarities among countries. However, the autoregressive structure of constraints imposed to the model parameters of the time trends analysis have the role of auto-regressive random effects into the classic APC models[214].

Magnitude of these random effects in the spatial analyses has been modelled via their prior standard error. It has been assumed noninformative Uniform distributions for the prior standard errors of the parameters, due to it is expect to perform well than Gamma priors unless the number of levels of the variable is approximately below 5[162]. Unlike spatial analysis, we have modelled precision of the model parameters with Gamma distributions for the time trends analysis, as it has been described in the literature. This translates to a uniform distribution for the precission on the log-scale[217]. However, it should be also considered to evaluate the effect of other prior distributions on the precisions of model parameters, although it has been described that autoregressive smoothing eliminates the random fluctuation that occurs with maximum likelihood estimates[215,219,217].

Model selection has been performed via DIC and pD criterions[160]. Differences in DIC across models are meaningful, such that a model with a low DIC value than another is preferred. However, this criterion has not been used for the analysis of mixtures due to there are some possible inconsistencies in the definition of DIC for mixture models. The most notable is the occurence of negative dimension parameters, so negative pD (see DeIorio and Robert)[251] values. In addition Richardson[252] presented an alternative notion of DIC in the context of mixture models. The most important difficulty is related with the notion of deviance, which affects DIC, taking equally acceptable meanings. At present date, Celeux et al. have investigated the typology of DIC for missing data models, mixture models and random effects models[253]. They described that some of the extensions of the DIC notion are not

adequately to evaluate the complexity and fit of these models studied due to those exhibit too much variability from one model to the next and possibly negative pD values[253]. For that reason we have proposed the sequential approach to estimate the number of components and parameters of the mixture of Poisson distributions described in Chapter 3.

In the spatial analysis it was detected little differences in DIC values across models, and due to that fact, it has been also evaluated the value of the posterior standard error of the unstructured random effects. The model with the lowest posterior standard error of these random effects indicates that its covariates explain better variability of data. In the time trends analysis, APC models for some causes of incidence/death showed the lowest DIC values. However, we have observed that DIC should be evaluated jointly with pD, as we have observed in the analysis of cholangiocarcinoma mortality. For this cause of death, APC models showed the lowest DIC values and the lowest number of effective parameters, pD, indicating that few parameters are sufficient to explain variability with the complete autoregressive structure.

But we should note that estimation of DIC depends on pD, and these values have been obtained with those reported by the *bugs* function of the R2WinBUGS library. The implementation of pD in this library is based on the estimation of the variance of the *a posteriori* deviance described by *Gelman et al.*[137], whereas the original of *Spiegelhalter et al.*[160] is based on $pD = E(D) - D(\bar{\theta})$. Although both estimations are derived on asymptotic χ^2 distributions, it is beyond the objective of this monograph to investigate if both estimations could lead to different conclusions.

In summary, the flexibility of the Bayesian approach allowed us to model extra-Poisson variability in three statistical analyses, applying different models, and addressing relevant aspects that should be taken into account in each problem.

Challenging statistical issues in the framework of Bayesian applied modelling are i) the selection of prior distributions for model parameters, which is related with convergence of the model, and ii) model selection procedures, which warrant more research in this area.

6.2. Epidemiological analysis

We have observed that PLC risk worldwide strongly depends on the geographical area, being those areas with the highest risk locaded in Eastern Asia and Middle-Western Africa. We have described the geographical variation of PLC risk taking into account the effect of the HBV and HCV prevalences in different world areas. It has been estimated first the risk of death by PLC among HBV male carriers worldwide. In Europe, it has been determined which are the countries with highest risk of PLC taking into account the effect of both hepatitis infections. In Spain, a high risk country for PLC among European countries, it has been estimated the time trends of PLC and liver cirrhosis during the last 15 years period with data available.

In high risk areas for PLC risk it has been also estimated higher HBV population prevalences than that of HCV. We have shown that RR of death by PLC among HBV male carriers was 23.5 (95% CRI:14.9 - 44.5) on the basis of data reported in cohort studies carried out in Europe, Asia and America. However, this RR should be explained by geographical area and control group selected in the cohort study. The highest RRs were observed in the studies carried out in developing countries of Southeast Asia, where HBV transmission mainly occurs at younger ages or by perinatal infection. In these countries, HBV transmission and exposure to aflatoxins in food, which is also common in these areas, suggests an increase of risk of progression to PLC [73,150,87,151,152].

We have also quantified for the first time the "healthy donor effect" due to the fact that we have observed that RR of death by PLC among HBV male carriers is between 2-fold (low risk areas for PLC) and 5-fold (high risk areas for PLC) higher when comparing studies which used WBD as control group versus those which use GP as control group. This finding could lead to underestimate the RR of death by PLC among HBV male carriers in those studies which used GP as control group.

The decreasing role of HBV in the etiology of PLC has been reflected after the universal HBV vaccination programs initiated in most countries[158]. In developed countries, such as Japan and U.S., recent studies have shown that PLC cases linked with HBV remained faily constant, whereas PLC cases linked to HCV seems to increase dramatically[254,255]. In the geographical analysis of PLC in European countries, we have described that HCV seems to have the predominant role in the etiology of PLC in Europe, being Southern and Eastern European countries those areas with the highest RRs of incidence and death by PLC during 2002. Those countries have shown high posterior probabilities of RR of PLC higher than the European mean taking into account the effect of both HBV and HCV prevalences. Although this results of our study should take into account several limitations related with data quality, we have described the pattern of PLC distribution in Europe at the beginning of the 21st century.

The increasing incidence of PLC in developed countries ant the end of the 20th century has been also reported in European countries such as U.K, France and Italy[26,78,193,194,47,49,42]. However, time trends by PLC in Spain did not reflect any increase by this disease until present date. For the first time in Spain it has been analysed incidence data of PLC taking into account information of five Spanish cancer registries.

We have found an increase of incidence and mortality rates of hepatocellular carcinoma during the period 1983-97 in Spain. Exposure to HCV 30 years ago would be also related to hepatocellular carcinoma trends as it has been described in other studies of PLC[26,79,245,246,47,249,49]. We have also found an increasing trend of cholangiocarcinoma mortality in the same line as it has been found in Japan, Australia, U.K. and U.S.[26,47,249,49]. This increasing trend of cholangiocarcinoma mortality could be attributed to improvement in diagnosis from better imaging and diagnostic techniques[47]. However, we have not detected a significant increasing trend

of cholangiocarcinoma incidence, mainly due to the low number of cases reported by the Spanish Cancer Registries.

We have observed a decreasing trend of cirrhosis mortality in both sexes during the study period, although younger cohorts did not show this pattern. This cohort effect suggests that younger cohorts could be exposed to some additional risk factors and not only alcohol consumption. HIV and HCV or HBV co-infection[242,243] and intravenous drug addiction[240,241] could also explain the increase of liver cirrhosis mortality among younger cohorts.

Chapter 7

7. Conclusions

7.1. Statistical methods

Meta-analysis of cohort studies of liver cancer risk of death among HBV carriers

7.1.1) To estimate a combined measure of interest by means of mixtures of probability distributions is an alternative method to investigate which are the sources of heterogeneity in meta-analysis and it also allows to do not necessarily assume the gaussian distribution for the measure of interest.

7.1.2) The "stick-breaking" method allows to determine the number of components of the mixture of probability distributions on the basis of a sequential process.

7.1.3) In the "stick-breaking" method, the variability of the number of components of the mixture of probability distributions depends on a sampling procedure from a $Beta(1, \delta)$ distribution, being necessary to consider δ as a random variable.

Geographical distribution of primary liver cancer in Europe during 2002

7.1.4) The bayesian framework allowed to detect low-risk areas and high risk areas of a map through the *a posteriori* probability of RRs.

7.1.5) The inclusion of HBV and HCV prevalences significantly decreased the magnitude of the posterior standard deviation of the unstructured random effects.

7.1.5) A sensibility analysis on the selection of the prior standard error or precission of the unstructured random effects is necessary.

Time trends of liver disease in Spain during the period 1983-97

7.1.6) There are no significant differences between APC1 and APC2 models in the estimation of deviations from linearity in an age-period-cohort analysis.

7.1.7) It has been observed a dramatical decrease in the effective number of parameters, pD, for the APC1 and APC2 models in those causes with small number of cases.

7.1.8) Despite of the flexibility of the Bayesian modelling, the APC model was selected only for those causes of death with the largest number of causes.

Sensibility analysis for prior distributions

7.1.9) Although the Bayesian methods overcome some non-estimability problems that may occur when applying maximum likelihood estimation, a crucial point when using those is the specification of the prior distributions for the model parameters. A sensibility analysis on the choice of the prior distributions for the precission of model parameter is necessary in order to assess the inference from these models.

Choice of model

7.1.10) The concept that models with smaller *DIC* should be preferred to models with larger *DIC* should be considered taking into account the number of effective parameters (pD value) of the model.

7.1.11) In the model selection procedure of random effects models, there should be evaluated the decrease of the posterior standard error of these random effects jointly with the *DIC* and pD criterion.

7.2. Epidemiological analysis

Meta-analysis of cohort studies of liver cancer risk of death among HBV carriers

7.2.1) The pooled RR of death by PLC among HBV carriers is 23.5 (95% CRI: 14.9 – 44.5) being necessary to explain this RR by an effect of geographical area and by design of study.

7.2.2) The studies carried out with general population as comparison group showed a RR of 10.2 (95% CRI:7.9 – 12.8) for those studies carried out in low risk areas for PLC, whereas the RR was 18.6 (95% CRI: 17.3 – 20.1) for those studies carried out in high risk areas for PLC.

7.2.3) The studies carried out with workers or blood donors as comparison group showed a RR of 20.1 (95% CRI:13.5 – 21.6) for those studies carried out in low risk areas for PLC, whereas the RR was 103.0 (95% CRI: 88.6 – 118.3) for those studies carried out in high risk areas for PLC.

7.2.4) It has been quantified the "healthy donor effect" in longitudinal studies: a selection bias could be introduced in the estimation of RR of death when mortality in workers or blood donors is compared with that of general population.

Geographical distribution of primary liver cancer in Europe during 2002

7.2.5) High risk areas for PLC in Europe are located in Southern and Eastern European countries.

7.2.6) Northern European countries have the lowest risk of PLC in men with a RR of incidence of 0.48 (95% CRI: 0.37 – 0.61) and a RR of death of 0.51 (95% CRI: 0.41 – 0.60).

7.2.7) Northern and Western European countries have the lowest risk of PLC in women. The RR of incidence was 0.55 (95% CRI: 0.42 – 0.71) in Northern countries and 0.65 (95% CRI: 0.49 – 0.85) in Western countries. The RR of death in Northern countries was 0.65 (95% CRI: 0.53 – 0.79) whereas in Western countries it was 0.71 (95% CRI: 0.56 – 0.89).

7.2.8) In Europe, HCV seems to play the predominant role in PLC risk when adjusting for both HBV and HCV.

7.2.9) Those countries with HCV prevalence greater than 2% showed an increase of RR of incidence of 1.78-fold (95% CRI: 1.15 – 2.73) in men and 1.36-fold (95% CRI: 1.09 – 2.25) in women, whereas the increase of RR of death was 1.48-fold (95% CRI: 1.14 – 1.93) in men and 1.28-fold (95% CRI: 1.05 – 1.75) in women.

7.2.10) There could be an underestimation of PLC risk in Eastern European countries due to low PLC risks compared with high HBV and HCV seroprevalences. It is necessary the implementation of population-based cancer registries in Eastern European countries in order to assess the impact of PLC in these areas.

7.2.11) It should be carried out studies to determine HBV and HCV prevalences across Europe by sex and age-groups in order to predict the future impact of these infections on the PLC trends in incidence and mortality.

Time trends of liver disease in Spain during the period 1983-97

7.2.12) In Spain, it has been observed a decrease of liver cirrhosis mortality in Spain with an APCH in men of -3.1% (95% CRI: -5.1, -1.9%), whereas in women the APC was -2.9% (95% CRI: -6.2%, -1.3%). However, cirrhosis mortality did not decrease for men younger than 35 years during the study period.

7.2.13) The decrease of cirrhosis mortality in Spain could be attributed to advances in the treatment of cirrhosis and a reduction in alcohol consumption.

7.2.14) The increase of cirrhosis mortality among young cohorts could be partially attributed to intravenous drug addiction joinltly with co-infection with HBV or HCV and HIV.

7.2.15) In Spain, we have constated the increase of hepatocellular carcinoma incidence (APCH in men: 6.6%, 95% CRI: 5.8, 8.1: APCH in women: 4.5%, 95% CRI: 1.4%, 7.3%) and mortality (APCH in men: 6.8%, 95% CRI: 5.8%, 8.1%; APCH in women: 5.1%, 95% CRI: 3.5%, 6.3%).

7.2.16) The increase of incidence and mortality of hepatocellular carcinoma could be attributed to improvements in diagnostic techniques implemented during the period of study and to an important exposure to HCV in the oldest cohorts.

7.2.17) It has been detected an increase in mortality by cholangiocarcinoma (APCH in men: 17.1%, 95% CRI: 13.5%, 21.2%; APCH in women: 15.0%, 95% CRI: 11.5%, 19.5%), whereas this increase was not significant for incidence data.

7.2.18) Better imaging and diagnostic techniques could explain the increase of cholangiocarcinoma mortality.

7.2.19) In order to prevent a future burden of increase in the incidence of liver cancer in Spain is warranted to carry out screening studies to detect liver disease among young populations.

Appendix

Appendix A.1. WinBUGS code for models of Chapter 3

A.1.1 "Stick Breaking" Model using $\mu_j \sim Gamma(0.01, 0.001)$ and $\delta \sim Gamma(1,1)$

```
model
{        for( i in 1 : N ) {
         S[i] ~ dcat(p[])
         for (j in 1:C) {SC[i,j] <- equals(j,S[i])}
                 theta[i] <- mu[S[i]]*e[i]
                 x[i] ~ dpois(theta[i])}

#theta[k]: RR of k-th subpopulation
################################################
###### Constructive DPP
################################################
          pi[1] <- r[1]
          ac[1]<-1
          for (j in 2:C) {
                  pi[j] <- r[j]*(1-r[j-1])*pi[j-1]/r[j-1]
                       }

#Ordering constraint===> it is necessary > 2 components
  mu[1] ~dgamma(0.01,0.001)
    for (k in 2:C){
             mu[k]~dgamma(0.01,0.001)|(mu[k-1],)
             r[k] ~ dbeta(1,delta[k])
  # scaling to ensure sum to 1
             p[k] <- pi[k]/sum(pi[]) }
             p.s <- sum(p[])
################################################
###### End of constructive DPP
################################################

# Counts total clusters    (nonempty)
                  K <- sum(cl[])
                  for (j in 1:C) {
                          cl[j] <- step(sum(SC[,j])-1)
                          }
# Calculates mixture: mix
                  for (m in 1:C) {
                          smix[m]<-p[m]*mu[m]
                                }
                  mix<-sum(smix[])
# Prior on delta
                  for (h in 1:C)
                  {
                  delta[h] ~dgamma(1,1)
                  }

}
```

213

A.1.2.	"Stick Breaking" Model using $\log(\mu_j) \sim Normal(0, 0.001)$ and $\delta \sim Gamma(1,1)$.

```
model
{        for( i in 1 : N ) {
         S[i] ~ dcat(p[])
         for (j in 1:C) {SC[i,j] <- equals(j,S[i])}
                 theta[i] <- mu[S[i]]*e[i]
                 x[i] ~ dpois(theta[i])}

#theta[k]: RR of k-th subpopulation

# Constructive DP
         pi[1] <- r[1]
         ac[1]<-1
         for (j in 2:C) {
                 pi[j] <- r[j]*(1-r[j-1])*pi[j-1]/r[j-1]
                     }


   for (k in 1:C){
             log(mu[k])<-lmu[k]
             lmu[k]~dnorm(0,0.001)
             r[k] ~ dbeta(1,delta[k])
  # Ensure sum to 1
             p[k] <- pi[k]/sum(pi[]) }
             p.s <- sum(p[])

# End of constructive process


# Counts total clusters    (nonempty)
                 K <- sum(cl[])
                 for (j in 1:C) {cl[j] <- step(sum(SC[,j])-1)}
# Calculates mixture: mix
                 for (m in 1:C) {smix[m]<-p[m]*mu[m]}
                 mix<-sum(smix[])
                 for (i in 1:N){RR[i]<-mu[i]/e[i]}
# Prior on delta
                 for (h in 1:C)
                 {
                 delta[h] ~dgamma(1,1)
                 }

                 }
```

A.1.3. Sources of heterogeneity: Generalized linear model.

```
model
      {
      #X: High Risk Areas, Y: Low risk Areas
              for( i  in 1 : N ) {
                 mu[i]<-exp(alpha[1]*Y1[i]+beta[1]*X1+alpha[2]*Y2[i]+beta[2]*X2)
                 mitpo[i]<-mu[i]*e[i]
                         y[i] ~ dpois(mitpo[i])
              }
              for (j in 1:2)
              {
              alphak[j]~dnorm(0.0,0.01)
              beta[j]~dnorm(0.0,0.001)
              }
              for (j in 1:2)
              {
      alphar[j]<-exp(alphak[j])
      betark[j]<-exp(beta[j])
      }
Q1<-alphar[2]/alphar[1]
Q2<-betark[2]/betark[1]
Q3<-betark[1]/alphar[1]
Q4<-betark[2]/alphar[2]
      }
```

Appendix A.2. Sensibility analysis of prior distributions for random effects of Chapter 4 models

Extra-Poisson variability in the Spatial Analysis of Chapter 4 has been captured by means of the unstructured random effects. In order to evaluate the effect of different prior distributions for the precision of the unstructured random effects, we have evaluated four different prior distributions for that precision (τ_h). Sensibility analysis will be shown below for PLC incidence data in men. Similar conclusions could be derived for data related with the incidence in women, and mortality in both men and women.

Table A.2.1 Note that posterior median values for the standard error of the spatially unstructured random effects, $(\tau_h)^{-\frac{1}{2}}$, and its standard error appear to be more similar between Gamma (0.001,0.001) prior and Uniform (0,100). In addition, SRF converges to 1 for both Gamma (0.001,0.001) and Uniform (0,100) prior distributions after 60,000 iterations in all models M1-M6.

Table A.2.1. Different prior distributions for the precision of the unstructured random effects. Posterior standard error of the spatially unstructured random effects: Median value and standard error (in brackets) jointly with its Scale Reduction Factor (SRF) for data related with PLC Incidence Men.

	M1	SRF	M2	SRF	M3	SRF
Gamma (0.01,0.0001)	0.57 (0.18)	1.25	0.39 (0.13)	1.23	0.36 (0.14)	1.27
Gamma (0.1,0.0001)	0.55 (0.13)	1.14	0.37 (0.14)	1.09	0.37 (0.11)	1.12
Gamma (0.001,0.001)	0.49 (0.13)	1.05	0.31 (0.12)	1.05	0.30 (0.10)	1.02
Uniform (0,100)*	0.47 (0.11)	1.01	0.28 (0.07)	1.01	0.28 (0.04)	1.01

	M4	SRF	M5	SRF	M6	SRF
Gamma (0.1,0.0001)	0.34 (0.11)	1.22	0.31 (0.12)	1.18	0.29 (0.11)	1.17
Gamma (0.1,0.0001)	0.34 (0.12)	1.15	0.32 (0.13)	1.13	0.31 (0.11)	1.14
Gamma (0.001,0.001)	0.27 (0.09)	1.04	0.26 (0.08)	1.03	0.26 (0.08)	1.03
Uniform (0,100)*	0.24 (0.05)	1.01	0.24 (0.06)	1.01	0.24 (0.05)	1.01

* Prior distribution for the sd of the unstructured random effects

It has been evaluated the effect of that prior distributions on each of the models fitted. Table A.2.2 shows posterior median value, standard deviation and Scale Reduction Factor of the hepatitis viruses parameters of the model selected M5 (in log-scale).

Table A.2.2. Posterior median value, standard deviation (sd) and scale reduction factor (SRF) of M5 covariates (in log-scale) for PLC Incidence in Men. Sensibility analysis based on different prior distributions for the precision of the unstructured random effects.

	β_1			β_2			β_3		
	Median	sd	SRF	Median	sd	SRF	Median	sd	SRF
Gamma (0.01,0.0001)	0.0351	0.3911	1.2496	0.0215	0.2993	1.2975	0.7314	0.3365	1.2935
Gamma (0.1,0.0001)	0.0260	0.1911	1.1249	0.0125	0.2131	1.1514	0.6783	0.3483	1.1721
Gamma (0.001,0.001)	0.0231	0.1313	1.0114	0.0162	0.1635	1.0127	0.6314	0.2825	1.0137
Uniform (0,100)*	0.0193	0.1143	1.0012	0.0109	0.1316	1.0074	0.5712	0.2159	1.0096

β_1**:** Covariate related with HBV >2%; β_2**:** Covariate related with HCV prevalence between 1-2%; β_3**:** Covariate related with HCV >2%; * Prior distribution for the sd of the unstructured random effects.

Taking into account that each covariate has the same prior distribution, $\beta \sim N(0, 0.0001)$, the posterior parameter estimates and standard errors show a slight dependence on the prior distribution used for τ_h. After 60,000 iterations convergence of the β parameters appears to be not achieved when Gamma(0.01,0.0001) and Gamma(0.1,0.0001) distributions were used. However, convergence for the β parameters seems to be achieved in the framework of the Gamma(0.0001,0.0001) prior distribution used for τ_h, and Uniform(0,100) for the prior standard deviation of these random effects.

Appendix A.3. WinBUGS code for model M5 of Chapter 4

A.3.1 Model with HBV and HCV terms (Final model Chosen)

```
# A: Geographical Area Effect
# b.HCV: HCV effect , 3 levels, reference level 1
# b.HBV: HBV effect, 2 levels, reference level 1
#pp1: posterior probability of RR>1
#h: unstructurec random effects

model
{
  for (i in 1 : N) {
     O[i]  ~ dpois(theta[i])
     log(theta[i]) <-log(E[i])+ A[AREA[i]]+b.HCV[HCV[i]]+b.HBV[HBV[i]]+h[i]
     RR1[i]<-theta[i]/E[i]
     h[i] ~ dnorm(0, tau.h)        # Unstructured random effects
          pp1[i]<-step(RR1[i]-1)

  }


  # Other priors:

  tau.h  ~ dunif(0,100)


  sigma.h <- sqrt(1 / tau.h)

  for (j in 1:4)
  {
  A[j]~dnorm(0.0,0.0001)
   }
  b.HCV[1]<-0
  b.HCV[2]~dnorm(0.0,0.0001)
  b.HCV[3]~dnorm(0.0,0.0001)
  b.HBV[1]<-0
  b.HBV[2]~dnorm(0.0,0.0001)
}
```

A.3.2 Model with HBV, HCV and interaction terms for both seroprevalences

```
# A: Geographical Area Effect
# b.HCV: HCV effect , 3 levels, reference level 1
# b.HBV: HBV effect, 2 levels, reference level 1
# b.INT: Interaction between HBV and HCV, 3 levels: 1) no interaction,
# 2) interaction HBV>2 and HCV 1-2, 3) interaction HBV>2 and HCV>2
#pp1: posterior probability of RR>1
#h: unstructurec random effects

model
{
 for (i in 1 : N) {
    O[i]  ~ dpois(theta[i])
    log(theta[i]) <-log(E[i])+ A[AREA[i]]+b.HCV[HCV[i]]+b.HBV[HBV[i]]+b.INT[INT[i]]+h[i]
    RR1[i]<-theta[i]/E[i]
    h[i] ~ dnorm(0, tau.h)        # Unstructured random effects
        pp1[i]<-step(RR1[i]-1)

 }


 # Other priors:

 tau.h  ~ dunif(0,100)


 sigma.h <- sqrt(1 / tau.h)

 for (j in 1:4)
 {
 A[j]~dnorm(0.0,0.0001)
 }
 b.HCV[1]<-0
 b.HCV[2]~dnorm(0.0,0.0001)
 b.HCV[3]~dnorm(0.0,0.0001)
 b.HBV[1]<-0
 b.HBV[2]~dnorm(0.0,0.0001)
 b.INT[1]<-0
 b.INT[2]~dnorm(0.0,0.0001)
 b.INT[3]~dnorm(0.0,0.0001)
}
```

Appendix A.4. WinBUGS code for APC models of Chapter 5

A.4.1 Model APC1: Berzuini, Clayton, Bray et al.

```
# NOTA PRECISSIO X 1.0E-6 Seguint Bray et al. i Bashir i Esteve
# N: TOTAL DADES
# I: Grups Edat
# M: Periodes
# K: Cohorts

model
{
for (n in 1:N-M*I) {
pmoh[n] ~ dpois(mu[n]);
log(mu[n]) <- log(popn[n]) + alpha[age[n]] + beta[period[n]]
+ gamma[cohort[n]];
pred.mu[n] <- exp(mu[n])
pred.rate[n] <- 100000*pred.mu[n]/popn[n];
}


total <- sum(pred.mu[]);
##############################################################
#### PERIOD
betamean[1] <- 0.0;
betaprec[1] <- taup*1.0E-6;
betamean[2] <- 0.0;
betaprec[2] <- taup*1.0E-6;
for (j in 3:J){
betamean[j] <- 2*beta[j-1] - beta[j-2];
betaprec[j] <- taup;
}

### REFERENCIA PERIODE 1983
beta[1]<- 0
beta[2]~dnorm(betamean[2],betaprec[2])
for (j in 3:J){
beta[j] ~ dnorm(betamean[j],betaprec[j]);
}
taup ~ dgamma(1.0E-3,1.0E-3);
sigmap <- 1/sqrt(taup);

##############################################################
#### AGE
alphamean[1] <-0;
alphamean[2] <-0;
for (i in 3:(I)){
alphamean[i] <- 2*alpha[i-1] - alpha[i-2];
}
alphaprec[1] <- taua*1.0E-6;
alphaprec[2] <- taua*1.0E-6;
for (i in 3:I){
alphaprec[i] <- taua*1.0E-6;
}


for (i in 1:I){
alpha[i] ~ dnorm(alphamean[i],alphaprec[i]);
```

```
}

da <- 0.0001 ;
ra <- 0.0001;
taua ~ dgamma(ra,da);
sigmaa <- 1/sqrt(taua);


##############################################
### COHORT
gammamean[1] <- 0.0;
gammaprec[1] <- tauc*1.0E-6;
gammamean[2] <- 0.0;
gammaprec[2] <- tauc*1.0E-6;
for (k in 3:K){
gammamean[k] <- 2*gamma[k-1] - gamma[k-2];
gammaprec[k] <- tauc*1.0E-6
}
### REFERENCIA COHORT 6 (1933)

gamma[6]<-0
for (k in 1:5){
gamma[k] ~ dnorm(gammamean[k],gammaprec[k]);
}
for (k in 7:(K)){
gamma[k] ~ dnorm(gammamean[k],gammaprec[k]);
}

tauc ~ dgamma(1.0E-3,1.0E-3);
sigmac <- 1/sqrt(tauc);
}
```

A.4.2 Model APC2: Bashir i Estève Modifies Berzuini, Clayton, Bray et al.

Modifies Age:

```
#### AGE
alphamean[1] <- 2*alpha[2] - alpha[3];
alphamean[2] <- (2*alpha[1] + 4*alpha[3] - alpha[4])/5;
for (i in 3:(I-2)){
alphamean[i] <- (4*alpha[i-1] + 4*alpha[i+1]- alpha[i-2]
- alpha[i+2])/6;
}
alphamean[I-1] <- (2*alpha[I] + 4*alpha[I-2] - alpha[I-3])/5;

alphamean[I] <- 2*alpha[I-1] - alpha[I-2];
Nneighsa[I] <- 1;
for (i in 1:I){
alphaprec[i] <- 1.0E-6 * taua;
}
for (i in 1:I){
alpha[i] ~ dnorm(alphamean[i],alphaprec[i]);
}
da <- 0.0001
ra <- 0.0001
taua ~ dgamma(ra,da);
sigmaa <- 1/sqrt(taua);
}
```

A.4.3 Holford Method:

```
# METODE HOLFORD: Drift en el periode, Escala Residual en la COHORT,
EDAT ESCALA PRIMARI
## K: Cohorts
# M: Periodes
#I: Edats
#nu[]: mediana dels efectes cohort APC
#beta[]: mediana dels efectes periode APC
#alpha[]: mediana dels efectes edat APC

model
{

#COHORTS
for (i in 1:K)
{
nu[i]~dnorm(mu.nu[i],tau.nu)
mu.nu[i]<-mu.c+delta.c*i
#residual extraient cohort 1933 (6) i extraient drift
nu.r[i]<-(nu[i]-mu.nu[i])-nu[6]
}
mu.c~dnorm(0,0.001)
delta.c~dnorm(0,0.001)
sd.nu~dunif(0,10)
tau.nu<-pow(sd.nu,-2)


#PERIODES
for (j in 1:M)
{
beta[j]~dnorm(mu.b[j],tau.b)
#afegeixo drift a periode i trec periode referencia
beta.r[j]<-mu.b[j]+delta.c(per[j]-1985)-mu.b[1]
mu.b[j]~dnorm(0,0.001)
}
sd.b~dunif(0,10)
tau.b<-pow(sd.b,-2)

#EDATS
for (h in 1:I)
{
alpha[h]~dnorm(mu.a[h],tau.a)
#COMPLETAR FACTORS
alpha.r[i]<-mu.a[h]+mu.d+delta.c*h+delta.c*1985+beta[1]
mu.a[h]~dnorm(0,0.001)
}
tau.a~pow(sd.a,-2)
sd.a~dunif(0,10)

# RECORDATORI: REPORTAR effectes edat:alpha.r, cohort:nu.r,
periode:beta.r

}
```

Appendix A.5. Sensibility Analysis of prior distributions of precission parameters in Chapter 5 for Holford method

In order to assess sensibility to prior distributions for the precission parameters used in the Holford method it has been tried different prior distributions. The age, period and cohort parameters of a full APC model (APC1 or APC2) has been obtained, and let τ_ν be the precission of the cohort parameters (in log-scale) of equation (42), τ_β be the precission of the period parameters and τ_α be the precission of the age parameters. As it has been done in Chapter 4, Table A.5.1 shows different prior distributions which have been used for these parameters. It has been evaluated their posterior median value, standard error and Scale Reduction Factor.

Table A.5.1. Prior distributions for the precission of the parameters of the Holford method.

Same Prior for each precission parameter	Posterior for τ_ν			Posterior for τ_β			Posterior for τ_α		
	Median	sd	SRF	Median	sd	SRF	Median	sd	SRF
Gamma (0.1,0.01)	17.325	12.359	1.1528	26.362	15.632	1.258	17.673	0.337	1.156
Gamma (0.001,0.001)	12.550	15.695	1.1152	28.963	12.544	1.159	21.584	0.348	1.199
Gamma (0.01,0.01)	9.685	5.234	1.0014	25.632	11.568	1.001	12.541	0.283	1.001
Uniform (0,100)*	9.354	7.584	1.0024	24.633	12.568	1.001	11.236	0.286	1.015
Uniform (0,10)*	1.968	6.251	1.0001	3.259	4.837	1.001	1.365	3.625	1.001

* Prior distribution for the square root inverse (sd) of the precission parameters

Note that these parameters do not appear to be robust due to the dependence of their prior precission. For both τ_ν and τ_α the prior Uniform(0,10) shows a SRF close to 1, but precission parameters show high variability compared with their median value. However, for prior distributions Gamma (0.01,0.01) and Uniform(0,100), posterior median values of

precissions and their standard errors are very similar. However, precission modeled with prior Gamma(0.01,0.01) seems to converge slightly better than Uniform(0,100), and posterior precisions show small standard errors than those obtained with prior Uniform(0,100).

Appendix A.6. Age-Period-Cohort Analysis with reference cohort 1952

In this section we describe the Age, Period and Cohort effects obtained with the APC2 model. For each cause of incidence and mortality related with PLC or with liver cirrhosis it has been represented the Incidence or Mortality Rate per 100,000 person-years and the Rate Ratios of the Cohort and Period effects. In this subanalysis, we have used a different reference cohort, cohort 1952, in order to evaluate variability of interpretations when reference categories change. Drift trend has been assumed for the Period effect whereas Cohort effect is the residual scale.

Table A.6.1 shows the Annual Percent Change (APCH) and its 95% CRI for PLC, Hepatocellular carcinoma, Cholangiocarcinoma and Cirrhosis. APCH has been estimated with the Age-drift model, for that reason results of Table A.6.1 slightly differs from those of Table 5.8, calculated from the APCH estimated with a model which used Age-Adjusted rates (equation 41 Chapter 5).

Table A.6.1. Annual Percent Change (APCH) and its 95% credibility intervals (95% CRI) for LC histologies and liver cirrhosis mortality. (Age-Drift Model)

	MEN		WOMEN	
	APCH	**95% CRI**	**APCH**	**95% CRI**
INCIDENCE				
Liver Cancer	2.9	(1.85 , 3.97)	0.69	(-0.88, 2.29)
Hepatocellular Carcinoma	6.57	(4.93 , 8.24)	5.81	(2.68 , 9.04)
Cholangiocarcinoma	-0.17	(-1.64 , 1.33)	-2.42	(-4.44, -0.36)
MORTALITY				
Liver Cancer	-0.27	(-0.54 , 0.10)	-3.78	(-4.13 , -3.42)
Hepatocellular Carcinoma	7.12	(6.55 , 7.44)	5.46	(4.74 , 6.19)
Cholangiocarcinoma	16.14	(14.23 , 18.09)	15.69	(13.65 , 17.78)
Cirrhosis	-3.41	(-3.58 , -3.24)	-2.92	(-3.20 , -2.65)

Figure A.6.1. Liver Cancer Incidence in Men (155)
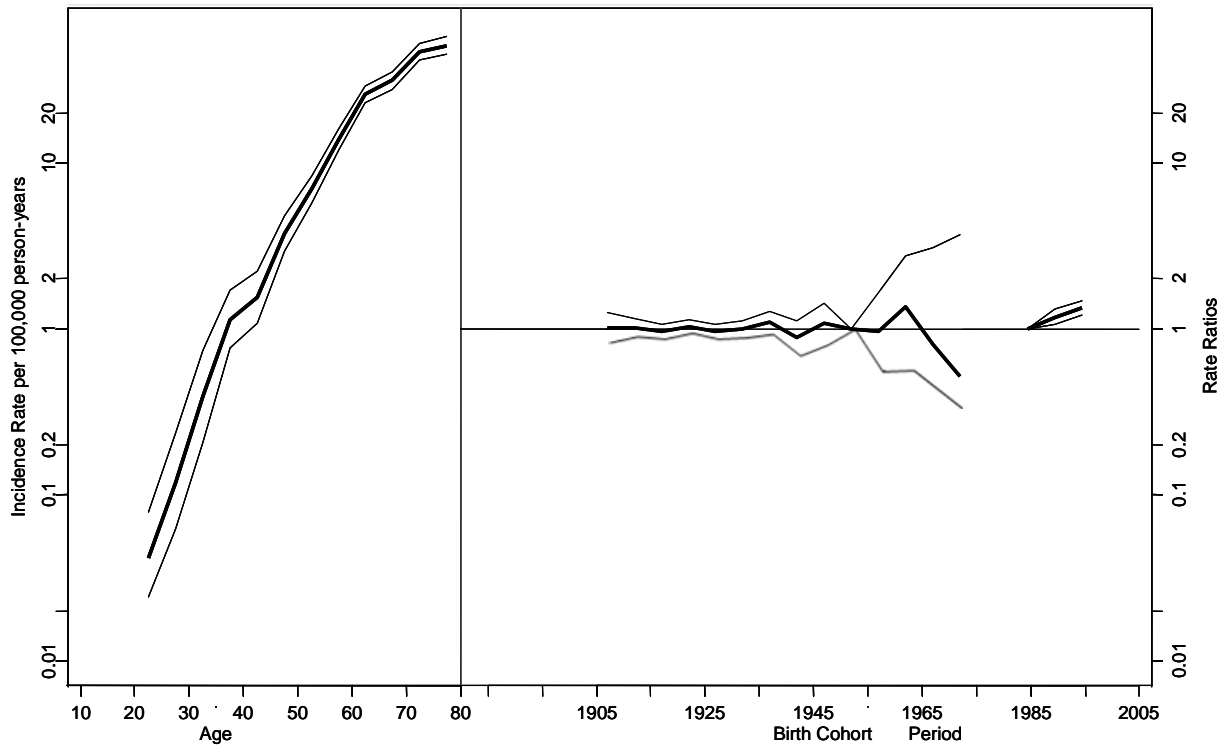

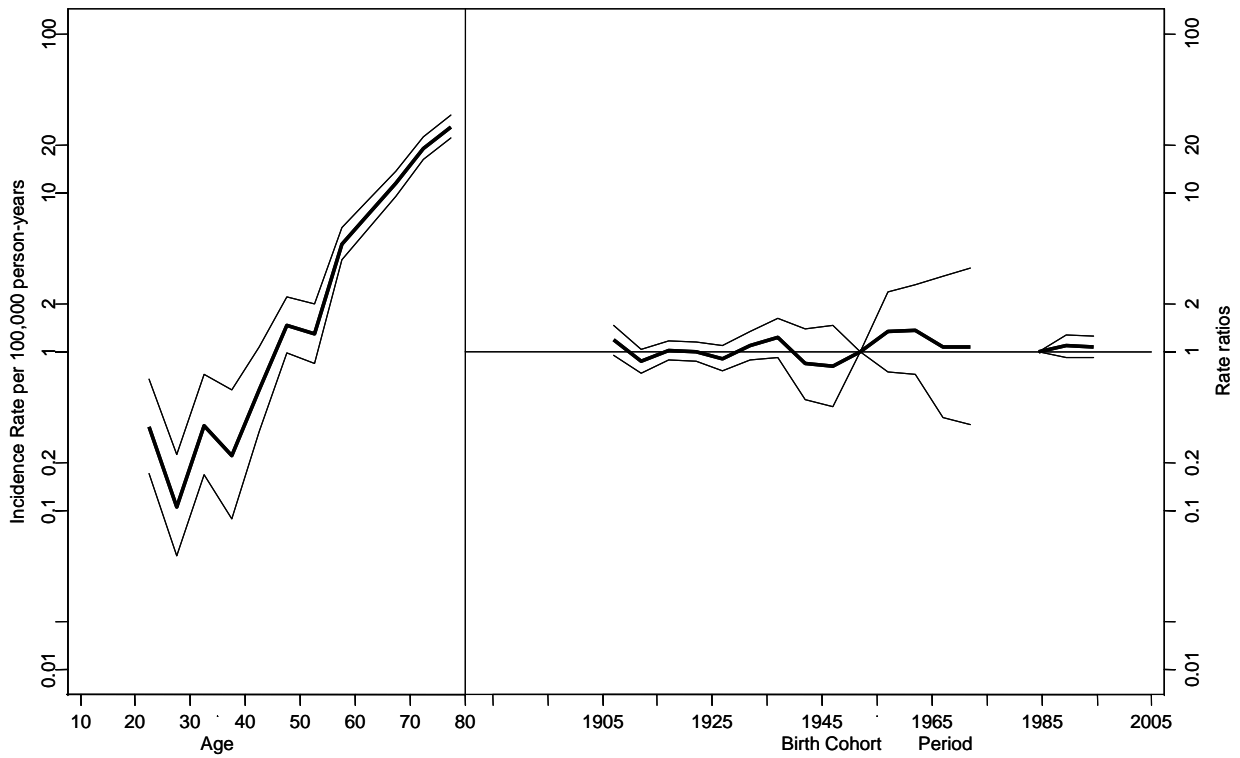**Figure A.6.2.** Liver Cancer Incidence in Women (155)

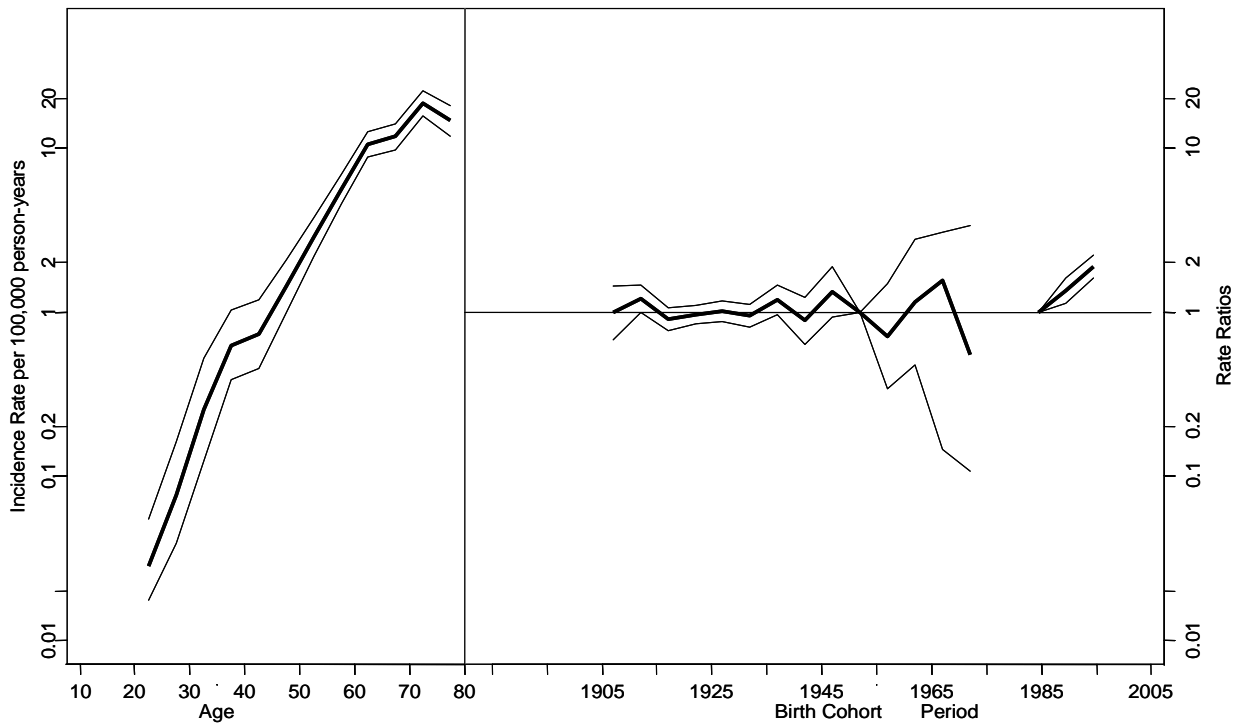**Figure A.6.3.** Hepatocellular Carcinoma Incidence in Men (1550)

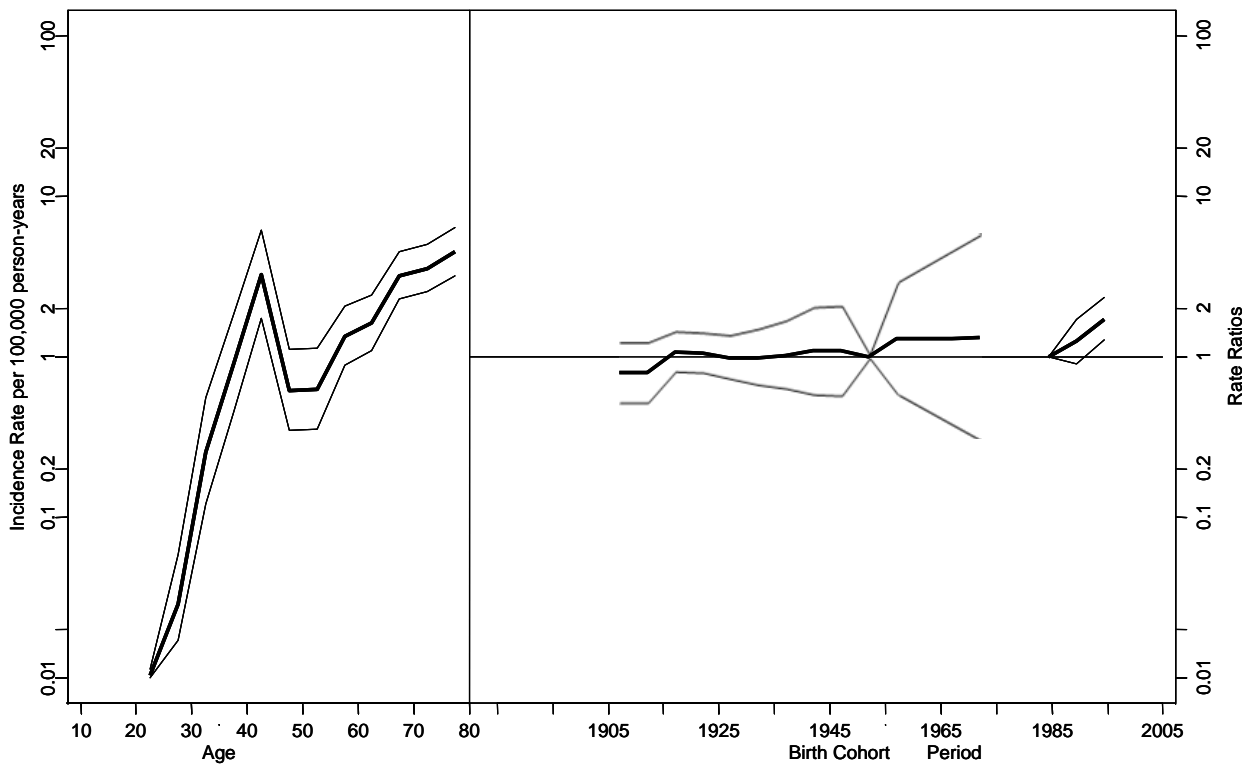**Figure A.6.4.** Hepatocellular Carcinoma Incidence in Women (1550)

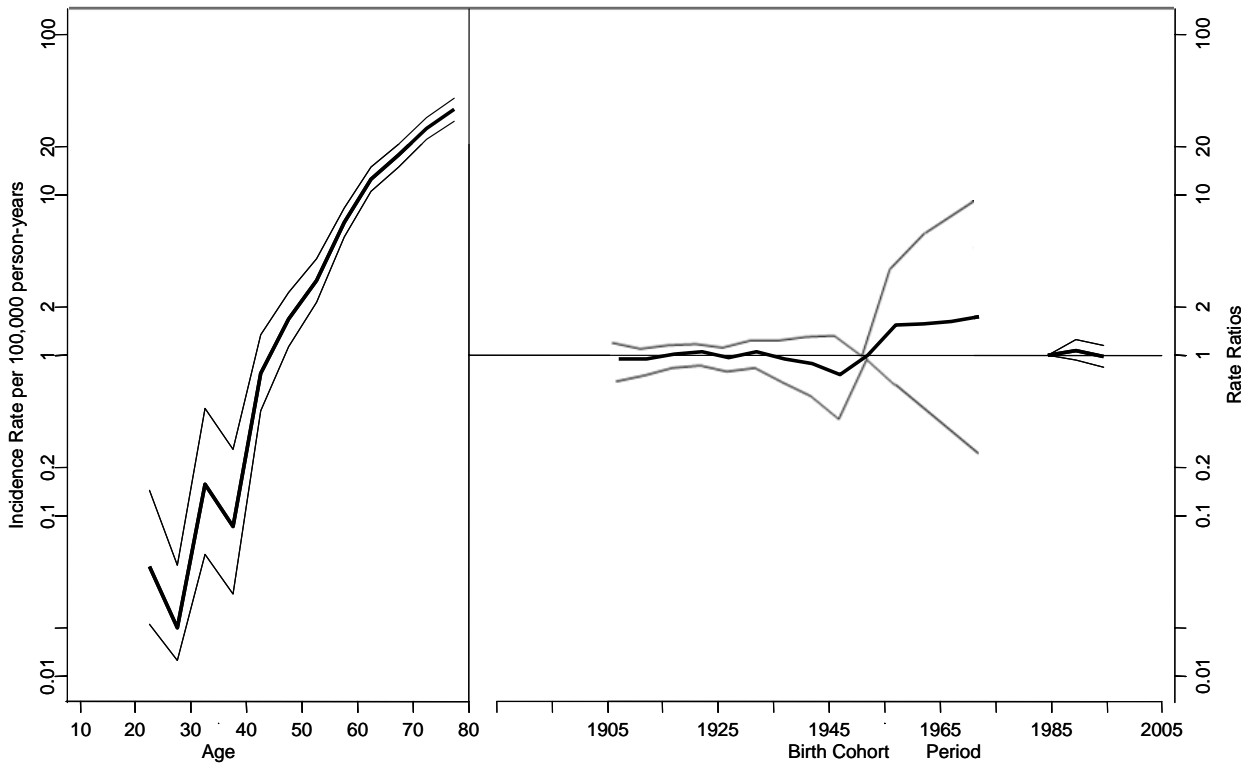**Figure A.6.5.** Cholangiocarcinoma Incidence in Men (1551)

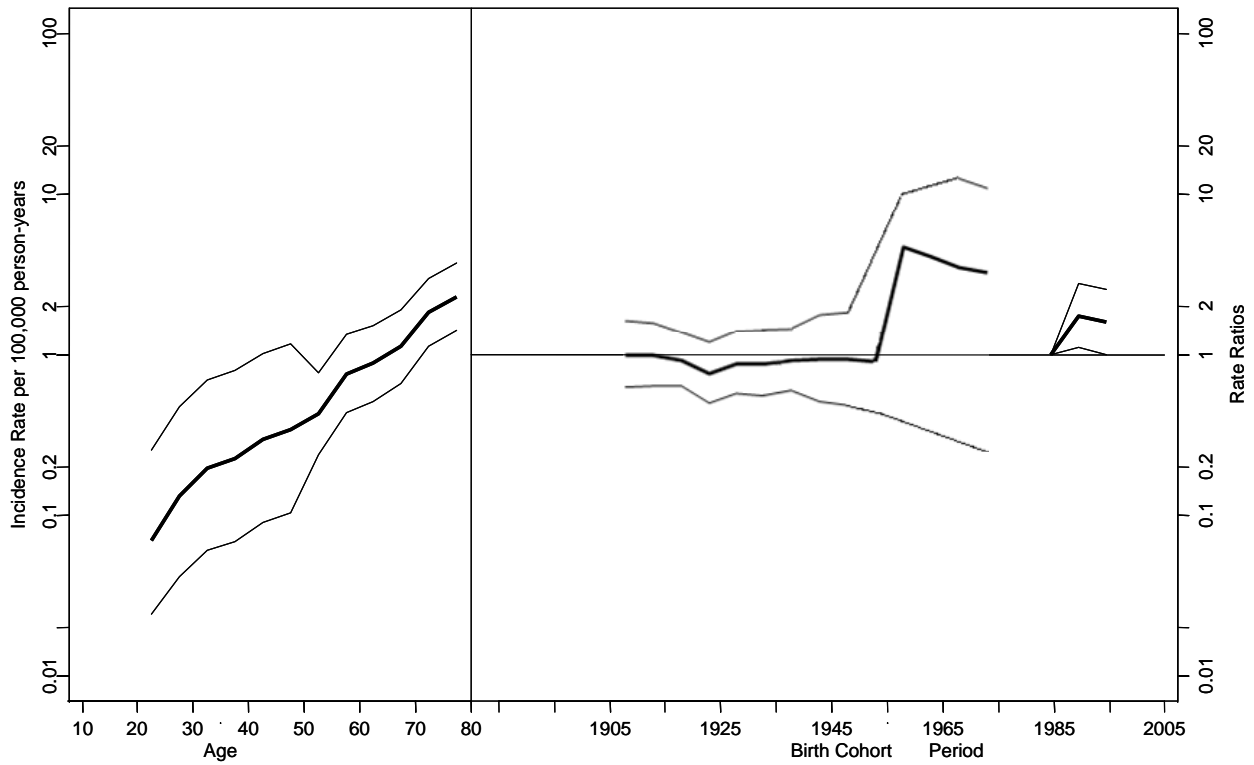**Figure A.6.6.** Cholangiocarcinoma Incidence in Women (1551)
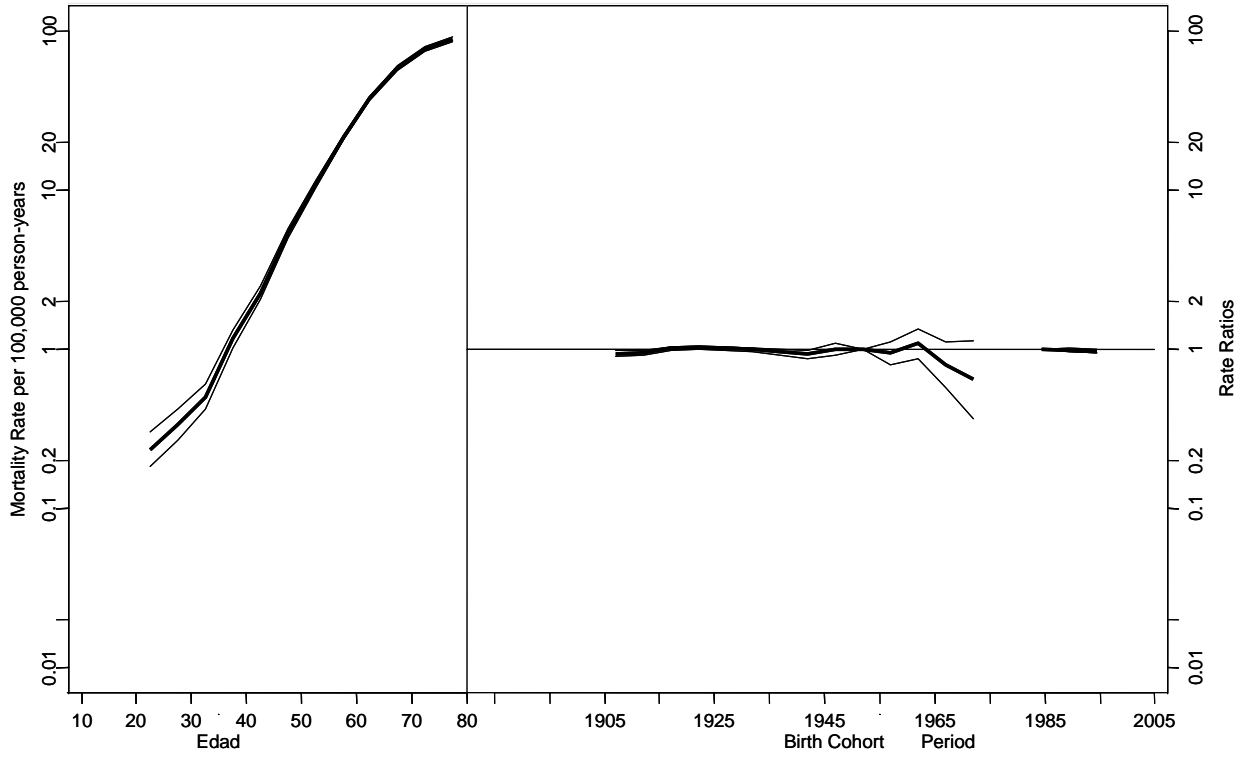
Figure A.6.7. Liver Cancer Mortality in Men (155)

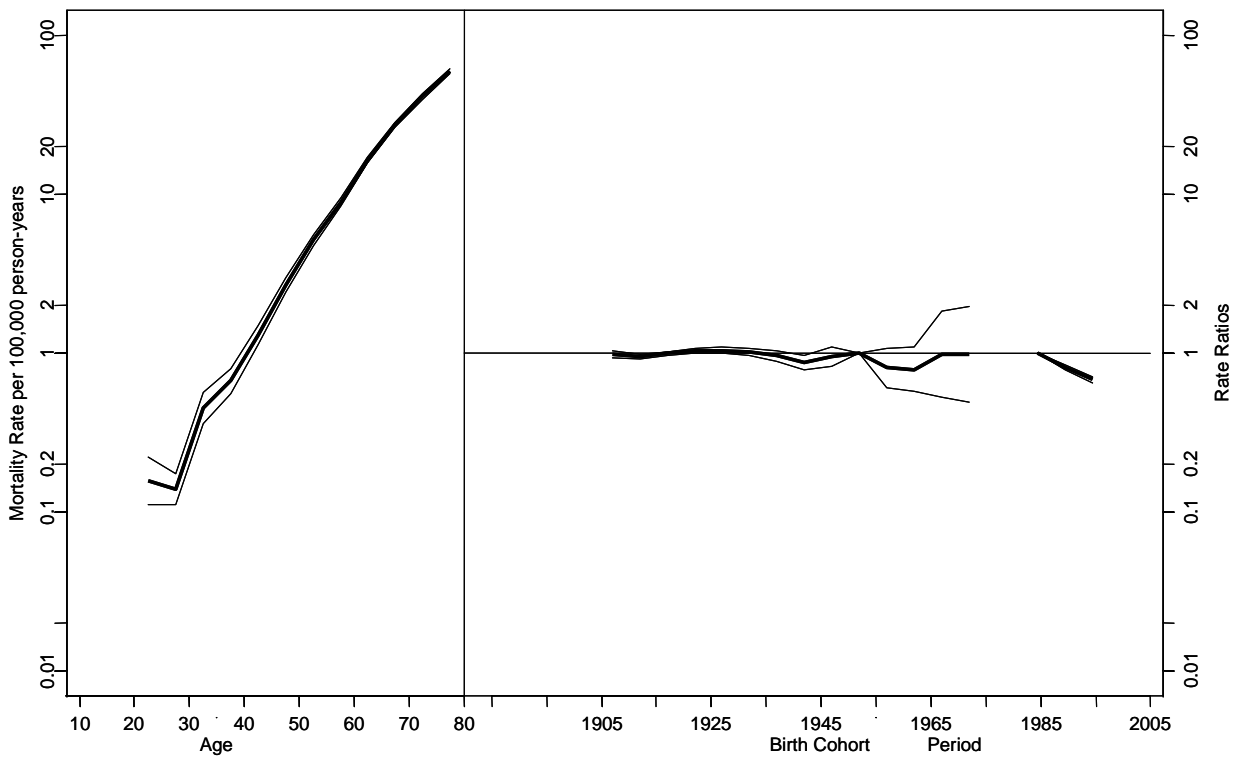**Figure A.6.8**. Liver Cancer Mortality in Women (155)

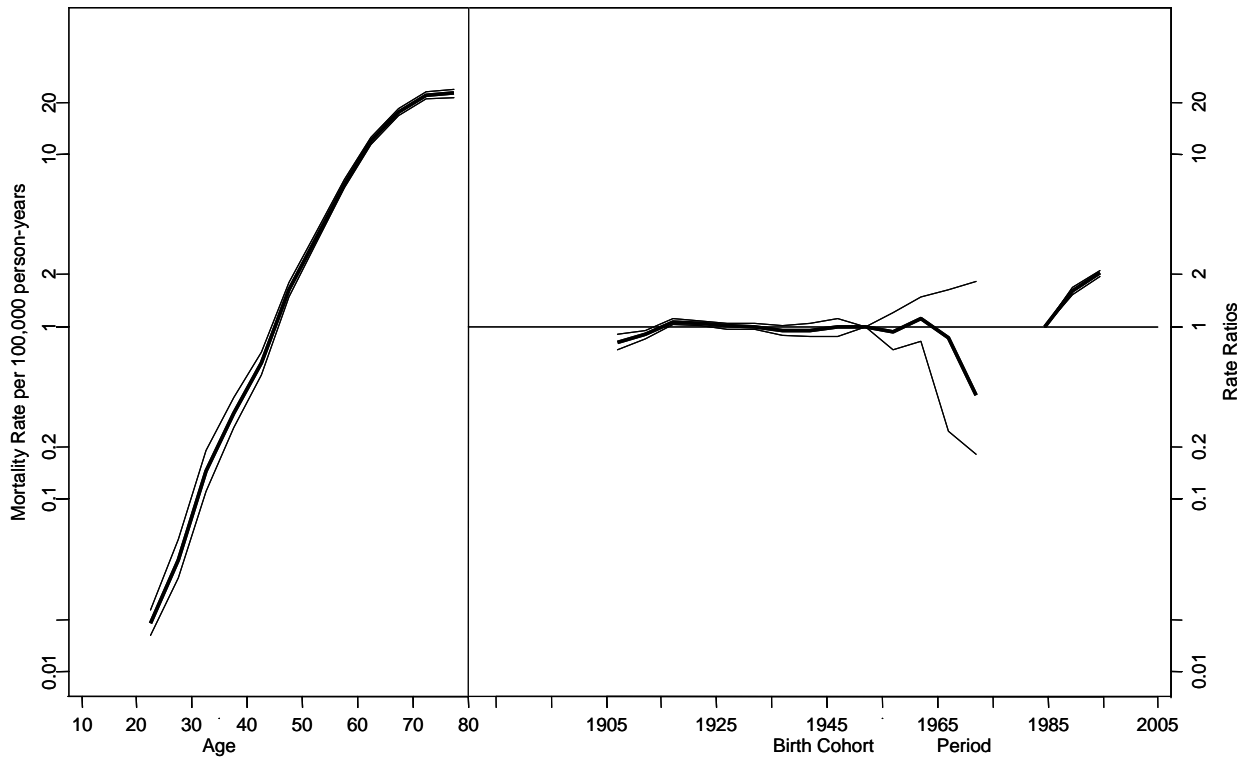**Figure A.6.9.** Hepatocellular Carcinoma Mortality in Men (1550)

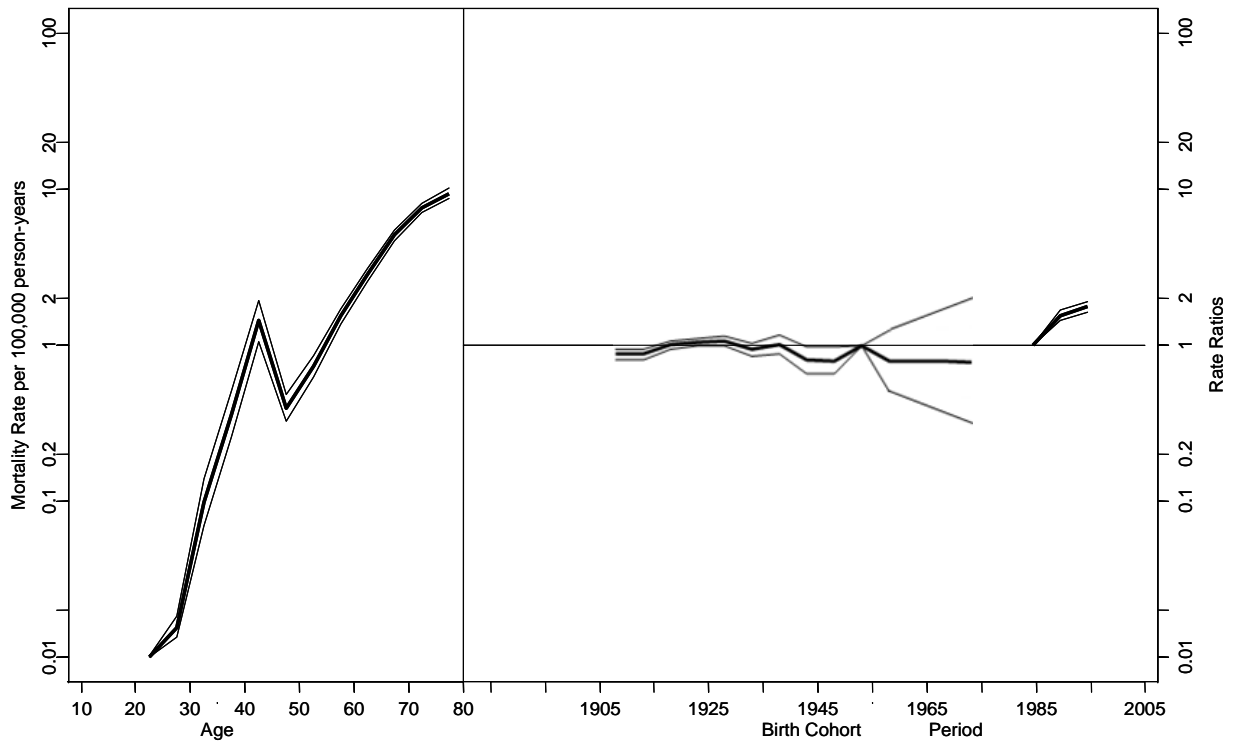**Figure A.6.10.** Hepatocellular Carcinoma Mortality in Women (1550)

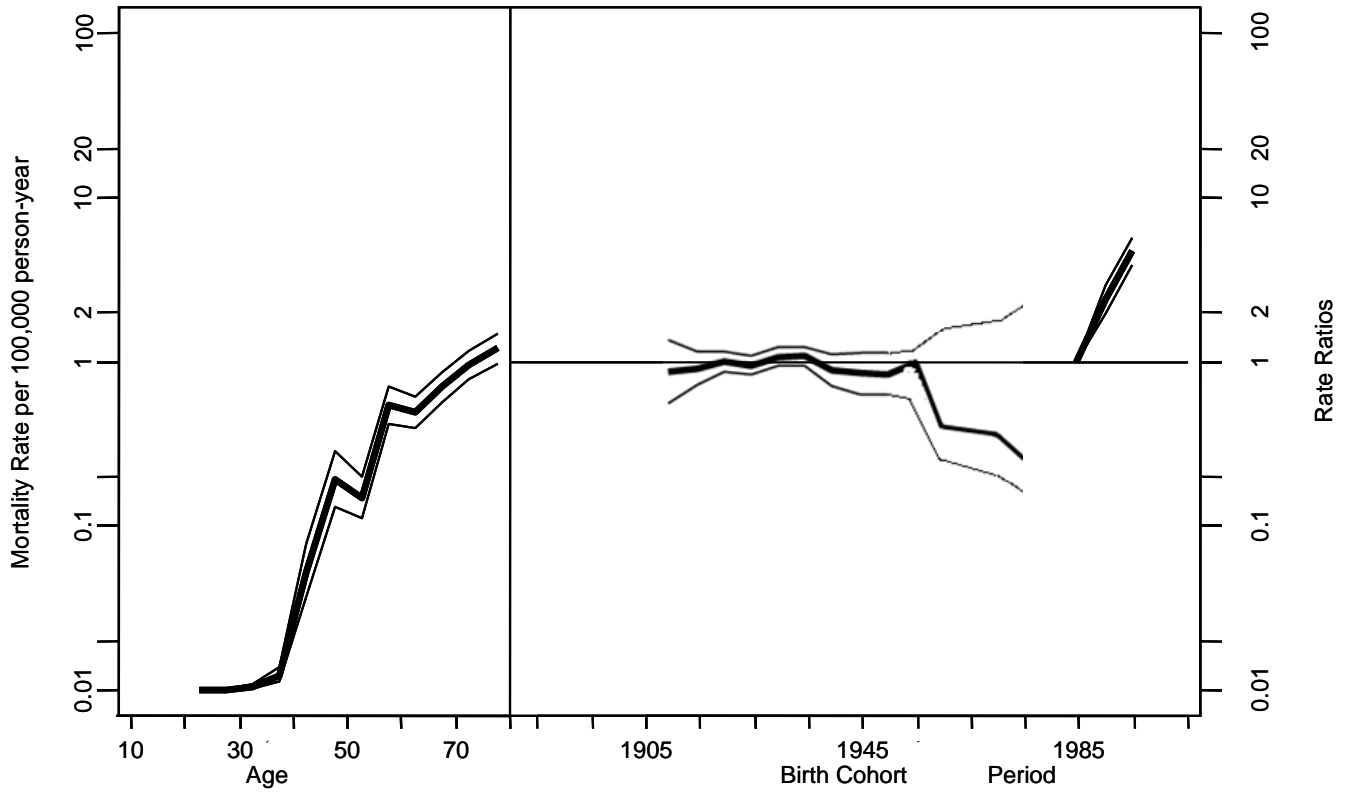**Figure A.6.11.** Cholangiocarcinoma Mortality in Men (1551)

Figure A.6.12. Cholangiocarcinoma Mortality in Women (1551)
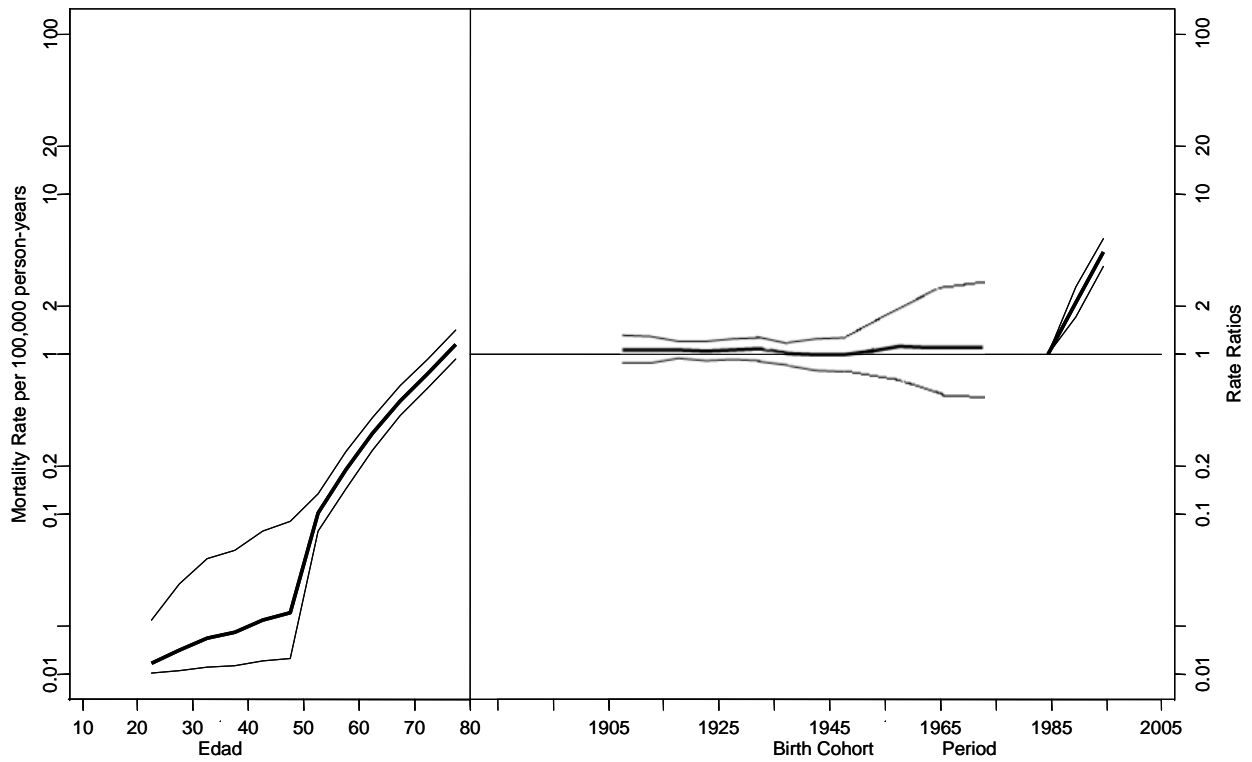
Figure A.6.13. Cirrhosis Mortality in Men (571)
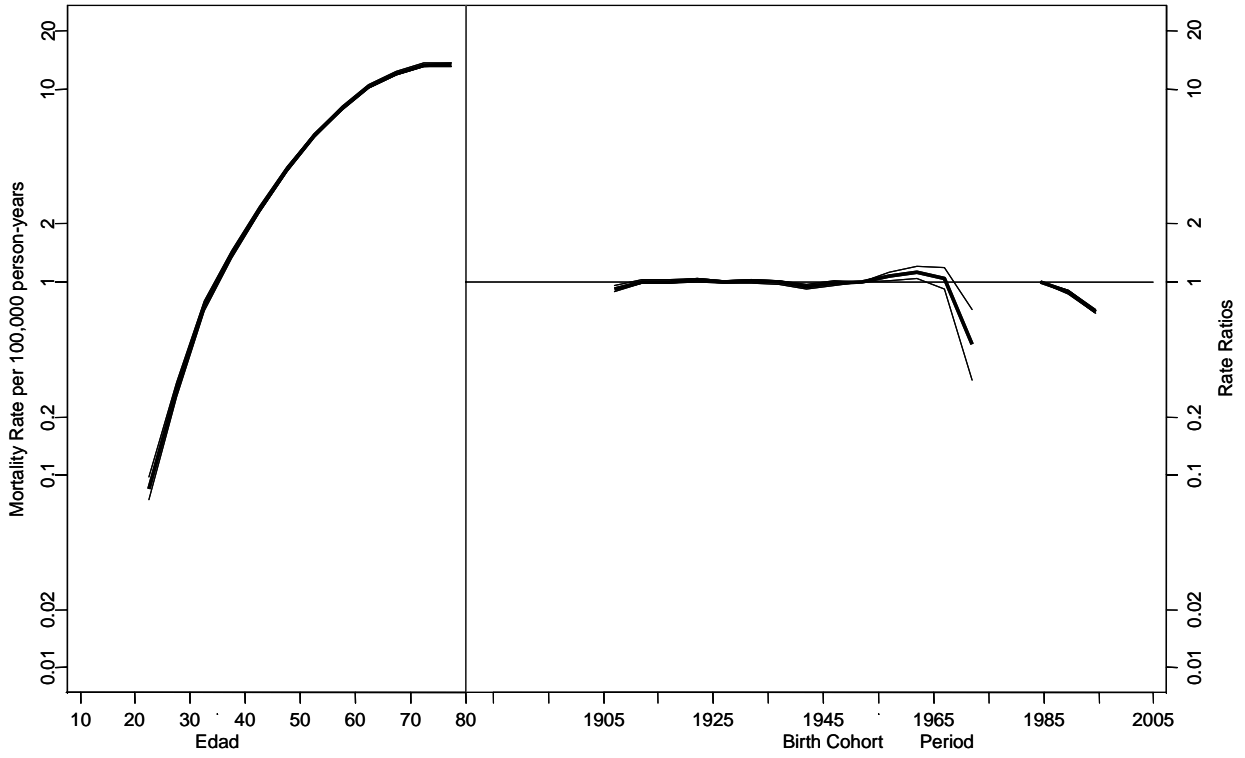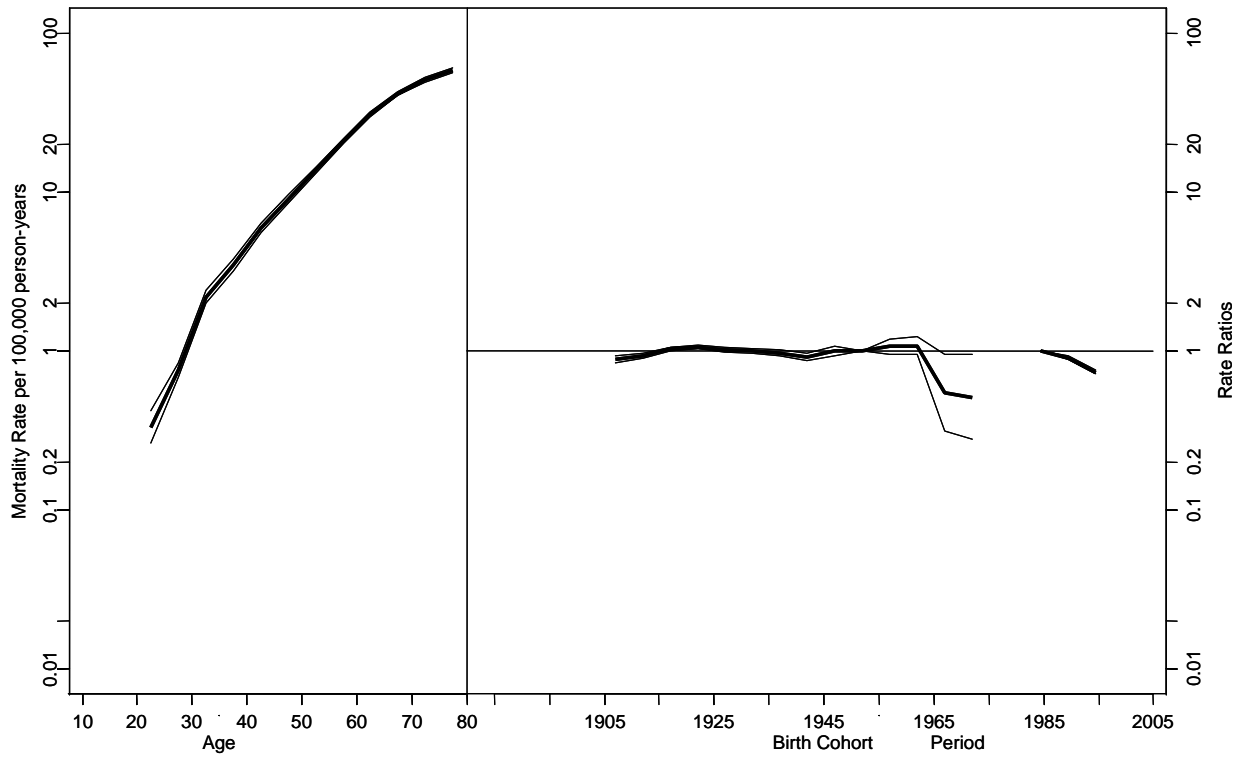

Figure A.6.14. Cirrhosis Mortality in Women (571)

Bibliography

Reference List

1. World Health Organization (WHO). Pathology and genetics of tumours of the digestive system.World Health Organisation classification of tumours series. Hamilton, S. R. and Aaltonen, L. A. 2000. Lyon, International Agency for Research on Cancer (IARC). ISBN: 9283224108.

2. Bosch, F. X., Ribes, J., Cleries, R., and Diaz, M. Epidemiology of hepatocellular carcinoma. Clin.Liver Dis. 9[2], 191-211, v. 2005.

3. IARC. IARC Monographs of the Evaluation of Carcinogenic Risk to Humans. Vol 59. Hepatitis Viruses. IARC. 1994. Lyon, International Agency for Research on Cancer.

4. Murray, P. R, Rosenthal K.S., Kobyashi G.S., and Pfaller M.A. Medical Microbiology. 5. 2005. Philadelphia, Mosby.

5. Lavanchy, D. Hepatitis B virus epidemiology, disease burden, treatment, and current and emerging prevention and control measures. J.Viral Hepat. 11[2], 97-107. 2004.

6. World Health Organization (WHO). Hepatitis B factsheet. http://www.who.int/mediacentre/factsheets/fs204/en/index.html . 2000. 1-6-2006.

7. Bosch, F. X., Ribes, J., Díaz, M., and Clèries, R. Primary Liver Cancer: Worldwide Incidence and Trends. Gastroenterology 127, S5-S16. 2004.

8. de Franchis, R., Hadengue, A., Lau, G., Lavanchy, D., Lok, A., McIntyre, N., Mele, A., Paumgartner, G., Pietrangelo, A., Rodes, J., Rosenberg, W., and Valla, D. EASL International Consensus Conference on Hepatitis B. 13-14 September, 2002 Geneva, Switzerland. Consensus statement (long version). J.Hepatol. 39 Suppl 1, S3-25. 2003.

9. Lok, A. S. and McMahon, B. J. Chronic hepatitis B: update of recommendations. Hepatology 39[3], 857-861. 2004.

10. World Health Organization (WHO). Countries using hepatitis B vaccines in their national infant immunization system. http://www.who.int/vaccines-surveillance/graphics/htmls/hepb.htm . 2005. 5-6-2006.

11. World Health Organization (WHO). Hepatitis B immunization. Introducing hepatitis B into national immunization services. www.who.int/vaccines-documents/DocsPDF01/www598.pdf . 2003. 1-6-2006.

12. World Health Organization (WHO). Hepatitis B vaccine. http://www.who.int/gpv-surv/graphics/htmls/hepb.htm . 1998. 1-1-2006.

13. World Health Organization (WHO). Geographic Patern of Hepatitis B Surface Antigen (HBsAg) Seroprevalence. http://www.who.int/vaccines-surveillance/graphics/htmls/HBsAgseroprevalence.htm . 1997. 1-6-2006.

14. World Health Organization (WHO). Hepatitis B and breastfeeding. http://www.who.int/child-adolescent-health/New_Publications/NUTRITION/updt-22.htm . 1996. 1-6-2006.

15. Shepard, C. W., Finelli, L., and Alter, M. J. Global epidemiology of hepatitis C virus infection. Lancet Infect.Dis. 5[9], 558-567. 2005.

16. World Health Organization (WHO). Hepatitis C fact sheet. http://www.who.int/mediacentre/factsheets/fs164/en/ . 2000. 1-1-2006.

17. Worman, H. J. The Hepatitis C Sourcebook. http://cpmcnet.columbia.edu/dept/gi/hepC.html . 2006. 1-6-2006.

18. World Health Organization (WHO). Hepatitis C. Weekly Epidemiological Record (WHO). 72, 65-69. 1997.

19. Jonas, M. M. Hepatitis C virus infection: clinical aspects and treatment with interferon alfa. Clin.Ther. 18 Suppl B, 110-125. 1996.

20. Ferlay, J., Bray, F., Pisani, P., and Parkin, D. M. GLOBOCAN 2002: Cancer Incidence, Mortality and Prevalence Worldwide. IARCPress IARC CancerBase. 2.0[No. 5. version]. 2004. Lyon.

21. Ribes, J., Cleries, R., Rubio, A., Hernandez, J. M., Mazzara, R., Madoz, P., Casanovas, T., Casanova, A., Gallen, M., Rodriguez, C., Moreno, V., and Bosch, F. X. Cofactors associated with liver disease mortality in an HBsAg-positive Mediterranean cohort: 20 years of follow-up. Int.J.Cancer 119:3, 687-694. 22-2-2006.

22. Hall, A. J., Winter, P. D., and Wright, R. Mortality of hepatitis B positive blood donors in England and Wales. The Lancet , 91-93. 1985.

23. Crook, P. D., Jones, M. E., and Hall, A. J. Mortality of hepatitis B surface antigen-positive blood donors in England and Wales. International Journal of Epidemiology 32, 118-124. 2003.

24. Manno, M., Cammà, C., Schepis, F., Bassi, F., Gelmini, R., Giannini, F., Miselli, F., Grottola, A., Ferretti, I., Vecchi, Ch., De Palma, M., and Villa, E. Natural history of chronic HBV carriers in Northern Italy: morbidity and mortality after 30 years. Gastroenterology 127(3), 756-63. 2004.

25. Bosch, F. X. and Ribes, J. Epidemiology of Liver Cancer in Europe. Can J Gastroenterol 14:7, 621-630. 2000.

26. McGlynn, K. A., Tsao, L., Hsing, A. W., Devesa, S. S., and Fraumeni jr, J. F. International trends and patterns of primary liver cancer. Int.J.Cancer 94, 290-296. 2001.

27. World Health Organization (WHO). Mortality Database. WHO statistical information system. http://www3.who.int/whosis/menu.cfm?path=whosis,mort . 2005. 6-1-2005.

28. Parkin, D. M., Whelan, S. L., Ferlay, J., Teppo, L., Thomas, D. B., and (eds). Cancer Incidence in Five Continents. Volume VIII. IARC Scientific Publications Nº 155. 2002. Lyon, France, International Agency for Research on Cancer; International Association of Cancer Registries.

29. Bosch, F. X., Ribes, J., Diaz, M., and Cleries, R. Primary liver cancer: Worldwide incidence and trends. Gastroenterology 127[5 Suppl 2], S5-S16. 2004.

30. Bosch, F. X., Ribes, J., and Borras, J. Epidemiology of Primary Liver Cancer. Seminars in Liver Diseases 19 (3), 271-285. 1999.

31. Bosch, F. X. Global epidemiology of hepatocellular carcinoma. Okuda, K. and Tabor, E. Liver cancer. 1[2], 13-28. 1997. London, Churchill Livingstone.

32. Donato, F., Tagger, A., Gelatti, U., Parrinello, G., Boffetta, P., Albertini, A., Decarli, A., Trevisi, P., Ribero, M. L., Martelli, C., Porru, S., and Nardi, G. Alcohol and Hepatocellular carcinoma: the effect of lifetime intake and hepatitis virus infections in men and women. American Journal of Epidemiology 155(4), 323-331. 2002.

33. Yu, M. W., Yang, Y. C, Yang, S. Y., Cheng, S. W., Liaw, Y. F., Lin, S. M., and Chen, C. J. Hormonal markers and Hepatitis B virus-related hepatocellular carcinoma risk: a nested case-control study among men. Journal of the National Cancer Institute 93(21), 1644-1651. 2001.

34. Coleman, M. P., Gatta, G., Verdecchia, A., Esteve, J., Sant, M., Storm, H., Allemani, C., Ciccolallo, L., Santaquilani, M., and Berrino, F. EUROCARE-3 summary: cancer survival in Europe at the end of the 20th century. Ann.Oncol. 14 Suppl 5, V128-V149. 2003.

35. National Cancer Institute. SEER Cancer Statistics Review, 1975-2002. Ries, L. A. G., Eisner, M. P., Kosary, C. L., Hankey, B. F., Miller, B. A., Clegg, L., Mariotto, A., Feuer, E. J., and Edwards, B. K. http://seer.cancer.gov/csr/1975_2002 . 8-1-2005. Bethesda,MD. 8-1-2005.

36. Hanley, A. J, Choi, B. C., and Holowaty, E. J. Cancer mortality among Chinese migrants: a review. International Journal of Epidemiology 24, 255-265. 1995.

37. Khlat, M., Bouchardy, C., and Parkin, D. M. Cancer mortality among near eastern migrants in Australia. Révue d'Épidémiologie et Santé Publique 41, 208-217. 1993.

38. Bouchardy, C., Parkin, D. M., and Khlat, M. Cancer mortality among Chinese and South-East Asian migrants in France. International Journal of Cancer 58, 638-643. 1994.

39. Fang, J., Madhavan, S., and Alderman, M. H. Cancer Mortality of Chinese in New York City 1988-1992. International Journal of Epidemiology 25, 907-912. 1996.

40. Grulich, A. E., Swerdlow, A. J., Head, J., and Marmot, M. G. Cancer mortality in African and Caribbean migrants to England and Wales. Br J Cancer 66, 905-911. 1992.

41. McCredie, M., Williams, S., and Coates, M. Cancer mortality in East and Southeast Asian migrants to New South Wales, Australia, 1975-1995. British Journal of Cancer 79, 1277-1282. 1999.

42. Benhamiche, A. M., Faivre, C., Minello, A., Clinard, F., Mitry, E., Hillon, P., and Faivre, J. Time trends and age-period-cohort effects on the incidence of primary liver cancer in a well-defined french population: 1976-1995. Journal of Hepatology 29, 802-806. 1998.

43. Taylor-Robinson, S. D., Foster, G. R., Arora, S., Hargreaves, S., and Thomas, H. C. Increase in primary liver cancer in the UK, 1979-94. The Lancet 350, 1142-1143. 1997.

44. Okuda, K., Fujimoto, I., Hanai, A., and Urano, Y. Changing incidence of hepatocellular carcinoma in Japan. Cancer Res 47, 4967-4972. 1987.

45. El-Serag, H. B. and Mason, A. C. Rising incidence of hepatocellular carcinoma in the United States. The New England Journal of Medicine 340, 745-750. 1999.

46. El-Serag, H. B., Davila, J. A., Petersen, N. J., and McGlynn, K. A. The continuing increase in the incidence of hepatocellular carcinoma in the United States: an update. Annals of Internal Medicine 139(10), 817-823. 2003.

47. Khan, S. A., Taylor-Robinson, S. D., Toledano, M. B., Beck, A., Elliott, P., and Thomas, H. C. Changing international trends in mortality rates for liver, biliary and pancreatic tumours. Journal of Hepatology . 2002.

48. Levi, F., Lucchini, F., Negri, E., Boyle, P., and la Vecchia, C. Cancer Mortality in Europe, 1995-1999, and overview of trends since 1960. International Journal of Cancer 110, 155-169. 2004.

49. Patel, T. Increasing Incidence and Mortality of Primary Intrahepatic Cholangiocarcinoma in the United states. Hepatology 33, 1353-1357. 2001.

50. Ribes, J., Cleries, R., Borras, J., Galceran, J., and Bosch, F. X. Time trends in incidence and mortality for chronic liver disease and liver cancer in the interval 1980-1997 in Catalonia, Spain. Eur.J.Gastroenterol.Hepatol. 16[9], 865-872. 2004.

51. Eighth revision, International Classification of Diseases, adapted for use in the United States. 1967. Washington, DC, Department of Health, Education and Welfare, Public Health Service Publication No. 1693.

52. International Classification of Diseases, 9th Revision, Clinical Modification. 1979. Washington, DC, Departament of Health and Human Services (DHHS No. (PHS) 80-1260).

53. International statistical classification of diseases and related health problems. 10th Revision. 1992. Geneve, WHO.

54. Hoel, D. G., Ron, E., Carter, R., and Mabuchi, K. Influence of Death Cerfitifact Errors on Cancer Mortality Trends. Journal of the National Cancer Institute 85, 1063-1068. 1993.

55. Llovet, J. M. and Beaugrand, M. Hepatocellular carcinoma: present status and future prospects. J.Hepatol. 38 Suppl 1, S136-S149. 2003.

56. Llovet, J. M., Burroughs, A., and Bruix, J. Hepatocellular carcinoma. The Lancet 362[9399], 1907-1917. 6-12-2003.

57. Parkin, D. M. The global health burden of infection-associated cancers in the year 2002. Int.J.Cancer 118[12], 3030-3044. 15-6-2006.

58. Prince, A. M. and Alcabes, P. The risk of development of hepatocellular carcinoma in hepatitis B virus carrier in New York. A preliminary estimate using death-records matching. Hepatology 2, 15-20. 1982.

59. Dodd, R. Y. and Nath, N. Increased risk for lethal forms of liver disease among HBsAg-positive blood donors in the United States. Journal of Virological Methods 17, 81-94. 1987.

60. Oshima, A., Tsukuma, H., Hiyama, T., Fujimoto, I., Yamamo, H., and Tanaka, M. Follow-up study of HBsAg-positive blood donors with special reference to effect of drinking and smoking on development of liver cancer. Int.J.Cancer 34, 775-779. 1984.

61. Sakuma, K., Saitoh, N., Kasai, M., Jitsukawa, H., Yoshino, I., Yamaguchi, M., Nobutomo, K., Yamumi, M., Tsuda, F., Komazawa, T., Nakamura, T., Yoshida, Y., and Okuda, K. Relative risks of death due to liver disease among Japanese male adults having various statuses for hepatitis B s and e antigen/antibody in serum: a prospective study. Hepatology 8(6), 1642-1646. 1988.

62. Sakuma, K., Takahara, T., Okuda, K., Tsuda, F., and Mayumi, M. Prognosis of Hepatitis B Virus Surface Antigen Carriers in Relation to Routine Liver Function tests: A Prospective Study. Gastroenterology 83, 114-117. 1982.

63. Iijima, T., Saitoh, N., Nobutomo, K., Nambu, M., and Sakuma, K. A prospective cohort study of hepatitis B surface antigen carriers in a working population. Gann 75, 571-573. 1984.

64. Tokudome, S., Ikeda, M., Matsushita, K., Maeda, Y., and Yoshinari, M. Hepatocellular carcinoma among female japanese hepatitis b virus carriers. Hepato-gastroenterol 34, 246-248. 1987.

65. Tokudome, S., Ikeda, M., Matsushita, K., Maeda, Y., and Yoshinari, M. Hepatocellular Carcinoma among HBsAg positive blood donors in Fukuoka, Japan. Eur J Cancer Clin Oncol 2, 235-239. 1988.

66. Beasley, R. P., Hwang, L. Y., Lin, C. C., and Chien, S. C. Hepatocellular carcinoma and hepatitis B virus: A prospective study of 22707 men in Taiwan. The Lancet ii, 1129-1133. 1981.

67. Beasley, R. P. Hepatitis B Virus. The major etiology of hepatocellular carcinoma. Cancer 61, 1942-1956. 1988.

68. Yang, H. I, Lu, S.-N., Liaw, Y.-F., You, S.-L., Sun, C.-A., Wang, L.-Y., Hsiao, C. K., Chen, P.-J., Chen, D.-S., and Chen, C.-J. Hepatitis B e antigen and the risk of hepatocellular carcinoma. New England Journal of Medicine 347(3), 168-174. 2002.

69. Tu, J. T., Gao, R. N., Zhang, D. H., and Gu, B. C. Hepatitis B virus and primary liver cancer on Chongming Island, People's Republic of China. Natl Cancer Inst Monogr 69, 213-215. 1985.

70. Yeh, F. S., Yu, M. C., Mo, Ch. Ch., Luo, S., Tong, M. J., and Henderson, B. E. Hepatitis B virus, aflatoxins, and hepatocellular carcinoma in Southern Guangxi, China. Cancer Research 49, 2506-2509. 1989.

71. Ross, R. K., Yuan, J. M., Yu, M. C., Wogan, G. N., Qian, G. S., Tu, J. T., Groopman, J. D., Gao, Y. T., and Henderson, B. E. Urinary aflatoxin biomarkers and risk of hepatocellular carcinoma. The Lancet 339[8799], 943-946. 18-4-1992.

72. Evans, A. A., Chen, G., Ross, E. A., Shen, F. M., Lin, W. Y., and London, W. T. Eight-year follow-up of the 90,000-person Haimen City cohort: I. Hepatocellular carcinoma mortality, risk factors, and gender differences. Cancer Epidemiol Biomarkers Prev 11[4], 369-376. 2002.

73. Alter, M. J. Epidemiology of hepatitis B in Europe and worldwide. Journal of Hepatology 39, S64-S69. 2003.

74. Evans, A. A., O'Connell, A. P., Pugh, J. C., Mason, W. S., Shen, F., Chen, G. C., Lin, W. Y., Dia, A., M'Boup, S., Dramé, B., and London, W. T. Geographic variations in viral load among Hepatitis B carriers with

differing risks of hepatocellular carcinoma. Cancer Epidemiology, Biomarkers & Prevention 7, 559-565. 1998.

75. Tang, B., Kruger, W. D., Chen, G., Shen, F., Lin, W. Y., Mboup, S., London, W. T., and Evans, A. A. Hepatitis B viremia is associated with increased risk of hepatocellular carcinoma in chronic carriers. J.Med.Virol. 72[1], 35-40. 2004.

76. Fattovich, G. Natural history and prognosis of hepatitis B. Semin.Liver Dis. 23[1], 47-58. 2003.

77. Abdel-Aziz, F., Habib, M., Mohamed, M. K., Abdel-Hamid, M., Gamil, F., Madkour, S., Mikhail, N. N., Thomas, D., Fix, A. D., Strickland, G. T., Anwar, W., and Sallam, I. Hepatitis C virus (HCV) infection in a community in the Nile Delta: population description and HCV prevalence. Hepatology 32[1], 111-115. 2000.

78. Yoshizawa, H. Hepatocellular carcinoma associated with hepatitis C virus infection in Japan: projection to other countries in the foreseeable future. Oncology 62(suppl 1), 8-17. 2002.

79. Wasley, A. and Alter, M. J. Epidemiology of hepatitis C: geographic differences and temporal trends. Semin Liver Dis 20, 1-16. 2000.

80. Frank, C., Mohamed, M. K., Strickland, G. T., Lavanchy, D., Arthur, R. R., Magder, L. S., El-Khoby, T., Abdel Wahab, Y., Aly Ohn, E. S., Anwar, W., and Sallam, I. The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. The Lancet 355, 887-891. 2000.

81. Okuda, K. Hepatitis C and Hepatocellular Carcinoma. Okuda, K. and Tabor, E. Liver Cancer. [4], 39-50. 1997. New York, Churchill Livingstone.

82. El-Serag, H. B. and Mason, A. C. Risk factors for the rising rates of primary liver cancer in the United States. Archives of Internal Medicine 160, 3227-3230. 2000.

83. Brechot, C., Jaffredo, F., Lagorce, D., Gerken, G., Meyer zum, Buschenfelde K., Papakonstontinou, A., Hadziyannis, S., Romeo, R., Colombo, M., Rodes, J., Bruix, J., Williams, R., and Naoumov, N. Impact of HBV, HCV and GBV-C/HGV on hepatocellular carcinomas in Europe: results of a European concerted action. J Hepatol. 29[2], 173-183. 1998.

84. Donato, F., Boffetta, P., and Puoti, M. A meta-analysis of epidemiological studies on the combined effect of hepatitis B and C virus infections in causing hepatocellular carcinoma. Int.J.Cancer 75, 347-354. 1998.

85. Kao, J. H., Chen, P. J., Lai, M. Y., and Chen, D. S. Basal core promoter mutations of hepatitis B virus increase the risk of hepatocellular carcinoma in hepatitis B carriers. Gastroenterology 124[2], 327-334. 2003.

86. Chu, C. J., Keeffe, E. B., Han, S. H., Perrillo, R. P., Min, A. D., Soldevila-Pico, C., Carey, W., Brown, R. S., Jr., Luketic, V. A., Terrault, N., and Lok, A. S. Prevalence of HBV precore/core promoter variants in the United States. Hepatology 38[3], 619-628. 2003.

87. Henry, S. H. and Bosch, F. X. Foodborne disease and mycotoxin epidemiology. Hui, Y. H., Smith, R. A., and Spoerke jr, D. G. Foodborne diseases handbook, vol 3: plant toxicants. Second Edition, revised and expanded[14], 593-626. 2000. New York . Bassel, Marcel Dekker, Inc.

88. di Bisceglie, A. M., Carithers, R. L., and Gores, G. J. Hepatocellular carcinoma. Hepatalogy 28, 1161-1165. 1998.

89. Donato, F., Tagger, A., Chiesa, R., Ribero, M., Tomasoni, V., Fasola, M., Gelatti, U., Portera, G., Boffetta, P., and Nardi, G. Hepatitis B and C virus infection, Alcohol drinking, and Hepatocellular carcinoma: a case- control study in Italy. Hepatology 26,3, 579-584. 1997.

90. Kuper, H., Tzonou, A., Kaklamani, E., Hsieh, C. C., Lagiou, P., Adami, H. O., Trichopoulos, D., and Stuver, S. O. Tobbaco smoking, alcohol consumption and their interaction in the causation of hepatocellular carcinoma. Int.J.Cancer 85(4), 498-502. 2000.

91. Chen, Z. M., Liu, B. Q., Boreham, J., Wu, Y. P., Chen, J. S., and Peto, R. Smoking and liver cancer in China: case-control comparison of 36,000 liver cancer deaths vs. 17,000 cirrhosis deaths. Int.J.Cancer 107(1), 106-112. 2003.

92. Wang, L. Y., You, S. L., Lu, S. N., Ho, H. C., Wu, M. H., Sun, C. A., Yang, H. I., and Chien-Jen, C. Risk of hepatocellular carcinoma and habits of alcohol drinking, betel quid chewing and cigarette smoking: a cohort of 2416 HBsAg-seropositive and 9421 HBsAg-seronegative male residents in Taiwan. Cancer Causes & Control 14(3), 241-250. 2003.

93. la Vecchia, C., Tavani, A., Franceschi, S., and Parazzini, F. Oral contraceptives and cancer. Drug Safety 14, 260-272. 1996.

94. Kew, M. C., Song, E., and Mohammed, A. Contraceptive steroids as a risk factor for hepatocellular carcinoma: a case control study in south African black. Hepatology 11, 298-302. 1990.

95. Mant, J. W. F. and Vessey, M. P. Trends in mortality from primary liver cancer in England and Wales 1975-92: influence of oral contraceptives. British Journal of Cancer 72, 800-803. 1995.

96. Waetjen, L. E. and Grimes, D. A. Oral contraceptives and primary liver cancer: temporal trends in three countries. Obstetrics & Gynecology 88, 945-949. 1996.

97. Beral, V., Hermon, C., Kay, C., Hannaford, P., Darby, S., and Reeves, G. Mortality associated with oral contraceptive use: 25 year follow up of cohort of 46 000 women from Royal College of General Practitioners' oral contraception study [see comments]. BMJ. 318, 96-100. 9-1-1999.

98. The Collaborative MILTS Project Team. Oral contraceptives and liver cancer. Results of the Multicentre International Liver Tumor Study (MILTS). Contraception 56[5], 275-284. 1997.

99. Yu, M. W., Chang, H. C., Chang, S. C., Liaw, Y. F., Lin, S. M., Liu, C. J., Lee, S. D., Lin, C. L., Chen, P. J., Lin, S. C., and Chen, C. J. Role of reproductive factors in hepatocellular carcinoma: impact on hepatitis B- and C-related risk. Hepatology 38(6), 1393-1400. 2003.

100. Yu, M. W., Chiang, Y. Ch., Lien, J. P., and Chen, C. J. Plasma antioxidant vitamins, chronic hepatitis B virus infection and urinary aflatoxin B1-DNA adducts in healthy males. Carcinogenesis 18, 1189-1149. 2001.

101. Inoue, M., Yoshimi, I., Sobue, T., and Tsugane, S. Influence of coffee drinking on subsequent risk of hepatocellular carcinoma: a prospective study in Japan. J.Natl.Cancer Inst. 97[4], 293-300. 16-2-2005.

102. Shimazu, T., Tsubono, Y., Kuriyama, S., Ohmori, K., Koizumi, Y., Nishino, Y., Shibuya, D., and Tsuji, I. Coffee consumption and the risk of primary liver cancer: pooled analysis of two prospective studies in Japan. Int.J.Cancer 116[1], 150-154. 10-8-2005.

103. Gelatti, U., Covolo, L., Franceschini, M., Pirali, F., Tagger, A., Ribero, M. L., Trevisi, P., Martelli, C., Nardi, G., and Donato, F. Coffee consumption reduces the risk of hepatocellular carcinoma independently of its aetiology: a case-control study. J.Hepatol. 42[4], 528-534. 2005.

104. Kurozawa, Y., Ogimoto, I., Shibata, A., Nose, T., Yoshimura, T., Suzuki, H., Sakata, R., Fujita, Y., Ichikawa, S., Iwai, N., and Tamakoshi, A. Coffee and risk of death from hepatocellular carcinoma in a large cohort study in Japan. Br.J.Cancer 93[5], 607-610. 5-9-2005.

105. Sharp, G. B., Lagarde, F., Mizuno, T., Sauvaget, C., Fukuhara, T., Allen, N., Suzuki, G., and Tokuoka, S. Relationship of hepatocellular carcinoma to soya food consumption: a cohort-based, case-control study in Japan. Int.J.Cancer 115[2], 290-295. 10-6-2005.

106. Casanovas-Taltavull, T., Ribes, J., Berrozpe, A., Jordan, S., Casanova, A., Sancho, C., Valls, C., and Bosch, F. X. Patient with hepatocellular carcinoma related to prior acute arsenic intoxication and occult HBV: epidemiological, clinical and therapeutic results after 14 years of follow-up. World J.Gastroenterol. 12[12], 1972-1974. 28-3-2006.

107. Haddow, J. E., Palomaki, G. E., McClain, M., and Craig, W. Hereditary haemochromatosis and hepatocellular carcinoma in males: a strategy for estimating the potential for primary prevention. J.Med.Screen. 10[1], 11-13. 2003.

108. Olynyk, J. K., Trinder, D., and Milward, E. Hereditary haemochromatosis: to screen or not to screen? J.Hepatol. 43[1], 9-10. 2005.

109. Kumagi, T., Horiike, N., Abe, M., Kurose, K., Iuchi, H., Masumoto, T., Joko, K., Akbar, S. F., Michitaka, K., and Onji, M. Small hepatocellular carcinoma associated with Wilson's disease. Intern.Med. 44[5], 439-443. 2005.

110. Kumagi, T., Horiike, N., Michitaka, K., Hasebe, A., Kawai, K., Tokumoto, Y., Nakanishi, S., Furukawa, S., Hiasa, Y., Matsui, H., Kurose, K., Matsuura, B., and Onji, M. Recent clinical features of Wilson's disease with hepatic presentation. J.Gastroenterol. 39[12], 1165-1169. 2004.

111. Lagiou, P., Kuper, H., Stuver, S. O., Tzonou, A., Trichopoulos, D., and Adami, H. O. Role of Diabetes Mellitus in the etiology of hepatocellular carcinoma. Journal of the National Cancer Institute 92(13), 1096-1099. 2000.

112. Wideroff, L., Gridley, G., Mellemkjaer, L., Chow, W. H., Linet, M., Keehn, S., Borch-Johnsen, K., and Olsen, J. H. Cancer Incidence in a population-based cohort of patients hospitalized with diabetes mellitus in Denmark. Journal of the National Cancer Institute 89, 1360-1365. 1997.

113. Adami, H. O., Chow, W. H., Nyrén, O., Berne, Ch., Linet, M. S., Ekbom, A., Wolk, A., McLaughlin, J. K., and Fraumeni jr, J. F. Excess risk of primary liver cancer in patients with diabetes mellitus. Journal of the National Cancer Institute 88, 1472-1477. 1996.

114. El Serag, H. B., Tran, T., and Everhart, J. E. Diabetes increases the risk of chronic liver disease and hepatocellular carcinoma. Gastroenterology 126[2], 460-468. 2004.

115. Muto, Y., Sato, S., Watanabe, A., Moriwaki, H., Suzuki, K., Kato, A., Kato, M., Nakamura, T., Higuchi, K., Nishiguchi, S., Kumada, H., and Ohashi, Y. Overweight and obesity increase the risk for liver cancer in patients with liver cirrhosis and long-term oral supplementation with branched-chain amino acid granules inhibits liver carcinogenesis in heavier patients with liver cirrhosis. Hepatol.Res. 29-5-2006.

116. Sorensen, H. T., Mellemkjaer, L., Jepsen, P., thulstrup, A. M., Baron, J., Olsen, J. H., and Vilstrup, H. Risk of cancer in patients hospitalized with fatty liver: a Danish cohort study. J.Clin.Gastroenterol. 36[4], 356-359. 2003.

117. Whittle, H. C., Maine, N., Pilkington, J., Mendy, M., Fortuin, M., Bunn, J., Allison, L., Howard, C., and Hall, A. Long-term efficacy of continuing hepatitis B vaccination in infancy in two Gambian villages. The Lancet 345, 1089-1092. 1995.

118. Hsu, H. M., Chen, D. S., Chuang, C. H., Lu, J. C. F., Two, D. M., Lee, C. C., Lu, H. C., Cheng, S. H., Wang, Y. F., Wang Chen, C., Lo, K. J., Shih, C. J., Sung, J. L., and The hepatitis control committee, The hepatitis steering committee. Efficacy of a mass hepatitis B vaccination program in Taiwan. Journal of the American Medical Association 260, 2231-2235. 1988.

119. Chang, M. H., Chen, C. J., Lai, M. S., Hsu, H. M., Wu, T. C., Kong, M. S., Liang, D. C., Shau, W. Y., Chen, D. S., and for the Taiwan Childhood Hepatoma Study Group. Universal Hepatitis B vaccination in Taiwan and the incidence of Hepatocellular carcinoma in children. The New England Journal of Medicine 336, 1855-1907. 1997.

120. Lee, M. S., Kim, D. W., Lee, H. S., Kim, Ch. Y., Park, T. S., Yoo, K. Y., Park, B. J., and Ahn, Y. O. Hepatitis B vaccination and reduced risk of primary liver cancer among male adults: a cohort study in Korea. International Journal of Epidemiology 27, 316-319. 1998.

121. Tabor, E., Yu, M-Y W., Hewlett, I., and Epstein, J. S. Summary of a workshop on the implementation of NAT to screen donors of blood and plasma for viruses. Transfusion 40, 1273-1275. 2000.

122. Morales Suarez-Varela, M. M., LlopisGonzalez, L. L., and Tejerizo Perez, M. L. Mortality rates of liver cancer in Spain: effect of age, birth cohort, and period of death. J.Environ.Pathol.Toxicol.Oncol. 13[1], 43-47. 1994.

123. la Vecchia, C., Lucchini, F., Franceschi, S., Negri, E., and Levi, F. Trends in mortality from primary liver cancer in Europe. European Journal of Cancer 36, 909-915. 2000.

124. Blumberg, B. S., Alter, H. J., and Visnich, S. A "new" antigen in leukemia sera. Journal of the American Medical Association 191, 101-106. 1965.

125. Dos Santos-Silva, I. Epidemiología del Cáncer: Principios y Métodos. 1999. Lyon, Agencia Internacional de Investigación sobre el Cáncer.

126. Ijima, T., Saitoh, N., Nobutomo, K., Mambu, M., and Sakuma, K. A prospective cohort study of hepatitis B surface antigen carriers in a working population. Gann 75, 571-573. 1984.

127. Mori, M., Hara, M., Wada, I., Yamamoto, K., Honda, M., and Naramoto, J. Prospective study of hepatitis B and C viral infections, cigarette smoking, alcohol consumption and other factors associated with hepatocellular carcinoma risk in Japan. American Journal of Epidemiology 151, 131-139. 2000.

128. McMahon, B. J., Alberts, S. R., Wainwright, R. B., Bulkow, L., and Lanier, A. P. Prospective study in 1400 hepatitis B surface antigen-positive Alaska native carriers. Archives of Internal Medicine 150, 1051-1054. 1990.

129. Nomura, A., Stemmermann, G. N., and Wasnich, R. D. Presence of hepatitis B surface antigen before primary hepatocellular carcinoma. Journal of the American Medical Association 247[16], 2247-2249. 23-4-1982.

130. Beasley, R. P. and Hwang, L. Y. Overview of the epidemiology of hepatocellular carcinoma. Hollinger, F. B., Lemon, S. M., and Margolis, H. S. Viral Hepatitis and Liver Disease. 532-535. 1991. Baltimore, Williams & Wilkins.

131. Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldon, T. A., and Song, F. Methods for Meta-Analysis in Medical Research. 2002. Chichester, John Wiley & Sons, LTD.

132. Song, F., Sheldon, T. A., Sutton, A. J., Abrams, K. R., and Jones, D. R. Methods for exploring heterogeneity in meta-analysis. Eval.Health Prof. 24[2], 126-151. 2001.

133. Sutton, A. J., Abrams, K. R., and Jones, D. R. An illustrated guide to the methods of meta-analysis. J.Eval.Clin.Pract. 7[2], 135-148. 2001.

134. Breslow, N. E. and Day, N. E. Statistical methods in Cancer research. Vol II. The design and analysis of cohort studies. Davis, W. 1987. Lyon, IARC Scientific Publications Nº 32.

135. McCullagh, P. and Nelder, J. Generalized linear models. 1989. London, Chapman & Hall.

136. Bohning, D. Computer Assisted Analysis of Mixtures and Applications. Meta-Analysis, Disease Mapping and Others. Chapman & Hall, CRC. 1999.

137. Gelman, A., Carlin, J. B., Stern, H., and Rubin, D. B. Bayesian Data Analysis. 2nd. 2003. London, Chapman & Hall.

138. Bernardo J.M. and Smith A.F.M. Bayesian Theory. 1st. 1994. Chichester (UK), Wiley.

139. Gilks, W. R, Richardson, S., and Spiegelhalter, D. J. Markov Chain Monte Carlo in Practice. 1996. London, Chapman & Hall.

140. Spiegelhalter, D. J., Thomas, A., Best, N., and Lunn, D. WinBUGS User Manual, version 1.4. MRC Biostatistics Unit. 2002. Cambridge (England).

141. R Development Core Team. R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org . 2006.

142. Congdon, P. Bayesian Statistical Modelling. 2001. Wiley.

143. Congdon, P. Applied Bayesian Modelling. 2003. Wiley.

144. Ishwaran, H. and James, L. F. Gibbs sampling methods for stick-breaking priors. J Amer Stat Assoc 96, 161-173. 2001.

145. Sethuraman, J. A constructive definition of the Dirichlet process prior. Statistica Sinica 2, 639-650. 1994.

146. Viallefont, V., Richardson, S., and Green, P. J. Bayesian analysis of poisson mixtures. Nonparametr Stat 59, 181-202. 2002.

147. Banerjee, S., Carlin, B. R., and Gelfand A.E. Hierarchical modeling and analysis for spatial data. 2004. Chapman & Hall /CRC.

148. Lawson A.B, Browne W.J., and Vidal-Rodeiro C. Disease mapping with WinBUGS and MLwiN. 2003. Wiley.

149. Richardson, S. and Green, P. J. On Bayesian analysis of mixtures with an unknown number of components (with discussion). J R Stat Soc Ser A-Stat Soc. 59, 731-792. 1997.

150. Ross, R. K., Yuan, J. M., Yu, M. W., Wogan, G. N., Tu, J. T., Groopman, J. D., Gao, Y. T., and Henderson, B. E. Urinary aflatoxin biomarkers and risk of hepatocellular carcinoma. The Lancet 339, 943-946. 1992.

151. Qian, C., Camps, J., Maluenda, M. D., Civeira, M. P., and Prieto, J. Replication of hepatitis C virus in peripheral blood mononuclear cells. Effect of alpha-interferon therapy. Journal of Hepatology 16, 380-383. 1992.

152. Ming, L., Thorgeirsson, S. S., Gail, M. H., Lu, P., Harris, C. C., Wang, N., Shao, Y., Wu, Z., Liu, G., Wang, X., and Sun, Z. Dominant role of hepatitis B virus and cofactor role of aflatoxin in hepatocarcinogenesis in Qidong, China. Hepatology 36[5], 1214-1220. 2002.

153. Spiegelman, D., Schneeweiss, S., and McDermott, A. Measurement error correction for logistic regression models with an "alloyed gold standard". American Journal of Epidemiology 145, 184-196. 15-1-1997.

154. Monson, R. R. Occupational Epidemiology. 54-59. 1980. Boca Raton, FL, CRC Press.

155. Monson, R. R. Observations on the healthy worker effect. J.Occup.Med. 28[6], 425-433. 1986.

156. Parkin, D. M. The global health burden of infection-associated cancers in the year 2002. Int.J.Cancer . 10-1-2006.

157. Ribes J, Clèries R., and Bosch, F. X. Epidemiología del cáncer hepático en Europa. Gastroenterología y Hepatología Continuada 1 (6), 263-272. 2002.

158. World Health Organization (WHO). Hepatitis C - global prevalence (Update). Wkly Epidemiol Rec 75, 18-19. 2001.

159. Ferlay, J., Bray, F., Pisani, P, and Parkin, D. M. GLOBOCAN 2000: Cancer incidence, mortality and prevalence worldwide, Version 1.0. IARC Cancer Base No. 5. 2001. Lyon, IARC Press.

160. Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde A.J. Bayesian Measures of Model Complexity and Fit (with Discussion). J Roy Stat Soc, Series B 64 (3), 583-639. 2002.

161. English, D. Geographical epidemiology and ecological studies. Elliot, P., Cuzick, J., English, D., and Stern, R. Geographical and environmental epidemiology: methods for small area studies. 3-13. 1992. Oxford University Press.

162. Gelman, A. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 1, 1-19. 2005.

163. Besag, J. Spatial interaction and the statistical analysis of lattice systems. J Roy Stat Soc, Series B 36, 192-236. 1974.

164. Clayton, D. G. and Kaldor, J. M. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics 43, 671-682. 1987.

165. Bernardinelli, L., Clayton, D., and Montomoli, C. Bayesian estimates of disease maps: how important are priors? Stat.Med. 14[21-22], 2411-2431. 15-11-1995.

166. Besag, J., York J.C., and Mollié, A. Bayesian image restoration, with two applications in spatial satatistics (with discussion). Ann.Inst.Statist.Math. 43, 1-59. 1991.

167. Scali, J., Richard, A., and Gerber, M. Diet profiles in a population sample from Mediterranean southern France. Public Health Nutr. 4[2], 173-182. 2001.

168. Esteve, J., Benhamou, E., and Raymond, L. Statistical methods in cancer research. Volume IV. Descriptive epidemiology. IARC Sci.Publ. [128], 1-302. 1994.

169. Rosenbaum, P. R. and Rubin, D. B. Difficulties with regression analyses of age-adjusted rates. Biometrics 40[2], 437-443. 1984.

170. Biggeri, A., Marchi, M., Lagazio, C., Martuzzi, M., and Bohning, D. Non-parametric maximum likelihood estimators for disease mapping. Stat Med. 19[17-18], 2539-2554. 15-9-2000.

171. Held, L., Natario, I., Fenton, S. E., Rue, H., and Becker, N. Towards joint disease mapping. Stat.Methods Med.Res. 14[1], 61-82. 2005.

172. Jin, X., Carlin, B. P., and Banerjee, S. Generalized hierarchical multivariate CAR models for areal data. Biometrics 61[4], 950-961. 2005.

173. Stroffolini, T., Andreone, P., Andriulli, A., Ascione, A., Craxi, A., Chiaramonte, M., Galante, D., Manghisi, G., Mazzanti, R., Medaglia, C., Pilleri, G., Rapaccini.G.L., Simonetti, R. G., Taliani, G., Tosti, M. E., Villa, E., and Gasbarrini, G. Characteristics of hepatocellular carcinoma in Italy. Jounal of Hepatology 29, 944-952. 1998.

174. Stroffolini, T. Etiological factor of hepatocellular carcinoma in Italy. Minerva Gastroenterol.Dietol. 51[1], 1-5. 2005.

175. Kuper, H. E., Tzonou, A., Kaklamani, E., Hadziyannis, S., Tasopoulos, N., Lagiou, P., Trichopoulos, D., and Stuver, S. Hepatitis B and C viruses in the etiology of hepatocellular carcinoma; a study in Greece using third-generation assays. Cancer, Causes & Control 11, 171-175. 2000.

176. Raptis, I., Koskinas, J., Emmanouil, T., and Hadziyannis, S. Changing relative roles of hepatitis B and C viruses in the aetiology of hepatocellular carcinoma in Greece. Epidemiological and clinical observations. J.Viral Hepat. 10[6], 450-454. 2003.

177. Sypsa, V., Touloumi, G., Tassopoulos, N. C., Ketikoglou, I., Vafiadis, I., Hatzis, G., Tsantoulas, D., Akriviadis, E., Delladetsima, J., Demonakou, M., and Hatzakis, A. Reconstructing and predicting the hepatitis C virus epidemic in Greece: increasing trends of cirrhosis and hepatocellular carcinoma despite the decline in incidence of HCV infection. J.Viral Hepat. 11[4], 366-374. 2004.

178. Sypsa, V., Touloumi, G., Papatheodoridis, G. V., Tassopoulos, N. C., Ketikoglou, I., Vafiadis, I., Hatzis, G., Tsantoulas, D., Akriviadis, E., Koutsounas, S., and Hatzakis, A. Future trends of HCV-related cirrhosis and hepatocellular carcinoma under the currently available treatments. J.Viral Hepat. 12[5], 543-550. 2005.

179. Garcia-Torres, M. L., Zaragoza, A., Giner, R., Primo, J., and del Olmo, J. A. Incidence and epidemiological factors of hepatocellular carcinoma in Valencia during the year 2000. Rev.Esp.Enferm.Dig. 95[6], 385-4. 2003.

180. Rodriguez-Vidigal, F. F., Baz, M. J., Romero, J., and del Puerto, M. [Epidemiology of hepatocellular carcinoma in a rural area. Role of hepatotrophic viruses on survival.]. An.Med.Interna 22[4], 162-166. 2005.

181. Deuffic, S., Buffat, L., Poynard, T., and Valleron, A-J. Modeling the Hepatitis C virus Epidemic in France. Hepatology 29, 1596-1601. 1999.

182. Schoniger-Hekele, M., Muller, C., Kutilek, M., Oesterreicher, C., Ferenci, P., and Gangl, A. Hepatocellular carcinoma in Austria: aetiological and clinical characteristics at presentation. Eur.J.Gastroenterol.Hepatol. 12[8], 941-948. 2000.

183. Rabe, C., Pilz, T., Klostermann, C., Berna, M., Schild, H. H., Sauerbruch, T., and Caselmann, W. H. Clinical characteristics and outcome of a cohort of 101 patients with hepatocellular carcinoma. World J.Gastroenterol. 7[2], 208-215. 2001.

184. Widell, A., Verbaan, H., Wejstal, R., Kaczynski, J., Kidd-Ljunggren, K., and Wallesrstedt, S. Hepatocellular carcinoma in Sweden: its association with viral hepatitis, especially with hepatitis C viral genotypes. Scandinavian Journal of Infectious Diaseases 32, 147-152. 2000.

185. Verhoef, C., Visser, O., de Man, R. A., de Wilt, J. H., IJzermans, J. N., and Janssen-Heijnen, M. L. Hepatocellular carcinoma in the Netherlands incidence, treatment and survival patterns. Eur.J.Cancer 40[10], 1530-1538. 2004.

186. Grgicevic, D., Balija, M., Pirc-Tiljak, D., Mihaljevic, I., Gjenero-Margan, I., Zupancic-Salek, S., and Macek, P. Decreasing risk of viral transfusion-transmitted diseases in Croatia. Croat.Med.J. 41[2], 191-196. 2000.

187. Ivan, A., Azoicai, D., Grigorescu, R., Pisica, G., Groll, M., Trifan, M., Barhala, M., Mitroi, I., Chiriac, L., Scripcaru, D., and . [Seroepidemiological considerations of the incidence of viral hepatitis C in

some population groups]. Rev.Med.Chir Soc.Med.Nat.Iasi 99[3-4], 129-138. 1995.

188. Vladutiu, D. S., Cosa, A., Neamtu, A., State, D., Braila, M., Gherman, M., Patiu, I. M., and Dulau-Florea, I. Infections with hepatitis B and C viruses in patients on maintenance dialysis in Romania and in former communist countries: yellow spots on a blank map? J.Viral Hepat. 7[4], 313-319. 2000.

189. Kostritsa, S. S., Kuzin, S. N., Khukhlovich, P. A., Shakhgil'dian, I. V., Doroshenko, N. V., Blinov, V. M., Loparen, V. N., Sukhanova, L. L., Gol'tsov, V. A., Drobeniuk, ZhA, and . [The detection of antibodies to the hepatitis C virus and to HBsAg in donors with viral hepatitis and cancerous diseases in the Republic of Moldova]. Vopr.Virusol. 39[2], 93-94. 1994.

190. Atanasova, M. V., Haydouchka, I. A., Zlatev, S. P., Stoilova, Y. D., Iliev, Y. T., and Mateva, N. G. Prevalence of antibodies against hepatitis C virus and hepatitis B coinfection in healthy population in Bulgaria. A seroepidemiological study. Minerva Gastroenterol.Dietol. 50[1], 89-96. 2004.

191. Khoronzhevskaia-Muliar, I. S., Martyniuk, G. A., Shakhgil'dian, I. V., Schastnyi, E. I., Kuzin, S. N., Sharlai, M. M., Koval'chuk, I. S., Kharitaniuk, R. A., and Kucheruk, E. F. [Spread of infection with hepatitis B and C viruses in different population groups of north-western Ukraine]. Zh.Mikrobiol.Epidemiol.Immunobiol. [5], 35-39. 2000.

192. Rehm, J., Room, R., Graham, K., Monteiro, M., Gmel, G., and Sempos, C. T. The relationship of average volume of alcohol consumption and patterns of drinking to burden of disease: an overview. Addiction 98[9], 1209-1228. 2003.

193. Taylor-Robinson, D., Thomas, H. C., Arora, S., and Hargreaves, S. Increased mortality from liver cancer in England and Wales is not related to hepatitis C. BMJ 319, 640. 1999.

194. El-Serag, H. B. Hepatocellular Carcinoma and Hepatitis C in the United States. Hepatology 36, S74-S83. 2003.

195. Deuffic, S., Poynard, T., Buffat, L., and Valleron, A. J. Trends in primary liver cancer (letter). The Lancet 351, 214-215. 1998.

196. Suárez, A., Viejo, G., Navascués, C. A., García, R., Díaz, G., Saro, C., and Román, F. J. Prevalencia de marcadores frente a los virus A,B y C de la hepatitis en la población de Gijón entre 26 y 65 años de edad. Gastroenterol Hepatol 20, 347-352. 1997.

197. Garcia-Fulgueiras, A., Tormo, M. J., Rodriguez, T., Perez-Flores, D., Chirlaque, D., and Navarro, C. Prevalence of hepatitis B and C markers in the south-east of Spain: an unlinked community-based serosurvey of 2,203 adults. Scand.J.Infect.Dis. 28[1], 17-20. 1996.

198. Solà, R., Cruz de Castro, E., Hombrados, M., Planas, R., Coll, S., Jardí, R., Sunyer, J., Covas, I., Marrugat, J., and S.C.D. Prevalencia de las hepatitis B y C en diversas comarcas de Cataluña: estudio transversal. Medicina Clínica 119(3), 90-95. 2002.

199. Calvet, X., Bruix, J., Bru, C., Gines, P., Vilana, R., Sole, M., Ayuso, M. C., Bruguera, M., and Rodes, J. Natural history of hepatocellular carcinoma in Spain. Five year's experience in 249 cases. Journal of Hepatology 10(3), 311-317. 1990.

200. Sharp, G. B., Cologne, J. B., Fukuhara, T., Itakura, H., Yamamoto, M., and Tokuoka, S. Temporal changes in liver cancer incidence rates in Japan: Accounting for death certificate inaccuracies and improving diagnostic techniques. International Journal of Cancer 93(5), 751-758. 2001.

201. IARC. CI5 I-VIII: Updated Database. Cancer Mondial Statistics Information System. IARC . 2005. 15-10-2005.

202. Holford, T. R. The estimation of age, period and cohort effects for vital rates. Biometrics 39, 311-324. 1983.

203. Robertson, C., Gandini, S., and Boyle, P. Age-period-cohort models: a comparative study of available methodologies. J.Clin.Epidemiol. 52[6], 569-583. 1999.

204. Holford, T. R. Understanding the effects of age, period, and cohort on incidence and mortality rates. Annu.Rev.Public Health 12, 425-457. 1991.

205. Holford, T. R. Approaches to fitting age-period-cohort models with unequal intervals. Stat.Med. 5-9-2005.

206. Clayton, D. and Schifflers, E. Models for temporal variation in cancer rates. I: age-period and age-cohort models. Statistics in Medicine 6, 449-467. 1987.

207. Clayton, D. and Schifflers, E. Models for temporal variation in cancer rates. II: Age-period-cohort models. Stat.Med. 6[4], 469-481. 1987.

208. Holford, T. R. Analysing the temporal effects of age, period and cohort. Stat.Methods Med.Res. 1[3], 317-337. 1992.

209. Osmond, C. and Gardner, M. J. Age, period and cohort models applied to cancer mortality rates. Stat.Med. 1[3], 245-259. 1982.

210. Tango, T. and Kurashina, S. Age, period and cohort analysis of trends in mortality from major diseases in Japan, 1955 to 1979: peculiarity of the cohort born in the early Showa Era. Stat.Med. 6[6], 709-726. 1987.

211. Tarone, R. E. and Chu, K. C. Evaluation of birth cohort patterns in population disease rates. Am.J.Epidemiol. 143[1], 85-91. 1-1-1996.

212. Decarli, A. and la Vecchia, C. Age, period and cohort models: a review of knowledge and implementation in GLIM. Riv Stat Applic 20, 397-410. 1987.

213. Robertson, C. and Boyle, P. Age, period and cohort models: the use of individual records. Stat.Med. 5[5], 527-538. 1986.

214. Breslow, N. E. and Clayton, D. G. Approximate inference in generalized linear mixed models. J Am Statist Ass 88, 9-25. 1993.

215. Berzuini, C. and Clayton, D. Bayesian analysis of survival on multiple time scales. Stat Med. 13[8], 823-838. 30-4-1994.

216. Bray, I., Brennan, P., and Boffetta, P. Recent trends and future projections of lymphoid neoplasms--a Bayesian age-period-cohort analysis. Cancer Causes and Control 12[9], 813-820. 2001.

217. Bray I. Application of Markov chain Monte Carlo methods to projecting cancer incidence and mortality. Appl Statist 51, 151-164. 2002.

218. Baker, A. and Bray, I. Bayesian projections: what are the effects of excluding data from younger age groups? Am J Epidemiol. 162[8], 798-805. 15-10-2005.

219. Bashir, S. A. and Estève, J. Projecting cancer incidence and mortality using Bayesian age-period.cohort models. J Epidemiol Bioestat 6, 287-296. 2001.

220. Besag, J. E, Green, P. J., Higdon, D. M., and Mengersen, K. L. Bayesian computation and stochastic systems (with discussion). Statis Sci 10, 3-66. 1995.

221. Knorr-Held, L. and Rainer, E. Projections of lung cancer mortality in West Germany: a case study in Bayesian prediction. Biostatistics. 2[1], 109-129. 2001.

222. Holford, T. R. An alternative approach to statistical age-period-cohort analysis. J.Chronic.Dis. 38[10], 831-840. 1985.

223. Puig, P. and Valero, J. Count data distributions: some characterizations with applications. J Am Stat Assoc 101, 332-340. 2006.

224. Corrao, G., Rubbiati, L., Zambon, A., and Aricò, S. Alcohol-attributable and alcohol-preventable mortality in Italy. A balance in 1983 and 1996. European Journal of Public Health 12, 214-223. 2002.

225. Singh, G. K. and Hoyert, D. L. Social epidemiology of chronic liver disease and cirrhosis mortality in the United States, 1935-1997: trends and differentials by ethnicity, socioeconomic status, and alcohol consumption. Hum Biol 72 (5), 801-820. 2000.

226. Hurwitz, E. S., Holman, R. C., Strine, T. W., and Chorba, T. L. Chronic liver disease mortality in the United States, 1979 through 1989. American Journal of Public Health 85 (9), 1256-1260. 1995.

227. Halliday, M. L., Coates, R. A., and Rankin, J. G. Changing trends of cirrhosis mortality in Ontario, Canada, 1911-1986. International Journal of Epidemiology 20 (1), 199-208. 1991.

228. Trends in factors that may cause or prevent cancer. Swerdlow, A., Dos Santos-Silva, I., and Doll, R. Cancer incidence and mortality in England and Wales. Trends and Risk Factors. First[5], 19-70. 2001. Oxford, Oxford University Press.

229. Vass, A. Rates of liver cirrhosis rise in England, fall in Europe. British Medical Journal 323(7326), 1388. 2001.

230. Leon, D. A. and McCambridge, J. Liver cirrhosis mortality rates in Britain from 1950 to 2002: an analysis of routine data. The Lancet 367[9504], 52-56. 7-1-2006.

231. la Vecchia, C., Levi, F., Lucchini, F., Franceschi, S., and Negri, E. Worldwide patterns and trends in mortality from liver cirrhosis, 1955 to 1990. Annals of Epidemiology 4(6), 480-486. 1998.

232. Noble, J. A., Caces, M. F., steffens, R. A., and Stinson, F. S. Cirrhosis hospitalization and mortality trends, 1970-87. Public Health Rep 108 (2), 192-197. 1993.

233. Miranda, B., Canon, J., Cuende, N., Garrido, G., Naya, M. T., and Fernandez-Zincke, E. Disparities in access to liver transplantation in Spain. Transplantation 76[9], 1398-1403. 15-11-2003.

234. Cuende, Melero N., Miranda, Serrano B., Canon Campos, J. F., Naya, Nieto T., and Garrido, Cantarero G. [Priority criteria for the access to transplantation. The case of liver transplant in Spain]. Med.Clin.(Barc.) 120[10], 380-386. 22-3-2003.

235. Yoon, Y. H., Yi, H., and Hilton, M. E. Surveillance Report #70: Liver Cirrhosis Mortality in the United States, 1970-2002. Bethesda, MD: NIAAA, Division of Epidemiology and Prevention Research, Alcohol Epidemiologic Data System, August 2005, e-pub @. NIH . 2006. 23-3-2006.

236. Rodriguez-Artalejo, F., de Andres, Manzano B., Guallar-Castillon, P., Banegas, Banegas, Jr., and del Rey, Calero J. Association of moderate consumption of alcohol with rates of heavy drinking and abstinence in Spain. Alcohol Clin.Exp.Res. 23[9], 1502-1506. 1999.

237. Rodriguez-Artalejo, F., de Andres, Manzano B., Banegas, Banegas, Jr., Guallar-Castillon, P., Villar, Alvarez F., and del Rey, Calero J. [The decline in moderate alcohol consumption has been associated with a decrease in heavy drinkers in Spain in the 1987-1993 period]. Med.Clin.(Barc.) 113[10], 371-373. 2-10-1999.

238. Benach J., Yasui Y., Rosa E., Pasarín MI., Benach, N., Español, E., Martínez JM., and Daponte A. Atlas of Mortality in Small Areas in Spain (1987-1995). 2001. Barcelona, Universitat Pompeu Fabra/MSD.

239. Ministerio de Cultura. Juventud y droga en España. Ministerio de Cultura. 1980.

240. Amela, C., del Amo, J., Bruguera, M., Castilla, J., Echevarría, J. M., Estban-Mur, J. I., and et al. Coinfection by HIV and hepatitis A, B and C virus in adult patients. Review and GESIDA/PNS recommendations. Practice guidelines for the management of HIV infection (2000-2002). GESIDA Consensus Conference [González-García,J.; Guerra,L.]. 2003.

241. Garcia-Samaniego, J., Rodriguez, M., Berenguer, J., Rodriguez-Rosado, R., Carbo, J., Asensi, V., and Soriano, V. Hepatocellular carcinoma in HIV-infected patients with chronic hepatitis C. Am.J.Gastroenterol. 96[1], 179-183. 2001.

242. Graham, C. S., Baden, L. R., Yu, E., Mrus, J. M., Carnie, J., Heeren, T., and Kozjel, M. J. Influence of human immunodeficiency virus infection on the course of hepatitis C virus infection: a meta-analysis. Clin Infect Dis 33, 562-569. 2001.

243. Thio, C. L., Seaberg, E. C., Skolasky, R., Jr., Phair, J., Visscher, B., Munoz, A., and Thomas, D. L. HIV-1, hepatitis B virus, and risk of liver-related mortality in the Multicenter Cohort Study (MACS). The Lancet 360[9349], 1921-1926. 14-12-2002.

244. Liew, C. T., Li, H. M., Lo, K. W., Leow, C. K., Lau, W. Y., Hin, L. Y., Lim, B. K., San Lai, P. B. Chan Y. H., Wang, X. Q., Wu, S., and Lee, J. J. Frequent allelic loss on chromosome 9 in hepatocellular carcinoma. International Journal of Cancer 81, 319-324. 1999.

245. Shibuya, K. and Yano, E. Regression analysis of trends in mortality from hepatocellular carcinoma in Japan, 1972-2001. Int.J.Epidemiol. 34[2], 397-402. 2005.

246. Davila, J. A., Morgan, R. O., Shaib, Y., McGlynn, K. A., and El Serag, H. B. Hepatitis C infection and the increasing incidence of hepatocellular carcinoma: a population-based study. Gastroenterology 127[5], 1372-1380. 2004.

247. Kim, W. R., Gores, G. J., Benson, J. T., Therneau, T. M., and Melton, L. J., III. Mortality and hospital utilization for hepatocellular carcinoma in the United States. Gastroenterology 129[2], 486-493. 2005.

248. Dominguez, A., Bruguera, M., Vidal, J., Plans, P., and Salleras, L. Community-based seroepidemiological survey of HCV infection in Catalonia, Spain. J Med Virol 65(4), 688-693. 2001.

249. Shaib, Y. H., Davila, J. A., McGlynn, K., and El Serag, H. B. Rising incidence of intrahepatic cholangiocarcinoma in the United States: a true increase? J.Hepatol. 40[3], 472-477. 2004.

250. Shaib, Y. H., El Serag, H. B., Davila, J. A., Morgan, R., and McGlynn, K. A. Risk factors of intrahepatic cholangiocarcinoma in the United States: a case-control study. Gastroenterology 128[3], 620-626. 2005.

251. Evans, T. G., Bonnez, W., Rose, R. C., Koenig, S., Demeter, L., Suzich, J. A., O'Brien, D., Campbell, M., White, W. I., Balsley, J., and Reichman, R. C. A Phase 1 study of a recombinant viruslike particle vaccine against human papillomavirus type 11 in healthy adult volunteers. J.Infect.Dis. 183[10], 1485-1493. 15-5-2001.

252. Richardson, S. Discussion of Spiegelhalter et al. J Roy Stat Soc, Series B 64, 631. 2002.

253. Celeux, G, Forbes, F, Robert, C. P., and Titterington, D. M. Deviance Information Criteria for missing data models. Bayesian analysis (to appear) . 2006.

254. Davis, D. L., Hoel, D., Fox, J., and López, A. International trends in cancer mortality in France, West Germany, Italy, Japan, England and Wales, and the USA. The Lancet 336, 474-481. 1990.

255. Hill, C., Benhamou, E., and Doyon, F. Trends in cancer mortality, France 1950-1985. British Journal of Cancer 63, 587-590. 1991.