DOES URSODEOXYCHOLIC ACID PROLONG SURVIVAL IN PATIENTS WITH PRIMARY SCLEROSING CHOLANGITIS: A META-ANALYSIS

STEVEN GRUCHY

Does Ursodeoxycholic Acid Prolong Survival In Patients With Primary Sclerosing Cholangitis: A Meta-Analysis

by

© Steven Gruchy, MD

A thesis submitted to the
School of Graduate Studies
in partial fulfillment of the
requirements for the degree of
Master of Science (Medicine)

Faculty of Medicine
Memorial University of Newfoundland

April 2010

St. John's Newfoundland

Abstract

Primary Sclerosing Cholangitis (PSC) is a progressive liver disease of unknown etiology. This disease can lead to many potential lethal clinical situations including liver cirrhosis. Ursodeoxycholic acid (UDCA) has been shown to be effective in other cholestatic liver diseases, most notably primary biliary cirrhosis. A number of randomized controlled trials (RCTs) using UDCA for the treatment of PSC have been carried out with varying results. The main objective of this study was to determine if the literature provides evidence that UDCA is effective at prolonging survival in patients with PSC.

Meta-analysis was used to evaluate the effect of UDCA on disease progression in patients with PSC. Only RCTs that compared UDCA to placebo in patients with PSC were included. Six fully published RCTs that met the inclusion criteria for this meta-analysis were identified in the literature. The outcome measurements used for this study included overall mortality and the requirement for liver transplant. Surrogate markers for the primary outcome of overall mortality were also analyzed and included worsening of liver histology, AST (U/L), ALP (U/L), albumin (g/L) and bilirubin (umol/L) levels. Subgroup analysis was also performed comparing high dose (>15mg/kg/day) to low/standard dosing (10-15mg/kg/day) of UDCA in patients with PSC for the primary outcome of overall mortality.

Pooling of the six fully published RCTs, identified a non-significant difference between treatment groups for the outcome of all cause mortality with an odds ratio of 0.859 and (95% confidence interval (CI), 0.365–2.022, p=0.728). A subgroup analysis of overall mortality stratified according to UDCA dosing did not identify any statistically significant difference in survival regardless of the dose of UDCA administered. A non-

significant difference between treatment groups for the pooled results of the outcome liver transplant required yielded an odds ratio 1.243 with (95%CI, 0.667-2.317, p=0.494). A non-significant difference between treatment groups was observed for the surrogate outcome worsening of liver histology with an odds ratio of 0.903 and (95%CI, 0.316-2.582, p=0.849). The pooled standardized mean difference (SMD) was calculated for the surrogate outcomes of liver biochemistry (AST (U/L), ALP (U/L)) and liver function (bilirubin (umol/L), albumin (g/L)). All surrogate outcomes suggested a benefit favoring UDCA, however, only AST (U/L), ALP (U/L) and bilirubin (umol/L) were statistically significant with a p-value <0.05.

These results indicate that although UDCA improved surrogate outcomes such as liver biochemistry and liver function, the results did not translate into a reduction in endpoints such as mortality or a need for liver transplant. This meta-analysis suggests that high dose UDCA (>15mg/kg/day) does not offer any treatment benefit over low/standard dosing (10-15mg/kg/day) for the outcome overall mortality in patients with PSC. Further research is needed to identify an effective medication to halt the progression of this disease. Future research may determine if starting therapy with UDCA at an earlier stage of disease translates into a survival advantage for patients with PSC.

Acknowledgements

A special thank you to my supervisor, Dr. John Fardy for his encouragement, expertise and guidance throughout the planning and writing of this thesis.

I would like to thank my supervisory committee, Drs. Mark Borgaonkar and John Harnett, for their input during the writing of this thesis, particularly during the final review.

Thank you to the division of Gastroenterology at Dalhousie University, especially Dr. Desmond Leddin for the encouragement and support I've received during this educational endeavor.

I would especially like to thank my wife, Jennette, as the completion of this thesis would not have been possible without her support and understanding.

Table of Contents

Abstract	ii
Acknowledgements	iv
Table of Contents	v
List of Tables	ix
List of Figures	x
List of Abbreviations	xii
List of Appendices	xiv
Chapter 1 Introduction	1
1.1 Primary Sclerosing Cholangitis	1
1.1.1 Definition and Overview	1
1.1.2 Epidemiology	2
1.1.3 Etiology and Pathogenesis	3
1.1.4 Diagnosis	4
1.1.5 Pathology	5
1.1.6 Natural History	6
1.1.7 Complications	7
1.1.8 Management	8
1.2 Ursodeoxycholic Acid	12
1.2.1 Structure and Overview	12
1.2.2 Mechanism of Action	13

1.2.3 Treatment for Cholestatic Liver Disease 14

1.3 Statement of Problem 17

1.4 Meta-Analysis 17

1.4.1 Objectives and Rationale 17

1.4.2 Design and Interpretation 18

1.5 Research Objectives 19

1.6 Research Questions 20

Chapter 2 Methods 21

2.1 Literature Search 21

2.2 Inclusion Criteria 22

2.3 Exclusion Criteria 22

2.4 Outcome Measurements 22

2.5 Study Quality 22

2.6 Data Abstraction 23

2.7 Combinability of Results 24

2.8 Combining the Results 25

Chapter 3 Results 27

3.1 Literature Review 27

3.2 Quality of Studies 29

3.3 Comparability of Patients Studied 32

3.4 Comparability of Randomized Controlled Trials 34

3.5 Comparability of Outcomes 36

3.6 Outcome Measurements - Pooling of Data and Exploration of Heterogeneity 39

3.6.1 All Cause Mortality 39

3.6.2 Sub Group Analysis – All Cause Mortality 43

3.6.3 Worsening of Liver Histology 47

3.6.4 Liver Transplant Required 51

3.6.5 Albumin 54

3.6.6 Bilirubin 57

3.6.7 AST 60

3.6.8 ALP 63

3.7 Sensitivity Analysis – Publication Bias 67

Chapter 4 Discussion 71

4.1 Quality of Studies – The Jadad Scoring System 71

4.2 Interpretation of Meta-Analysis 72

4.3 The Use of Surrogate Markers in Clinical Research 78

4.4 The Use of Mean Difference and Standardized Mean Difference in Meta-Analysis 83

4.5 Potential Pitfalls With The Test For Heterogeneity 85

4.6 The Fixed and Random Effects Models in Meta-Analysis 86

4.7 Publication Bias 87

4.8 Literature That May Have Added To This Meta-Analysis 88

4.9 Incomplete Data From Studies Included In This Meta-Analysis 90

4.10 Ursodeoxycholic Acid in Primary Sclerosing Cholangitis:
A Comparison of Two Meta-Analyses 94

Chapter 5 Conclusion 97

Bibliography 100

Appendix I Jadad Quality Score Calculation 110

Appendix II Data Extraction Sheet 112

Appendix III Statistical Formulae 115

List of Tables

Table 3.1	Code Numbers and Quality Scores For Published Randomized Controlled Trials	30
Table 3.2	Detailed Quality Scores For Each Trial	31
Table 3.3	Patient Characteristics	33
Table 3.4	Characteristics of Individual Trials	35
Table 3.5	Comparability of Outcome Measurements	38
Table 3.6	Outcome Measurement - Overall Mortality	41
Table 3.7	Overall Mortality in PSC Patients Receiving UDCA or Placebo, Subgroup Analysis According to UDCA Dosing	45
Table 3.8	Outcome Measurement - Worsening of Liver Histological Stage	49
Table 3.9	Outcome Measurement - Liver Transplant Required	52
Table 3.10	Outcome Measurement - Albumin (g/L)	55
Table 3.11	Outcome Measurement - Bilirubin (umol/L)	58
Table 3.12	Outcome Measurement - AST (U/L)	61
Table 3.13	Outcome Measurement - ALP (U/L)	65

List of Figures

Figure 3.1	A Forest Plot; Using Odds Ratio and 95% CI, Comparing All Cause Mortality in PSC Patients Receiving UDCA or Placebo	42
Figure 3.2	A Forest Plot; Using Odds Ratio and 95% CI, Comparing All Cause Mortality in PSC Patients Receiving UDCA or Placebo, Stratified According to UDCA Dosing	46
Figure 3.3	A Forest Plot; Using Odds Ratio and 95% CI, Comparing the Occurrence of Worsening Liver Histology in PSC Patients Receiving UDCA or Placebo	50
Figure 3.4	A Forest Plot; Using Odds Ratio and 95% CI, Comparing the Requirement of Liver Transplantation in PSC Patients Receiving UDCA or Placebo	53
Figure 3.5	A Forest Plot; Using Standardized Mean Difference and 95% CI, Comparing Albumin (g/L) Levels in PSC Patients Receiving UDCA or Placebo	56
Figure 3.6	A Forest Plot; Using Standardized Mean Difference and 95% CI, Comparing Bilirubin (umol/L) Levels in PSC Patients Receiving UDCA or Placebo	59
Figure 3.7	A Forest Plot; Using Standardized Mean Difference and 95% CI, Comparing AST (U/L) Levels in PSC Patients Receiving UDCA or Placebo	62

Figure 3.8 A Forest Plot; Using Standardized Mean Difference and 95% CI, Comparing ALP (U/L) Levels in PSC Patients Receiving UDCA or Placebo 66

Figure 3.9 Funnel Plot of Standard Error by Standardized Mean Difference, for the Outcome Bilirubin (umol/L) in PSC Patients Receiving UDCA or Placebo 69

Figure 3.10 Funnel Plot of Standard Error by Standardized Mean Difference, for the Outcome ALP (U/L) in PSC Patients Receiving UDCA or Placebo 70

List of Abbreviations

ACE – Angiotensin converting enzyme

ACG – American College of Gastroenterology

ALP – Alkaline phosphatase

ALT – Alanine aminotransferase

ANA – Antinuclear antibody

ARB – Angiotensin receptor blocker

AST – Aspartate aminotransferase

CBD – Common bile duct

CDDW – Canadian Digestive Diseases Week

CEA – Carcinoembryonic antigen

CF – Cystic fibrosis

CI – Confidence interval

CMA – Comprehensive Meta-Analysis Volume 2

DDW – Digestive Disease Week

ERCP – Endoscopic retrograde cholangiopancreatography

GI – Gastrointestinal

HDL – High-density lipoprotein

HLA – Human leukocyte antigen

IBD – Inflammatory bowel disease

JF – John Fardy

LDL – Low-density lipoprotein

MAPK – Mitogen activated protein kinase

MD – Mean difference

MELD – Model for end stage liver disease

MRCP – Magnetic resonance cholangiopancreatography

p-ANCA – perinuclear antineutrophil cytoplasmic autoantibody

PBC – Primary biliary cirrhosis

PSC – Primary Sclerosing Cholangitis

RCT – Randomized controlled trial

SD – Standard deviation

SE – Standard error

SG – Steven Gruchy

SMD – Standardized mean difference

UC – Ulcerative colitis

UDCA – Ursodeoxycholic acid

UEGW – United European Gastroenterology Week

χ^2 – Chi Square

List of Appendices

Appendix I	Jadad Quality Score Calculation	110
Appendix II	Data Extraction Sheet	112
Appendix III	Statistical Formulae	115

Chapter 1

Introduction

This chapter will provide an overview of the disease primary sclerosing cholangitis (PSC). The seriousness of this illness and the need for effective treatment will be discussed. Ursodeoxycholic acid (UDCA) will be introduced as an effective drug used in other cholestatic liver diseases. The usefulness of UDCA for patients in PSC will be explored and the results will lead to the development of a meta-analysis to determine if UDCA can prolong survival in patients with PSC. This chapter will also include the study objectives and research questions that will be investigated in this meta-analysis.

1.1 Primary Sclerosing Cholangitis

1.1.1 Definition and Overview.

PSC is a chronic cholestatic liver disease that affects both intra and extrahepatic bile ducts. Chronic inflammation and fibrosis leads to progressive destruction of the bile ducts (Silveira & Lindor, Clinical Features and Management of Primary Sclerosing Cholangitis, 2008). Bile is produced in the liver, stored in the gallbladder and secreted into the small bowel after ingesting a meal to aid in digestion. This process can only occur if there are adequate bile ducts present to transport bile from the liver to the small bowel. In PSC, the primary site of damage is the biliary epithelium which does not have the ability to regenerate like hepatocytes when injured (Mitchell & Chapman, 1997). This will lead to failure of biliary excretion if continued injury from inflammation and fibrosis occurs (Mitchell & Chapman, 1997). Once this happens, there is damage to the hepatocytes ultimately leading to cirrhosis and hepatic dysfunction. Cirrhosis is

characterized by fibrosis of the liver parenchyma. This in turn can lead to such complications as portal hypertension and hepatocellular carcinoma. Portal hypertension may manifest as ascites (an accumulation of fluid in the peritoneum), gastrointestinal (GI) bleeding from varices, hepatic encephalopathy and renal or pulmonary impairment. PSC has its own potential complications in addition to cirrhosis. Progressive destruction of the bile ducts can lead to sepsis from infection within the biliary tree (cholangitis). Jaundice can occur from a dominant stricture in the extrahepatic biliary system. PSC has also been associated with an increased incidence of certain malignancies. Cholangiocarcinoma, gallbladder cancer, hepatocellular carcinoma and colon cancer all occur in higher frequencies among patients with PSC (Silveira & Lindor, Clinical Features and Management of Primary Sclerosing Cholangitis, 2008).

1.1.2 Epidemiology.

A Canadian population based study has shown that the incidence of PSC is 0.92 cases per 100,000 person years (Kaplan, Laupland, Butzner, Urbanski, & Lee, 2007). The reported incidence is similar to other studies carried out in the United States and Europe. A population based estimate of the prevalence of PSC conducted in the United States revealed a rate of 13.6 per 100,000 people (Bambha, et al., 2003). Two thirds of patients with PSC are male with a mean age at diagnosis of forty years (Silveira & Lindor, Primary Sclerosing Cholangitis, 2008). PSC is highly associated with inflammatory bowel disease (IBD), particularly ulcerative colitis (UC). Approximately 75% of patients with PSC are found to have UC (Wiesner & LaRusso, 1980). However, only about 2-6% of patients with UC have PSC (Chapman, 2003). It is therefore not

surprising that even with a high prevalence of UC in the population, PSC remains a rare diagnosis.

1.1.3 Etiology and Pathogenesis.

There are currently several theories to explain the etiology and pathogenesis of PSC. Several investigators have hypothesized that a dysregulated immune system may lead to PSC. Reports of tissue lymphocyte populations, abnormal cytokines and the aberrant expression of human leukocyte antigen (HLA) class II molecules on the bile duct epithelium have provided some evidence that PSC is an immune mediated disease (Cullen & Chapman, 2005). This research may explain a potentially exaggerated cell mediated immune response targeted at the bile duct epithelium leading to the features seen in PSC.

Other researchers have sought to prove a genetic link to the pathogenesis of PSC. There is an increased incidence of PSC in patients who have first-degree relatives with the disease. Hazard ratios of 11.5, 11.1 and 2.3 were reported for the risk of developing PSC in an offspring, sibling or parent, respectively in a PSC patient cohort compared to a cohort of patients without PSC (Bergquist, et al., 2008). Patients with PSC have been found to have an increased prevalence of HLA-B8, -DR3, and -DRw52a (Van Milligen de Wit, Van Deventer, & Tytgat, 1995). DR2 appears to be associated with a younger onset of the disease whereas DR4 seems to identify more rapid disease progression (Portincasa, et al., 2005).

The strong association between PSC and UC has led investigators to consider the possibility of a bacterial etiology for the pathogenesis of PSC. The theory is that bacteria

can trans-locate across an inflamed colonic wall and migrate to the portal circulation where a chronic inflammatory response can occur in the biliary tract leading to fibrosis (Lee & Kaplan, 1995). This theory has not been validated in the literature and recent results have not shown any evidence to support the role of bacteria in the pathogenesis of PSC (Cullen & Chapman, 2005).

Another theory linking IBD and PSC suggests that mucosal lymphocytes produced in the colonic wall during an active flare of IBD persist as memory cells in the enterohepatic circulation. These gut derived lymphocytes can then become activated to produce biliary inflammation. This theory has been supported by literature showing that some lymphocyte homing receptors are shared by both the colon and liver (Grant, Lalor, Hubscher, Briskin, & Adams, 2001).

1.1.4 Diagnosis.

The diagnosis of PSC often occurs in asymptomatic patients presenting with raised cholestatic liver enzymes on laboratory evaluations. Approximately 15-55% of patients with PSC are asymptomatic at the time of diagnosis (Silveira & Lindor, Primary Sclerosing Cholangitis, 2008). An elevation of alkaline phosphatase (ALP) is the biochemical hallmark of PSC (Wiesner & LaRusso, 1980). Increases in the serum alanine aminotransferase (ALT) and aspartate aminotransferase (AST) are usually only mild to moderate. Once the diagnosis is suspected, usually in the patient with UC who has a rise in ALP, other serological tests can be ordered with differing sensitivities and specificities for the diagnosis of PSC. Antinuclear antibody (ANA) can be found in low titers in 20-60% of patients (Silveira & Lindor, Clinical Features and Management of

Primary Sclerosing Cholangitis, 2008). Anti-mitochondrial antibodies, seen frequently in patients with another cholestatic liver disease – primary biliary cirrhosis (PBC), are rarely seen in patients with PSC (Wiesner & LaRusso, 1980). Perinuclear antineutrophil cytoplasmic autoantibody (p-ANCA) can be found in patients with PSC with a sensitivity of 49% and a specificity of 89% making it a useful serologic test to help rule in the diagnosis of PSC (Bansi, Chapman, & Fleming, 1996).

PSC is usually confirmed radiographically. Modalities such as endoscopic retrograde cholangiopancreatography (ERCP) and magnetic resonance cholangiopancreatography (MRCP) are very sensitive and specific at diagnosing PSC. These imaging modalities will identify the classic features of PSC including diffuse multifocal strictures usually involving both the intra and extra hepatic bile ducts. The strictures seen are typically short and annular alternating with normal or minimally dilated segments to produce a characteristic "beaded" appearance (MacCarty, LaRusso, Wiesner, & Ludwig, 1983).

1.1.5 Pathology.

The histological changes in PSC can be very patchy. Anatomically, the bile ducts are distributed throughout the liver and can have varying amounts of inflammation and fibrosis. It is therefore understandable that a liver biopsy, which may only sample a few bile ducts, can demonstrate anything from normal bile ducts to frank biliary cirrhosis (Cullen & Chapman, 2005). This low sensitivity for liver biopsy makes it less favorable to be the sole method of diagnosing PSC. Liver histology is however useful for identifying the grade of PSC. Since 1981 a comprehensive grading system has been

created in order to determine the severity of PSC in liver samples. The histologic findings initially start in the portal triad (bile duct, portal vein, hepatic artery) and eventually spread to the hepatic parenchyma. There are four stages in the grading system for PSC. Stage 1 consists of enlargement, edema, and scarring of the portal triads, and mononuclear cell infiltration with some piecemeal necrosis and damage to isolated bile ducts. Stage II consists of expansion of portal triads with fibrosis extending into the surrounding parenchyma. Stage III is characterized by bridging fibrosis and stage IV represents cirrhosis (Ludwig, Barham, LaRusso, Elveback, Wiesner, & McCall, 1981).

1.1.6 Natural History.

PSC has a variable course from one patient to another; however, overall this chronic liver disease is progressive with a mean survival of 12 years (Farrant, et al., 1991). Disease progression and prognosis can be determined using a scoring system called the Mayo risk score. This is a scoring system that uses several surrogate markers in order to predict the natural history in an individual. A surrogate marker is defined as a laboratory investigation or any other intermediate substitute that is used to evaluate a treatment response on a clinically meaningful outcome measure. (Gluud, Brok, Gong, & Koretz, 2007). The Mayo risk score is calculated using serum bilirubin, AST, albumin, the patients' age and whether they have previously had variceal bleeding (Kim, Therneau, & Wiesner, 2000).

1.1.7 Complications.

Several complications can occur as PSC progresses in severity. Portal hypertension (ascites, GI bleeding, hepatic encephalopathy), liver failure, cholestasis, cholelithiasis and choledocholithiasis have all been reported (Portincasa, et al., 2005). GI bleeding can also occur in patients with UC from peristomal varices after proctocolectomy and ileal stoma formation (Chapman, 2003). Cholestasis can lead to fat malabsorption which in turn can result in deficiency of the fat-soluble vitamins A,D,E and K. It has been reported that vitamin A deficiency occurs in up to 50% of patients with PSC (Portincasa, et al., 2005). Osteoporosis is a known complication that may be related to the osteoblast inhibitors found in the serum of patients with cholestasis (Janes, Dickson, Okazaki, Bonde, McDonagh, & Riggs, 1995). Dominant extrahepatic strictures may occur in 15-20% of patients with PSC, leading to further complications such as jaundice, fever, pruritus and anorexia (Portincasa, et al., 2005). There is an increased risk of malignancy in patients with PSC. The most widely reported malignancy is cholangiocarcinoma that caries a lifetime risk of 10-15% (Lee & Kaplan, 1995). It may be difficult to distinguish between a dominant extrahepatic stricture and cholangiocarcinoma. The diagnosis of cholangiocarcinoma should be considered in patients who have rapid clinical deterioration including jaundice, weight loss and abdominal pain (Rosen, Nagorney, Wiesner, Coffey, & LaRusso, 1991). The young age of disease onset and inevitable progression of PSC along with the severe potential complications suggest a need for effective medical or surgical treatment that can halt the natural history of PSC.

1.1.8 Management.

The only effective treatment to date for PSC is liver transplantation. Five-year survival has been reported as high as 89% following liver transplantation (Farges, Malassagne, Sebagh, & Bismuth, 1995). The majority of liver transplants occur after there has been liver decompensation manifested by ascites, GI bleeding or hepatic encephalopathy. There has been debate about the optimal timing for liver transplant in patients with PSC as individuals with this disease can appear well and deteriorate quickly from complications such as cholangiocarcinoma. Patients with PSC are also considered for liver transplantation if they have intractable pruritus, fatigue or recurrent cholangitis (Silveira & Lindor, Primary Sclerosing Cholangitis, 2008). Unfortunately, liver transplantation does not provide a cure for PSC. There are reports of recurrent PSC in the liver graft in 2-40% of transplanted patients (Gautam, Cheruvattath, & Balan, 2006). Another potential consideration is the increased risk of colon cancer in patients with UC post transplant (Vera, et al., 2003). Liver transplantation is a surgical procedure with related morbidity and mortality. This procedure does not offer a cure of PSC and there is the potential increased risk of malignancy following transplantation. It is understandable that many physicians and patients are interested in delaying transplant as long as possible. The only way to successfully achieve this goal is to find a medication that can halt or reverse fibrosis and inflammation of the biliary tract as well as decrease the incidence of cholangiocarcinoma.

Throughout the past three decades, there has been a lot of interest in finding an effective medical treatment for patients with PSC. Treatments have focused on managing

the symptoms and complications of the disease. There is also ongoing research into treatments that can halt or reverse the course of this disease.

There are complications related to PSC that are common to all end stage liver disease. These include manifestations of portal hypertension such as ascites, encephalopathy, GI bleeding from varices as well as renal and pulmonary disease. These conditions are managed in the same way irrespective of the etiology of liver disease. A complication that is more common in patients with PSC includes bleeding from stomal varices post colectomy in UC patients (Wiesner R. , LaRusso, Dozois, & Beaver, 1986). Ideally, patients with PSC who require a colectomy for active UC should have an ileal-pouch-anal anastamosis procedure in order to avoid an ileal stoma (Mitchell & Chapman, 1998). In order to facilitate the construction of this pouch, the surgeon needs to leave some rectum behind for the anastamosis. There is an increased risk of colonic/rectal malignancy in patients with PSC and UC and therefore annual surveillance of the rectum and pouch with random biopsies using a flexible sigmoidoscope is recommended (Chapman, 2003).

Dominant biliary strictures are another complication that can occur in patients with PSC. These extrahepatic strictures may cause a reduction in the flow of bile leading to a more rapid clinical deterioration and biliary cirrhosis. In one study, dominant strictures occurred in 40% of patients with PSC over a five-year period (Stielh, Rudolph, Kloters-Plachky, Sauer, & Walker, 2002). At the present time, dominant strictures can be treated surgically or endoscopically. There are reports of good outcomes with biliary-enteric drainage of bile through surgical bypass of the strictured segment (Myburgh,

1994). However, surgical reconstruction and bypass are not done routinely for strictures as there are reports of scarring in the bile duct which may complicate future liver transplants (Chapman, 2003). There is also evidence to suggest that instead of a surgical reconstruction or bypass, liver transplantation may be a better option for the patient with five-year survivals of 89% post liver transplant (Farges, Malassagne, Sebagh, & Bismuth, 1995). Endoscopic treatments involve the use of ERCP in order to access the common bile duct (CBD) and to traverse the stricture with a guide wire in order to perform balloon dilatation and stenting of the stricture. This procedure facilitates the uninterrupted flow of bile. There are however reports of increased rates of cholangitis in patients with PSC undergoing ERCP (Mitchell & Chapman, 1998). It is likely that the introduction of bacteria into a biliary system containing intrahepatic strictures leads to this increased frequency of cholangitis.

Dominant strictures can often be difficult to differentiate from cholangiocarcinoma. ERCP permits brush cytology of the stricture in order to check for malignant cells; however, the sensitivity of a correct diagnosis is only 46% (Siqueira, et al., 2002). A combination of tumor markers using CA 19-9 and carcinoembryonic antigen (CEA) may be useful for identifying PSC patients with malignant strictures. In one study the combination of an abnormal CEA or CA 19-9 demonstrated a sensitivity of 100%, [95% confidence interval (CI) 65-100] and a specificity of 78.4%, [95% CI 63.1-89.7] (Siqueira, et al., 2002).

Several medications have been used to try and halt the progression of PSC. The rationale for these choices is largely based on the pathophysiology of PSC. Medications

that have been tried include those used in other cholestatic liver diseases, most notably primary biliary cirrhosis (PBC).

Increased copper levels have been detected in all patients with prolonged cholestasis. This observation led to the use of D-penicillamine in a randomized controlled trial (RCT) for patients with PSC. D-penicillamine is widely known as a copper chelator used to successfully treat Wilsons disease, a condition characterized by copper overload. In patients with PSC, no improvement on disease progression or overall survival was found in the treatment group. There were significant side effects including pancytopenia and proteinuria that led to drug discontinuation in 21% of patients (LaRusso, Wiesner, & Ludwig, 1988).

Colchicine is an antifibrogenic agent used in other cholestatic liver diseases with varying success. It has not been found to be successful to date in patients with PSC. In one study, 44 patients were randomized to 1mg/day of colchicine and 40 patients were randomized to placebo. At three years it was determined that there was no difference in liver biochemistry, liver histology or overall survival between the two groups (Olsson, et al., 1995).

Corticosteroids have been investigated for the treatment of patients with PSC. It was hypothesized that PSC is partly immune mediated and therefore corticosteroids used in other immune mediated conditions might provide some benefit to patients with PSC. The majority of patients with PSC also have UC. Steroids are the mainstay of treatment in active UC; however, when used in patients with PSC there has been very little improvement in the liver disease. There are also potentially severe long-term side effects

associated with corticosteroids that have discouraged large RCTs (Mitchell & Chapman, 1998). The use of corticosteroids is not currently recommended for the treatment of patients with PSC.

Several immunosuppressants have been used in the treatment of PSC. The most widely studied has been methotrexate. After encouraging findings in an open label study, an RCT compared oral pulse methotrexate at a dose of 15mg per week with a placebo group. The results showed that the only significant change was a fall in the liver enzyme ALP by 31%. Histology and cholangiographic parameters were unchanged (Knox & Kaplan, 1994). A pilot study using methotrexate and UDCA was performed by Lindor, et al; with similar negative results to the above RCT (Lindor, Jorgensen, & Anderson, 1996). Given the potential serious side effects of methotrexate including lung disease, hepatotoxicity and pancytopenia, there likely will not be a larger controlled trial using methotrexate in patients with PSC.

Other immunosuppressants reported in the literature to treat PSC have not been well studied and include azathioprine, cyclosporine and tacrolimus (Chapman, 2003). At the present time none of these medications are recommended routinely for the treatment of patients with PSC.

1.2 Ursodeoxycholic Acid

1.2.1 Structure and Overview.

UDCA is a medication that has been evaluated more than any other for the treatment of PSC. It is particularly attractive as a potential treatment for patients with PSC because of its known hepatoprotective properties. UDCA is a hydrophilic

dihydroxy (3α,7β-dihydroxy-5β-cholan-24oic acid) bile acid that accounts for 3% of the bile acid pool in adults (Paumgartner & Beuers, 2002). UDCA is not synthesized in the liver and is thought to be formed in the colon by bacterial 7β epimerization of the primary bile acid chenodeoxycholic acid. In turn, UDCA is then passively absorbed by the colonic mucosa and enters the portal circulation to become part of the bile acid pool (Angulo, 2002). Oral absorption of UDCA occurs through bile acid solubilization (Lazaridis, Gores, & Lindor, 2001). UDCA is taken up from the portal blood into the liver with a first pass extraction of about 50%. UDCA is then conjugated with glycine and taurine and actively secreted into bile (Paumgartner & Beuers, 2002). During continuous UDCA treatment, 19-64% of the total bile acids become UDCA, and up to 60% of the bile acids found in serum are UDCA depending on the daily dose used (Angulo, 2002). UDCA conjugates compete with endogenous bile acids for absorption in the ileum where they undergo enterohepatic circulation. The small amount of UDCA that is not absorbed in the terminal ileum is deconjugated in the colon and excreted in the feces. Less than 5% of the dose of UDCA that is excreted is found in urine samples (Paumgartner & Beuers, 2002).

1.2.2 Mechanism of Action.

There are several proposed mechanisms of action that are felt to contribute to the hepatoprotective properties of UDCA. The most widely accepted mechanism of action is the protection of cholangiocytes against cytotoxicity of hydrophobic bile acids (Paumgartner & Beuers, 2002). This likely occurs through the displacement of

hydrophobic bile acids from the bile salt pool during competitive absorption of UDCA in the ileum.

Stimulation of hepatobiliary secretion is another important mechanism of action for UDCA. This mechanism is thought to occur by the stimulation of transporter proteins in the hepatocyte along with the insertion and targeting of transporter molecules in the canalicular membrane, ultimately preventing the accumulation and retention of toxic bile acids that may lead to hepatocellular injury (Cullen & Chapman, 2005).

A third mechanism of action of UDCA is the protection of hepatocytes against bile acid induced apoptosis. Apoptosis is one of the main forms of hepatocyte death in cholestatic liver disease and occurs through the accumulation of hydrophobic bile acids in hepatocytes (Lazaridis, Gores, & Lindor, 2001). UDCA appears to have an antiapoptotic effect signaling a survival mechanism in hepatocytes. This survival mechanism is thought to occur through the activation of epidermal growth factor receptor and mitogen-activated protein kinases (MAPK) (Cullen & Chapman, 2005). Mitochondrial dysfunction by hydrophobic bile acids also leads to apoptosis. UDCA achieves cytoprotection to hepatocytes and cholangiocytes by preserving and stabilizing plasma membranes and mitochondria (Rodriques, Fan, & Wong, 1998).

1.2.3 Treatment for Cholestatic Liver Disease.

UDCA has been evaluated and is currently being used in several hepatocellular diseases especially those that cause cholestasis. One of the initial uses for UDCA was in gallstone dissolution. UDCA causes solubilization of cholesterol from the gallstone surface. Dissolution rates have been reported to be 30-60% (Tint, Salen, & Colalillo,

1982). This treatment has fallen out of favor with the advent of laparoscopic surgical procedures for cholecystectomy. UDCA is however, still used for high-risk surgical candidates (Angulo, 2002).

UDCA has been used in patients with cystic fibrosis (CF) who develop thick biliary secretions leading to biliary obstruction and potentially cirrhosis. Two studies have demonstrated improvement in biochemical and histological parameters when UDCA was used in this patient population. In 1998, Lindblad, Glaumann & Strandvik followed a cohort of CF patients with liver disease/cirrhosis based on liver biopsy. These patients received UDCA at a dose of 10-15mg/kg/day for two years. At the end of the study, the authors had shown statistically significant improvements in liver biochemistry and liver histology in the cohort, leading them to conclude that UDCA "modulates inflammation in CF-associated liver disease and indicates improvement in liver morphology during two years of treatment" (Lindblad, Glaumann, & Strandvik, 1998). In 1997, Van De Meeberg, et al; published an RCT in which thirty patients with CF and cholestatic liver disease were randomized to low dose (10mg/kg/day, n =17) or high dose (20mg/kg/day, n=13) UDCA. High dose UDCA produced a significant improvement and often complete response in liver biochemistry at two years compared to low dose UDCA. The authors concluded that because UDCA was so well tolerated, high dose UDCA should be the treatment of choice in this patient population (Van De Meeberg, Houwen, Sinaasappel, Heijerman, Bijleveld, & Vanberge-Henegouwen, 1997).

UDCA has been used as treatment for intrahepatic cholestasis of pregnancy. This condition affects females who are usually in their third trimester of pregnancy. It is

characterized by severe pruritus and an increase in liver enzymes in a cholestatic pattern. This condition has led to increased fetal distress, premature delivery and an increased risk of perinatal mortality (Angulo, 2002). In 1997, Palma, et al; performed an RCT in which patients were randomized to one gram of UDCA or placebo. This study found that UDCA improved the clinical symptoms of pruritus and also improved liver tests in this patient population. In this study, deliveries occurred at or near term in the treatment group and occurred earlier in the placebo group (Palma, et al., 1997).

Perhaps the greatest success of UDCA to date has been in patients with PBC. This is a cholestatic liver disease that shares some similarities to PSC. The disease most often occurs in middle-aged women and is characterized by destruction of small intralobular bile ducts and usually progresses to cirrhosis and liver failure if left untreated. There have been several well-performed RCT's that have demonstrated a survival benefit when UDCA was used in patients with PBC. In 1997 Poupon, et al; performed a combined analysis of the major RCTs using UDCA in patients with PBC. This analysis included a Canadian, French and Mayo clinic RCT. The results identified a survival benefit and a decreased need for liver transplant when UDCA was used for four years compared to patients receiving placebo, RR 1.9, 95% CI (1.3-2.8), $p<0.01$ (Poupon, Lindor, Cauch-Dudek, Dickson, Poupon, & Heathcote, 1997). There have been three meta-analyses published to date looking at the use of UDCA in PBC. Two of these studies did not reveal a survival benefit or a delay in time to transplant in PBC patients taking UDCA (Gong, Huang, Christensen, & Gluud, 2007), (Goulis, Leandro, & Burroughs, 1999). These two meta-analyses have been criticized for including studies with short duration of

follow up and for including studies using low doses of UDCA. Shi et al, in a third meta-analysis included only RCTs that had greater than two years of follow up and used doses of UDCA greater than 10 mg/kg/day. This meta-analysis showed a significant reduction in the incidence of liver transplantation (OR 0.65, p=0.01). The authors concluded that long term treatment with mid dose UDCA can improve liver biochemistry and survival free of liver transplantation in patients with PBC (Shi, Wu, Lin, Chen, Zhu, & Xie, 2006).

1.3 Statement of Problem

It is not surprising that based on prior experience with UDCA in cholestatic liver disease, physicians would try this medication for patients with PSC. There have been several non-randomized and randomized trials that have set out to answer the question of whether UDCA is efficacious in the treatment of patients with PSC. Many of these trials have recruited small numbers of patients, likely because of the low incidence of this disease in the population. Many trials used varying doses of UDCA in their protocols. These studies have displayed conflicting results. The question of whether UDCA prolongs survival or time to transplant in patients with PSC remains to be answered.

1.4 Meta Analysis

1.4.1 Objectives and Rationale.

A meta-analysis may be helpful in determining the usefulness of UDCA in patients with PSC. A meta-analysis critically reviews and statistically combines results of previous research. This type of study may add to our current knowledge of a given topic by providing a more precise estimate of the true effect than any one individual study (Tonelli, Hackam, & Garg, 2009). Sacks, et al; have outlined the main purposes and

strengths of a meta-analysis: 1) to increase statistical power for primary endpoints and for subgroups, 2) to resolve uncertainty when reports disagree, 3) to improve estimates of effect size, 4) to answer questions not asked at the start of a trial (Sacks, Berrier, Reitman, Ancona-Berk, & Chalmers, 1987). Using these criteria, it seems fitting that a meta-analysis may be used to answer a question such as "Does UDCA prolong survival in patients with PSC?"

1.4.2 Design and Interpretation.

There are two main aspects to consider in the design and interpretation of a meta-analysis. The first is the quantitative portion that deals with statistically combining results from individual trials. This is a very important part of the meta-analysis; however, the trials should not be combined unless their outcomes are similar enough to pass a statistical test of heterogeneity. If heterogeneity exists between the outcomes of the trials, the studies should be examined carefully to try and identify why the outcomes are not similar enough to be statistically combined (Hardy & Thompson, 1998). Occasionally reasons for heterogeneity are identified and a meta-analysis can still be performed through subgroup analysis of the included trials.

The second equally important component to the meta-analysis is the qualitative portion. A meta-analysis is much more robust if the author has investigated and described the study quality of each individual trial. It is important to record quality indicators such as whether randomization was complete, was blinding performed, were patients analyzed in an intent to treat fashion and was follow up adequate. The more information recorded in a trial the more robust the data ultimately leading to a stronger

study. Guidelines have been published to evaluate the quality of a meta-analysis. A high quality meta-analysis would provide positive answers for the following questions: 1) Is there evidence of a working protocol, 2) Are literature search strategies explicitly described, 3) Are inclusion and exclusion criteria specified, and reasons given for exclusions, 4) Are visual displays and tests of homogeneity done, 5) Are appropriate statistics and sensitivity analysis employed, 6) If the pooled analysis shows significant differences, is the issue of publication bias addressed, 7) Are conclusions drawn for treatment recommendations (L'abbe, Detsky, & O'Rourke, 1987).

1.5 Research Objectives

Using the above guidelines, a meta-analysis will be performed to achieve the following research objectives:

1) Determine if UDCA prolongs survival in patients with PSC.
2) Identify whether there is a survival advantage for using high dose (>15 mg/kg/day) UDCA in patients with PSC.
3) Determine if UDCA prevents worsening of liver histology in patients with PSC.
4) Identify whether UDCA decreases the need for liver transplantation in patients with PSC.
5) Determine the usefulness of the surrogate markers AST, ALP, albumin, and bilirubin, for predicting hard outcomes such as mortality or need for liver transplant in patients with PSC.

6) Explore the role of using standardized mean difference as an effect measure in meta-analysis for continuous outcome measurements.

1.6 Research Question

In order to accomplish the research objectives, a meta-analysis will be conducted to answer the primary research question,

"Does UDCA prolong survival in patients with PSC?"

A meta-analysis will be conducted to answer the following secondary research questions:

1) Does UDCA prevent the worsening of liver histology in patients with PSC?
2) Using subgroup analysis, does high dose (>15mg/kg/day) UDCA offer a survival advantage over low/standard dose (10-15mg/kg/day) UDCA in patients with PSC?
3) Does UDCA decrease the need for liver transplantation in patients with PSC?
4) In patients with PSC, do the liver enzymes ALP and AST significantly improve with UDCA compared to placebo?
5) In patients with PSC, do the liver functions albumin and bilirubin significantly improve with UDCA compared to placebo?
6) Does an improvement in surrogate markers such as liver enzymes and liver function correlate with a decrease in mortality or a need for liver transplantation in patients with PSC who are taking UDCA?

Chapter 2

Methods

This chapter will focus on the methodology used to perform this meta-analysis. Details of the literature search as well as the inclusion and exclusion criteria will be provided. The outcome measurements used in this meta-analysis will be listed. The method of assessing study quality will be explored. The details of the data abstraction to help determine if the outcome measurements are combinable will be presented. The statistical methods used for combining the results and determining outcome combinability will be provided.

2.1 Literature Search

The randomized controlled trials (RCTs) used in this meta-analysis were identified through a comprehensive literature search process. PubMed, Embase, the Cochrane library and Cinahl were the databases used for the literature search. Key words searched included primary sclerosing cholangitis, PSC, cholangitis, bile acids, ursodeoxycholic acid, ursodiol and UDCA. Limits placed on the search included clinical trials, randomized controlled trials, meta-analysis, human studies. All years and languages were included in the analysis. The reference section of each relevant study was reviewed in order to ensure that no RCTs were missed during the initial database search.

The abstracts for all major gastroenterology conferences were searched in an attempt to identify unpublished trials that were available in abstract form. The major conferences included, Canadian Digestive Diseases Week (CDDW), Digestive Disease

Week (DDW), United European Gastroenterology Week (UEGW) and the American College of Gastroenterology (ACG) Annual Scientific Meeting. These abstracts were manually searched from 1990-2009.

2.2 Inclusion Criteria

In order to meet inclusion criteria for this meta-analysis, the studies had to be RCTs that were fully published. Each RCT required a treatment arm (UDCA) and a placebo group used for comparison. It was decided a priori that published abstracts of RCTs would be included in the meta-analysis but would be pooled separately as part of a sensitivity analysis to minimize bias in the meta-analysis.

2.3 Exclusion Criteria

Studies were excluded from this analysis if they were not RCTs or if there was no placebo group. Studies were excluded if their outcome measurements did not include at least one of the outcome measurements used for this meta-analysis.

2.4 Outcome Measurements

The primary outcome measurement used for this meta-analysis was overall mortality. Secondary outcome measurements included worsening of liver histology, need for liver transplant, and changes in the liver enzymes (AST, ALP) and liver function (bilirubin, albumin).

2.5 Study Quality

The quality of each study was assessed using a protocol developed by Jadad (Jadad, et al., 1996) (Appendix I). This scoring system has been validated as appropriate for determining the methodological quality of individual trials. In fact, it has become the

most widely used scoring system in the world for determining the quality of trials (Haynes, Sackett, Guyatt, & Tugwell, 2008). Jadad, et al; have modified more lengthy protocols to include three key elements (randomization, blinding and withdrawals) that can be applied to RCTs and used to assess their overall methodological quality. The goal of the authors was to create a scoring system that could be used to assess a trials quality in a quick and efficient manner. The scoring system allocates a minimum score of zero and a maximum score of five. A trial with a score of less than three is felt to be of poorer quality and should be interpreted with caution. The benefit of the Jadad scoring system is that it is easy and quick to use. There are only five questions to answer in the scoring system. The protocol is, however, not without its criticism. Some authors believe that it is an oversimplified approach to determining methodological quality and that too much emphasis is placed on blinding (Berger, 2006).

2.6 Data Abstraction

Two independent critical appraisers (JF, SG) evaluated the assessed studies. Each evaluator was given a data abstraction sheet (appendix II) and Jadad's scoring system was applied for each study. Once a study was deemed appropriate for inclusion into the meta-analysis, a quality score from 0-5 was determined and the average of the two scores was taken as the final score. If the two appraisers had different viewpoints about a certain trial, discussion occurred and results were compared before a consensus was finally reached.

The key data collected included the dose of UDCA used, either at standard dosing of 10-15mg/kg/day or high dosing of >15mg/kg/day. The appraisers sought to determine

not only whether randomization was carried out but how was it accomplished. Each trial was assessed for blinding and whether appropriate follow up was performed. The reviewers determined if all patients were accounted for at the end of the trial. The primary outcomes of each study were noted. Baseline characteristics recorded included demographics such as average age, gender, and percent of patients with underlying IBD. Baseline liver function and biochemistry was recorded and compared to the liver function and biochemistry at the end of the trials.

When important information was not available but the study was deemed appropriate for the meta-analysis, attempts were made to contact the study authors for more information.

2.7 Combinability of Results

Each trial deemed appropriate for the meta-analysis was checked for similarities by comparing the study protocols and study populations. Appraisers manually checked the outcome measurements to determine combinability. The study drug was also assessed for homogeneity by determining if the dosage of UDCA was the same throughout all trials.

Statistical combinability was determined by formally checking for homogeneity of the outcomes in the trials. This was carried out by using a chi-square (χ^2) statistic for heterogeneity. The trials were felt to be combinable if the p-value for the statistical test for heterogeneity was >0.10. If the analysis showed heterogeneity, the individual studies were examined for their differences. The χ^2 test for heterogeneity is useful to serve as an indicator that the differences between trials may be due to more than just chance alone.

Attempts to explain the heterogeneity will be detailed when necessary. Sub group analysis will be attempted if differences in trials are felt to be clinically significant for the analysis.

2.8 Combining the Results

The data (liver biochemistries and liver function) from the trials were combined by calculating standardized mean differences (SMD) for each trial and then determining a pooled estimate of the SMD along with the 95% confidence interval (CI) for the pooled SMD. A p-value of <0.05 was determined to be statistically significant. Statistical analysis for the pooled SMD and 95% CI were carried out using Comprehensive Meta-Analysis Volume 2 (CMA) (Borenstein, Hedges, Higgins, & Rothstein, 2005).

Outcomes such as mortality, liver histology, and need for liver transplant were combined using a pooled estimate of the odds ratio and the 95% CI was then calculated for the pooled estimate of the odds ratio using CMA. A p-value of <0.05 was determined to be statistically significant.

There are two models used in meta-analysis to determine a combined effect size for the outcome of interest. The fixed effects model assumes that there is one effect size that is shared by all studies. The statistical pooling of the outcome leads to the estimate for this common effect size. In a fixed effects model, the only error in the estimate of the combined effect is the random error within each RCT (Spector & Thompson, 1991). This method of producing a combined effect size and confidence interval is often referred to as the liberal method and assumes that the trials are homogenous enough to be combined for analysis.

The other model used in meta-analysis is the random effects model which assumes that the true effect may vary from trial to trial. In this model there is not only variation within each trial but also variation between trials as well (Spector & Thompson, 1991). This model is often referred to as more of a conservative model. In the absence of heterogeneity, the fixed effects model and random effects model will often approximate each other (Alderson & Green, 2002). When a meta-analysis includes only a few trials, it is difficult to calculate the between trial variation and this model is not recommended for use. (Borenstein, Hedges, & Rothstein, 2007).

This current meta-analysis used a fixed effects model to determine the combined effect size for the various outcome measurements. If heterogeneity between the trials was discovered, attempts were made to resolve this heterogeneity. If the trials were felt to be combinable despite quantitative heterogeneity then a fixed effects model was still used because the small number of trials included in this meta-analysis would limit the ability of a random effects model to detect between trial variation.

Chapter 3

Results

In this chapter, the results of the literature review will be revealed and a qualitative analysis of the included RCTs will be performed. The RCTs will be compared in terms of overall quality, patient characteristics, trial methodology and outcome measurements.

A quantitative analysis of the included RCTs will be performed. The RCTs will be combined when appropriate to determine a pooled effect size and 95% CI for each outcome measurement. A sensitivity analysis will be conducted for statistically significant results between treatment groups.

3.1 Literature Review

A total of fifteen fully published studies were identified during the literature review. Although all languages were included in the review of the literature, all published studies suitable for this meta-analysis were written in English. Only six of these studies were fully published RCTs comparing UDCA versus placebo in patients with PSC (Beuers, et al., 1992), (Lindor, 1997), (Mitchell, Bansi, Hunt, Von Bergmann, Fleming, & Chapman, 2001), (Stiehl, Walker, Stiehl, Rudolph, Hofmann, & Theilmann, 1994), (Olsson, et al., 2005), (Lindor, et al., 2009). Another study was deemed suitable for the meta-analysis; however, there was no end point data listed in the published trial that could be used for this meta-analysis (De Maria, Colantoni, Rosenbloom, & Van Thiel, 1996). The principal investigator was contacted but did not respond to our request for more data. Three studies were excluded after it was determined that they compared

single dose UDCA to multi-dose UDCA in patients with PSC (Podda, et al., 1989), (Van de Meeberg, et al., 1996), (Van Hoogstraten, et al., 1998). One trial was excluded as it compared low dose UDCA, standard dose UDCA and high dose UDCA with no placebo group for comparison (Cullen, Rust, Flemming, Edwards, Beuers, & Chapman, 2008). Two pilot studies were not randomized and were therefore excluded from this meta-analysis (Harnois, Angulo, Jorgensen, LaRusso, & Lindor, 2001), (O'Brien, Senior, Arora-Mirchandani, Batta, & Salen, 1991). Two meta-analyses were discovered during the literature search. A meta-analysis of bile acid therapy in PSC was found in the Cochrane library and was published the same year that this meta-analysis was initially presented (Chen & Gluud, 2003), (Gruchy & Fardy, 2003). Another meta-analysis published in 2009 was identified after this current meta-analysis was completed (Shi, Li, Zeng, Lin, & Xie, 2009). The results will be reviewed and compared to this current meta-analysis in the discussion section.

The extended literature review, that included a review of the references for the relevant RCT's and searching the abstracts of all major gastrointestinal meetings in North America and Europe, identified five abstracts that might be suitable for our meta-analysis. After careful review of the published abstracts, we determined that two of the studies were not RCTs (Kim, Jorgemsen, Malinchoc, Benson, Dickson, & Lindor, 1997), (O'Brien, Craig, & Hatfield, 1993). Two studies were abstracts of preliminary results from fully published trials that have been reviewed and abstracted for consideration in this meta-analysis (Bansi, Christie, Fleming, & Chapman, 1996), (Van Thiel, Wright, & Gavaler, 1992). One abstract was deemed suitable for this meta-analysis; however, after

extensive investigation, we could not identify the necessary data for entry into the meta-analysis (Lo, et al., 1992). Our literature review could not determine whether this abstract was ever published as a full publication.

3.2 Quality of Studies

The quality of each of the six randomized controlled trials was assessed by the method previously outlined (Jadad, et al., 1996) (Appendix I). This method relies on explicit detail in the methodology section of each publication. It is difficult to accurately assign quality scores to abstracts given their inherent lack of detail. The quality scores of each of the six fully published RCT's are found in Table 3.1. A score between four and five in the Jadad system is generally felt to be of good methodological quality. A study with a quality score of three or less is felt to have some methodological flaws. Table 3.2 outlines the scores for the individual questions in each trial. The outline of the scoring system draws the readers' attention to the fact that the randomization and blinding processes are the main focus of this quality system. It is possible that important trial methodology may be underestimated in such a scoring system.

Table 3.1

Code Numbers and Quality Scores For Published

Author	Code Number	Quality Score[a]
Fully Published Trials		
Beuers, U; et al. 1992	F1	5/5
Lindor, K.D; et al. 1997	F2	4/5
Mitchell, S.A; et al. 2001	F3	4/5
Stiehl, A; et al. 1994	F4	3/5
Olsson, R; et al. 2005	F5	4/5
Lindor, K.D; et al. 2009	F6	5/5
Abstract Only		
Randomized Controlled Trials		
Lo, S.K; et al. 1992	A1	3/5

[a]Jadad Scoring System (see appendix I)

Table 3.2

Detailed Quality Scores For Each Trial

	F1	F2	F3	F4	F5	F6	A1
Was the study randomized?	1	1	1	1	1	1	1
Was the method used to generate the sequence of randomization described and appropriate?	1	1	0	1	0	1	0
Was the study double blind?	1	1	1	1	1	1	1
Was the method of double blinding described and appropriate	1	1	1	0	1	1	0
Was there a description of withdrawals and dropouts?	1	0	1	0	1	1	1
Inappropriate randomization (-1 point)	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Inappropriate blinding (-1 point)	n/a	n/a	n/a	n/a	n/a	n/a	n/a

3.3 Comparability of Patients Studied

The patient characteristics for the included trials are outlined in table 3.3. Patients with PSC are typically younger males who often have underlying inflammatory bowel disease (IBD). The age range for patients in these six trials is 38.5-52.0. The subtle difference in ages amongst the six trials may represent patients that are at different stages of disease. The majority of patients in the included trials are males with IBD. Four out of six trials (F1, F2, F3, F6) identify a majority of patients who are at stage I or II liver disease at entry into the trial. Trial F4 appears to have patients who have more advanced disease as 65% of patients already have stage III-IV liver disease at the onset of the study. Trial F5 does not include histological staging in their study. Without knowing the results of the trials, one might speculate that patients in the trial F4 may not have a response to UDCA since patients are already at an advanced histological stage and therefore the damage to the liver from PSC may be irreversible at that point. However, it is difficult to assess the severity of disease through biopsy alone in any individual patient as the disease is patchy and any one patient may have random biopsies showing anything from normal bile ducts to frank cirrhosis (Cullen & Chapman, 2005).

Table 3.3

Patient Characteristics

Trial Number	Mean Age	Male(%)	IBD(%)	Histologic Stage Number (%)			
				I	II	III	IV
F1	38.8	79	71	28.5	28.5	28.5	14.3
F2	42.8	58	81	18.5	30.5	33.5	18
F3	52.0	73	77	27.0	19.2	42.3	11.5
F4	38.5	Not listed	85	35		45.0	20.0
F5	43.3	70	85	---	---	---	---
F6	46.6	57	77	33.5	26.5	24.5	15.5

Note: IBD = Inflammatory bowel disease

3.4 Comparability of Randomized Controlled Trials

Clinical characteristics of the six trials are outlined in table 3.4. There is a large range in the size of the various trials, which is skewed towards the three larger trials that recruited 198 (F5), 150 (F6) and 105 (F2) patients each. The other three trials have much fewer patients. The overall effect size of the different outcome measurements for this meta-analysis will ultimately be impacted with a bias toward the larger trials.

Another important difference between the trials may be the dosage of UDCA used. Many experts believe that high dose UDCA (>15mg/kg/day) will have a greater impact on patients with PSC than using lower dosing (10-15mg/kg/day). Three trials (F3), (F5) and (F6) used high dose UDCA and met the criteria for inclusion in this meta-analysis. The remainder of the trials all used dosing <15mg/kg/day.

There was also a wide variety in the total duration of the RCTs. The recruitment phase of the trials lasted anywhere between three months and six years. Although one trial (F4) lasted only three months for the placebo-controlled portion of the study, the authors continued to follow patients for three years to determine the effect of UDCA on their outcome measurements. The ethical guidelines determined that this trial had to end after three months because there was a greater than two fold increase of serum transaminases in 80% of patients in the placebo group. All trials were similar in total length of follow up except for F5 and F6, which had a longer follow up (five years).

Table 3.4

Characteristics of Individual Trials

Trial	Number of Patients in Trial	Dosage of UDCA	Dosing (Single or Multidose)	Follow up Intervals (months)	Length of Follow up (months)	Duration of Trial (months)
F1	14	13-15 mg/kg/day	Multidose	6	12	12
F2	105	13-15 mg/kg/day	Multidose	3	24	72
F3	26	20 mg/kg/day	Multidose	3	24	24
F4	20	750mg/day	Single	3	36	3
F5	198	17-23 mg/kg/day	Multidose	6	60	60
F6	150	28-30 mg/kg/day	Multidose	3	60	72

Note: UDCA = Ursodeoxycholic acid

3.5 Comparability of Outcomes

The outcome measurements used in the included RCTs are listed in table 3.5. The effect size for the outcome measurements overall mortality, worsening of liver histology and liver transplant required was analyzed by calculating odds ratios.

The odds of death in the treatment group as compared to the placebo group, the odds of developing worsening liver histology in the treatment group as compared to the placebo group and the odds of requiring a liver transplant in the treatment group as compared to the placebo group were all calculated. These odds ratios and respective 95% CI were calculated as outlined in Appendix III. These odds ratios were then combined to give an overall pooled odds ratio and 95% CI for each dichotomous outcome measurement.

The odds ratio for each of the dichotomous outcomes along with their 95% CIs in the RCTs were tabulated in tables 3.6-3.9 and shown graphically in figures 3.1-3.4. These tables and figures also show the overall odds ratio for each dichotomous outcome along with its 95% CI.

Liver biochemistry results were analyzed by calculating an overall standardized mean difference (SMD) to allow for differences in laboratory values between the various studies. The mean values pre and post treatment for AST, ALP, bilirubin and albumin along with their standard deviations were used to calculate SMD's as outlined in Appendix III. These SMD's were then combined to give an overall pooled SMD and 95% CI.

The SMD's for each of the biochemical outcomes along with their 95% CI's in the trials were tabulated in Tables 3.10-3.13 and shown graphically in figures 3.5-3.8. These tables and figures also show the pooled SMD for each biochemical outcome with its 95% CI.

Table 3.5

Comparability of Outcome Measurements

RCT	Liver Enzymes/ Function	Histology	Endoscopic Abnormalities (ERCP)	Death	Liver Transplant Required	Tolerability of UDCA
F1	✔	✔		✔		
F2	✔	✔		✔	✔	✔
F3	✔	✔	✔	✔	✔	✔
F4	✔					
F5	✔			✔	✔	✔
F6				✔	✔	

Note: RCT = Randomized Controlled Trial; ERCP= Endoscopic Retrograde Cholangiopancreatography; UDCA = Ursodeoxycholic Acid

3.6 Outcome Measurements – Pooling of Data and Exploration of Heterogeneity

3.6.1 All Cause Mortality.

Five of the fully published trials listed overall mortality as an outcome measurement (F1,F2,F3,F5,F6). Examining figure 3.1, the forest plot demonstrates that four trials favored UDCA, while one trial (F6) favored placebo. The confidence intervals were large and they all crossed 1. None of the trials showed a significant difference between the UDCA group and the placebo group in terms of overall mortality and there was considerable overlap of the confidence intervals. The fixed effects model calculated a pooled odds ratio of 0.859 with 95% CI (0.365-2.022) and a non-significant p-value of 0.728 (table 3.6). As previously outlined, RCTs studying uncommon diseases require large numbers of patients in order to detect a difference in treatment groups for a rare outcome such as mortality. In this situation, the larger the study, the more events of interest occur resulting in a smaller confidence interval and a higher likelihood of achieving a statistical difference amongst treatment groups if a difference were truly present.

The formal test of heterogeneity (Appendix III) for this outcome yields a Q value of 1.480, with four degrees of freedom (df=4) and a p-value of 0.830. This would indicate that the trials are homogeneous for the outcome in question and are able to be combined.

In order to determine if the cause of overall mortality was related to the disease, study drug or another confounder, the specific cause of death in each patient needs to be examined. The overall mortality in the UDCA group was ten out of two hundred and

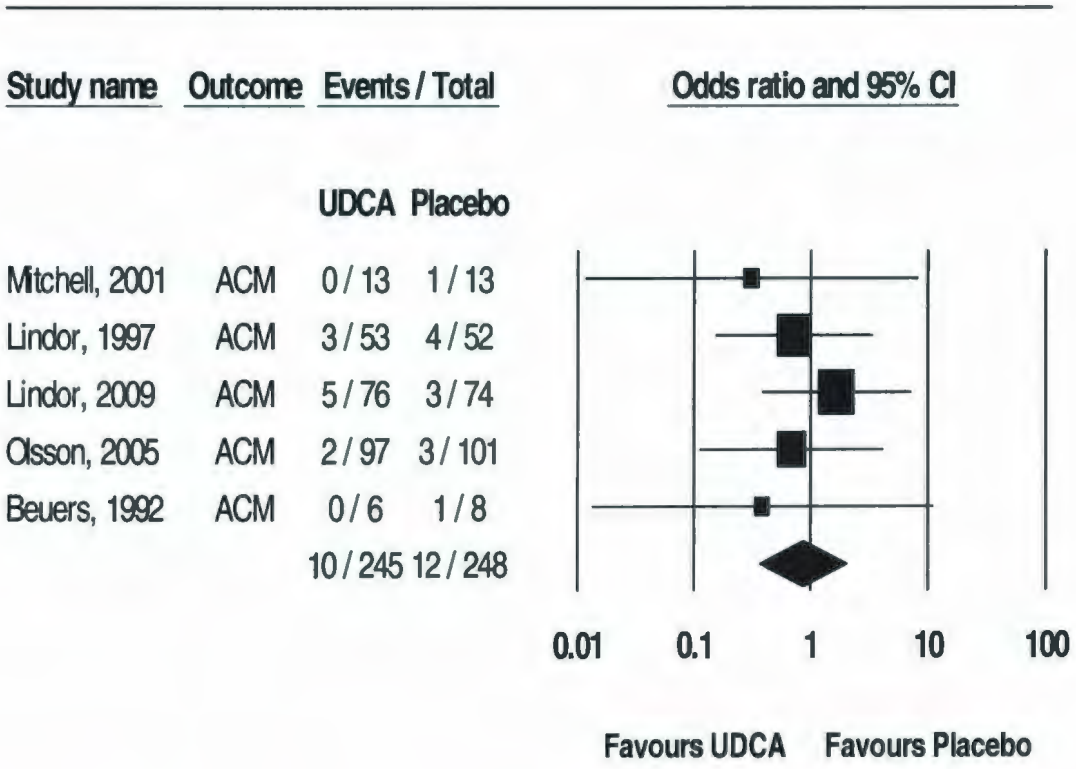
forty five patients. Five patients died of cholangiocarcinoma, a known complication from PSC, and five patients died of liver decompensation that can be explained from end stage liver disease secondary to PSC. In the placebo group twelve out of two hundred and forty eight patients died. Six patients died of cholangiocarcinoma, five died from liver decompensation and one person died from a cause listed as being unrelated to liver disease. This analysis would suggest that the mortality rates in both groups are likely attributable to having PSC and confounders are less likely. This is not surprising for this group of patients who tend to be younger and have less co-morbidities.

Table 3.6

Outcome Measurement - Overall Mortality

RCT	UDCA (n/nt)	Placebo (n/nt)	Model	OR	Lower 95% CI	Upper 95% CI	P-value
F1	0/6	1/8		0.385	0.013	11.168	0.578
F2	3/53	4/52		0.720	0.153	3.387	0.678
F3	0/13	1/13		0.309	0.011	8.300	0.484
F5	2/97	3/101		0.688	0.112	4.208	0.685
F6	5/76	3/74		1.667	0.384	7.239	0.495
Overall	**10/245**	**12/248**	**Fixed**	**0.859**	**0.365**	**2.022**	**0.728**

Note: RCT = Randomized controlled trial; UDCA = Ursodeoxycholic acid; n/nt = Number of patients affected with the outcome divided by the total number of patients in the group; OR = Odds ratio



Note: ACM = All Cause Mortality; UDCA = Ursodeoxycholic Acid

Figure 3.1: A Forest Plot; Using Odds Ratio and 95% CI, Comparing All Cause Mortality in PSC Patients Receiving UDCA or Placebo.

3.6.2 Sub Group Analysis - All Cause Mortality.

Once the analysis determined that there was no statistically significant difference between the UDCA group and placebo group in terms of the primary outcome of overall mortality, the data was closely inspected in order to determine if there was a specific sub group that may have a survival advantage from receiving UDCA. The RCTs were fairly homogenous in most baseline characteristics apart from the dose of UDCA that each trial used. Perhaps there was a difference in overall mortality between patients taking high dose UDCA (>15mg/kg/day) and those taking low/standard dose UDCA (10-15mg/kg/day).

A subgroup analysis arranged by dosing of UDCA was analyzed for the outcome of overall mortality. Three RCTs used high dose UDCA (F3,F5,F6) and two RCTs used low/standard dose UDCA (F1,F2). Figure 3.2 demonstrates a forest plot for this particular outcome. The graph demonstrates that in the high dose UDCA subgroup, two trials favor UDCA while one trial favors placebo. The CI's are large and they all cross 1. The fixed effects model for the subgroup of high dose UDCA calculated a pooled odds ratio of 1.017 with 95% CI (0.346-2.987) and a non-significant p-value of 0.976 (table 3.7). The formal test of heterogeneity (Appendix III) for the sub group of high dose UDCA for the outcome overall mortality yields a Q value of 1.118 with two degrees of freedom (df=2) and a p-value of 0.572.

The forest plot (figure 3.2) demonstrates that in the low/standard dose UDCA subgroup, both trials favor UDCA, although the CI's are large and both cross 1. The fixed effects model for the subgroup of high dose UDCA calculated a pooled odds ratio

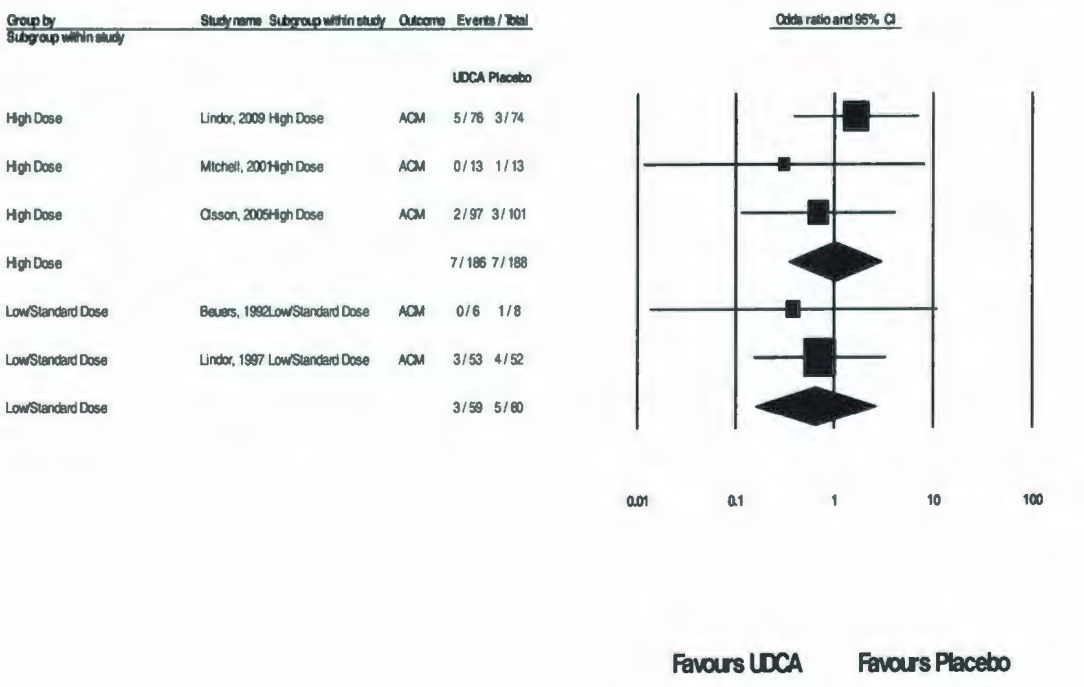
of 0.645 with 95% CI (0.158-2.635) and a non-significant p-value of 0.542 (table 3.7). The formal test of heterogeneity (Appendix III) for the sub group of low/standard dose UDCA for the outcome overall mortality yields a Q value of 0.110 with one degree of freedom (df=1) and a p-value of 0.740. This subgroup analysis suggests that there is no difference between UDCA and placebo in overall survival regardless of the dosing of UDCA administered.

Table 3.7

Overall Mortality in PSC Patients Receiving UDCA or Placebo, Subgroup Analysis According to UDCA Dosing

Model For Meta Analysis	Subgroup (UDCA Dosing)	RCT	UDCA (n/nt)	Placebo (n/nt)	OR	95% CI		P-value
						LL	UL	
	Low/Std Dose	F1	0/6	1/8	0.385	0.013	11.168	0.578
	Low/Std Dose	F2	3/53	4/52	0.720	0.153	3.387	0.678
Fixed	**Low/Std Dose**		**3/59**	**5/60**	**0.645**	**0.158**	**2.635**	**0.542**
	High Dose	F3	0/13	1/13	0.309	0.011	8.300	0.484
	High Dose	F5	2/97	3/101	0.688	0.112	4.208	0.685
	High Dose	F6	5/76	3/74	1.667	0.384	7.239	0.495
Fixed	**High Dose**		**7/186**	**7/188**	**1.017**	**0.346**	**2.987**	**0.976**

Note: UDCA = Ursodeoxycholic Acid; low/std = low and standard dosing; RCT= Randomized controlled trial; n/nt = Number of patients affected with the outcome divided by the total number of patients in the group;
OR = Odds ratio; LL = lower limit; UL = Upper limit



Note: ACM = All Cause Mortality; UDCA = Ursodeoxycholic Acid

Figure 3.2: A Forest Plot; Using Odds Ratio and 95% CI, Comparing All Cause Mortality in PSC Patients Receiving UDCA or Placebo, Stratified According to UDCA Dosing.

3.6.3 Worsening of Liver Histology.

Three of the fully published trials listed worsening of liver histology as an outcome measurement (F1,F2,F3). Two of the three trials (F1,F3) demonstrated a non-significant trend toward favoring the UDCA treatment group for improvement in liver histology. However, one trial (F2) demonstrated a trend toward worsening of liver histology in the UDCA group, although this trend was non-significant. Figure 3.3 demonstrates a forest plot for this particular outcome. The graph shows some degree of heterogeneity amongst the groups. The formal test of heterogeneity (appendix III) for this outcome yields a Q-value of 6.731, which with two degrees of freedom (df=2), leads to a p-value of 0.035. The point estimate for the odds ratio of worsening liver histology in the UDCA group as compared to the placebo group is 0.903 using a fixed effects model with a 95% CI (0.316-2.582) and non significant p-value of 0.849 (table 3.8). The small overall sample size, 67 patients in each group, may not be enough patients needed to detect a significant difference amongst treatment groups thus leading to a potential type II error.

The trials need to be examined in more detail to explain the potential heterogeneity. Two trials used the previously described Ludwig system for staging liver disease in patients with PSC (F2,F3). The third trial (F1) used a scoring system introduced by Poupon for staging patients with liver disease (Poupon, Balkau, Eschwege, & Poupon, 1991). All trials used a similar blinding system so that the pathologists were unaware of the treatment group the specimen belonged. The pathologist was also unaware of whether the liver biopsy was pre or post treatment in all studies. Only one

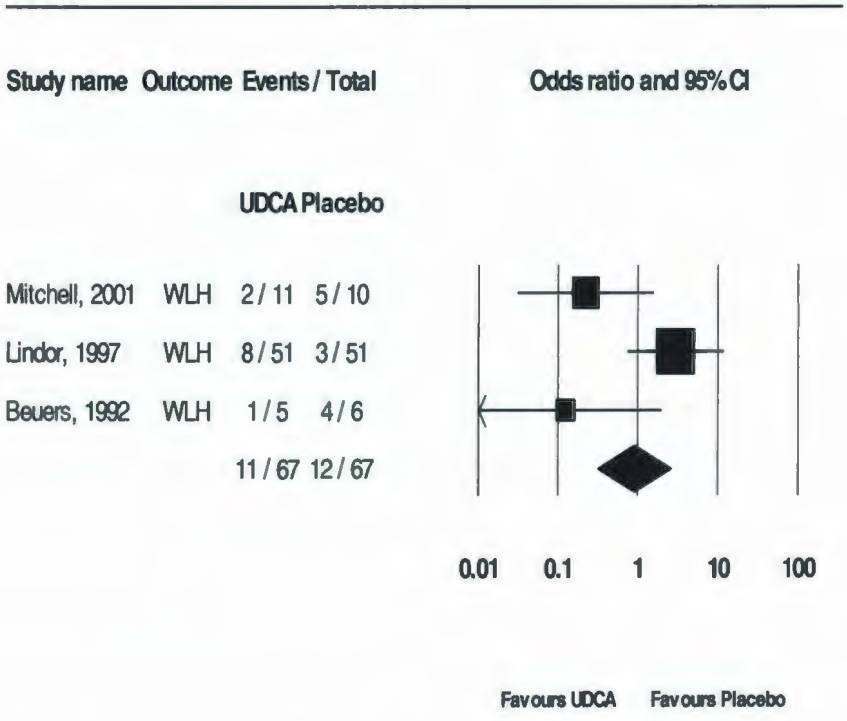
RCT (F3) used two pathologists who read all biopsies to decrease bias and improve quality control. Two trials (F2,F3) had biopsies analyzed at 24 months whereas one trial (F1) analyzed liver biopsies at 12 months. The difference in grading system may obviously introduce heterogeneity amongst the trials. The timing of liver biopsy may also introduce heterogeneity if the effect of UDCA on liver histology is more pronounced the longer the drug is administered to patients. However, this doesn't fully explain the results seen in this meta-analysis as the trial with the most patients and the longest follow up (F2), amongst the trials measuring histology as an outcome, was the trial that did not favor an improvement in liver histology for the UDCA group.

Table 3.8

Outcome Measurement – Worsening of Liver Histological Stage

RCT	UDCA (n/nt)	Placebo (n/nt)	Model	OR	Lower 95% CI	Upper 95% CI	P-value
F1	1/5	4/6		0.125	0.008	1.998	0.141
F2	8/51	3/51		2.977	0.742	11.942	0.124
F3	2/11	5/10		0.222	0.031	1.595	0.135
Overall	**11/67**	**12/67**	**Fixed**	**0.903**	**0.316**	**2.582**	**0.849**

Note: RCT = Randomized controlled trial; UDCA = Ursodeoxycholic acid; n/nt = Number of patients affected with the outcome divided by the total number of patients in the group; OR = odds ratio



Note: WLH = Worsening Liver Histology; UDCA = Ursodeoxycholic Acid

Figure 3.3: A Forest Plot; Using Odds Ratio and 95% CI, Comparing the Occurrence of Worsening Liver Histology in PSC Patients Receiving UDCA or Placebo.

3.6.4 Liver Transplant Required.

Four of the fully published trials listed a need for liver transplant as an outcome measurement (F2,F3,F5,F6). Three of the four trials (F2,F3,F6) demonstrated a non-significant trend toward an increased odds of requiring a liver transplant for the UDCA group; while one trial (F5) demonstrated a non-significant trend toward a decreased odds of requiring a liver transplant for the UDCA group. Figure 3.4 demonstrates a forest plot for this particular outcome. The formal test of heterogeneity (appendix III) for this outcome yields a Q-value of 2.922, with three degrees of freedom (df=3) and a p-value of 0.404. This would indicate that the trials are homogenous for the outcome in question and are able to be combined.

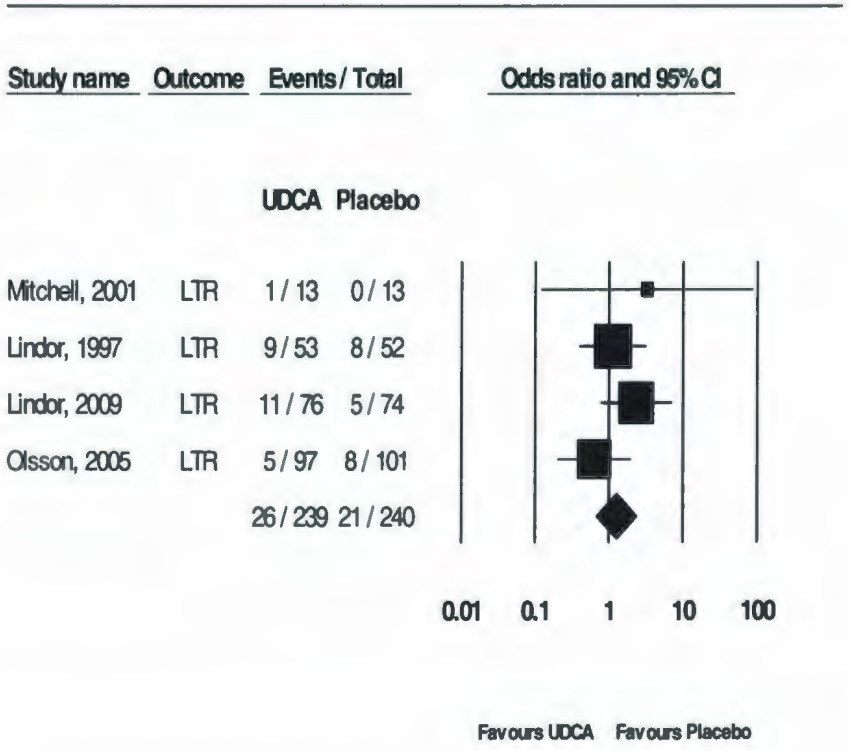
The point estimate for the odds ratio of requiring a liver transplant in the UDCA group as compared to the placebo group is 1.243 using a fixed effects model with a 95% CI (0.667-2.317) and non significant p-value of 0.494 (table 3.9). These results would suggest that UDCA does not decrease the need for liver transplant in patients with PSC. An alternative theory is that there are not enough patients requiring liver transplant in this meta-analysis to detect a significant difference between the treatment groups.

Table 3.9

Outcome Measurement – Liver Transplant Required

RCT	UDCA (n/nt)	Placebo (n/nt)	Model	OR	Lower 95% CI	Upper 95% CI	P-value
F2	9/53	8/52		1.125	0.398	3.183	0.824
F3	1/13	0/13		3.240	0.120	87.125	0.484
F5	5/97	8/101		0.632	0.199	2.003	0.435
F6	11/76	5/74		2.335	0.770	7.087	0.134
Overall	**26/239**	**21/240**	**Fixed**	**1.243**	**0.667**	**2.317**	**0.494**

Note: RCT = Randomized controlled trial; UDCA = Ursodeoxycholic acid; n/nt = Number of patients affected with the outcome divided by the total number of patients in the group; OR = Odds ratio



Note: LTR: Liver Transplant Received; UDCA = Ursodeoxycholic Acid

Figure 3.4: A Forest Plot; Using Odds Ratio and 95% CI, Comparing the Requirement of Liver Transplantation in PSC Patients Receiving UDCA or Placebo.

3.6.5 Albumin.

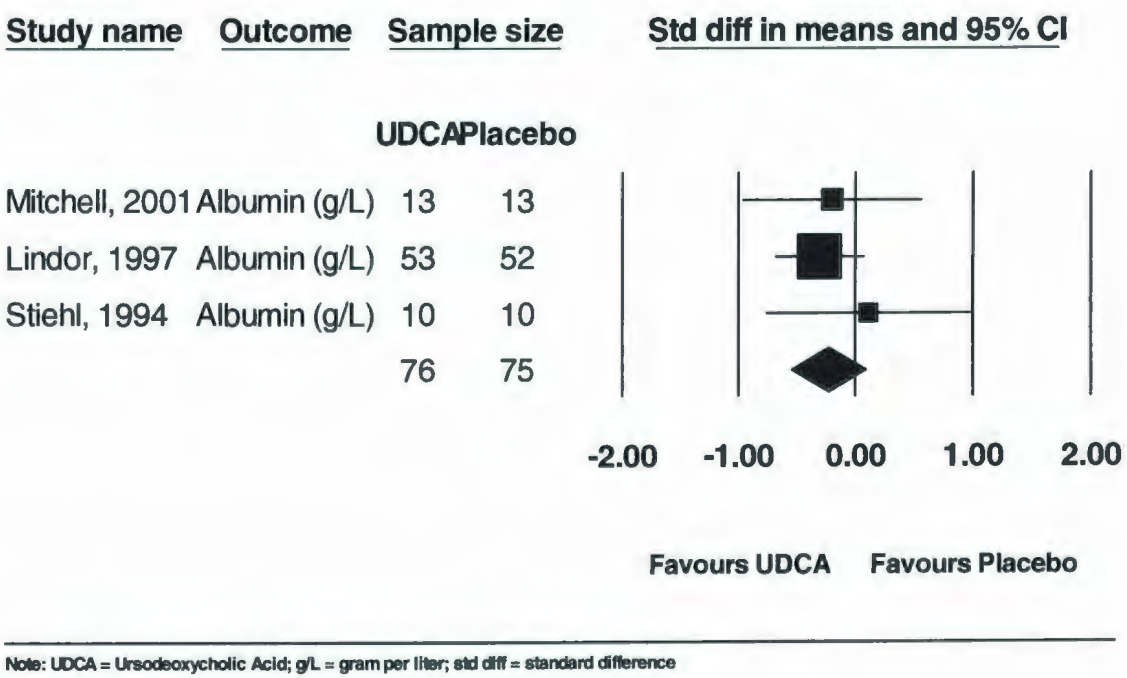
Three of the six fully published trials listed results for the liver function albumin to be analyzed in this meta-analysis (F2,F3,F4). Examining Figure 3.5, we can see that two of the trials (F2,F3) show a treatment effect favoring UDCA; while one trial (F4) demonstrates a treatment effect favoring placebo. Unfortunately, the treatment effect in the trials is small resulting in large 95% CI which cross 1. There are only three trials listing albumin as an outcome measurement and therefore pooling the results of each trial still does not include enough of a treatment effect to reach statistical significance. With a small treatment effect, many more patients would be needed to demonstrate a statistically significant effect. The formal test of heterogeneity (appendix III) for the outcome of these three trials yields a Q-value of 0.187, which with two degrees of freedom (df=2), leads to a p-value of 0.689. This would indicate that the included trials are statistically homogenous and are able to be combined. The pooled estimate of albumin using the standard difference in means of the UDCA group compared to the placebo group was -0.232 using a fixed effects model with a 95% CI (-0.553,0.088) and a non significant p-value of 0.156 (table 3.10).

Table 3.10

Outcome Measurement - Albumin (g/L)

RCT	UDCA (n)	Placebo (n)	Model	SMD	Lower 95% CI	Upper 95% CI	P-value
F2	53	52		- 0.307	-0.691	0.078	0.118
F3	13	13		- 0.200	-0.971	0.571	0.611
F4	10	10		0.113	-0.764	0.990	0.800
Overall	**76**	**75**	**Fixed**	**- 0.232**	**-0.553**	**0.088**	**0.156**

Note: RCT = Randomized controlled trial; UDCA = Ursodeoxycholic acid; n = The total number of patients in the group; SMD = standardized mean difference



Note: UDCA = Ursodeoxycholic Acid; g/L = gram per liter; std diff = standard difference

Figure 3.5: A Forest Plot; Using Standardized Mean Difference and 95% CI, Comparing Albumin (g/L) levels in PSC Patients Receiving UDCA or Placebo.

3.6.6 Bilirubin.

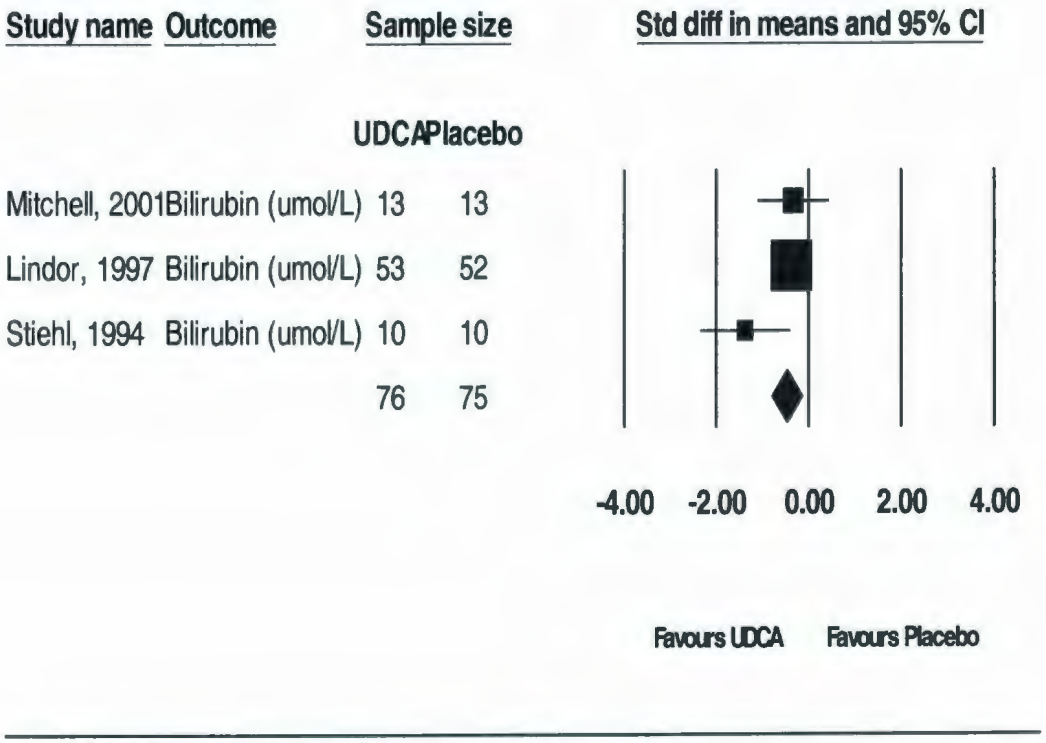
Four of the six fully published trials listed the results for the liver function bilirubin to be analyzed in this meta-analysis (F2,F3,F4,F6). However, one trial (F6) normalized the laboratory values and was not suitable for pooling in this meta-analysis. Examining Figure 3.6, we can see that one of the trials (F4) yields a statistically significant difference for bilirubin levels in favor of the UDCA group. Although two trials (F2,F3) trend toward a difference in bilirubin levels in favor of the UDCA group, the results do not reach statistical significance. The formal test of heterogeneity (appendix III) for the outcome of these three trials yields a Q-value of 3.696, which with two degrees of freedom (df=2), leads to a p-value 0.158. This would indicate the included trials are statistically homogeneous and are able to be combined. The pooled estimate of bilirubin using the SMD of the UDCA group compared to the placebo group was -0.472 using a fixed effects model with a 95% CI (-0.798, -0.147) and a statistically significant p-value of 0.004 (table 3.11).

Table 3.11

Outcome Measurement - Bilirubin (umol/L)

RCT	UDCA (n)	Placebo (n)	Model	SMD	Lower 95% CI	Upper 95% CI	P-value
F2	53	52		- 0.367	-0.752	0.019	0.063
F3	13	13		- 0.329	-1.103	0.445	0.404
F4	10	10		- 1.372	-2.346	-0.398	0.006
Overall	**76**	**75**	**Fixed**	**- 0.472**	**-0.798**	**-0.147**	**0.004**

Note: RCT = Randomized controlled trial; UDCA = Ursodeoxycholic acid; n = Total number of patients in the group; SMD = standardized mean difference



Note: UDCA = Ursodeoxycholic Acid; umol/L = micro mole per liter; std diff = standard difference

Figure 3.6: A Forest Plot; Using Standardized Mean Difference and 95% CI, Comparing Bilirubin (umol/L) levels in PSC Patients Receiving UDCA or Placebo.

3.6.7 AST.

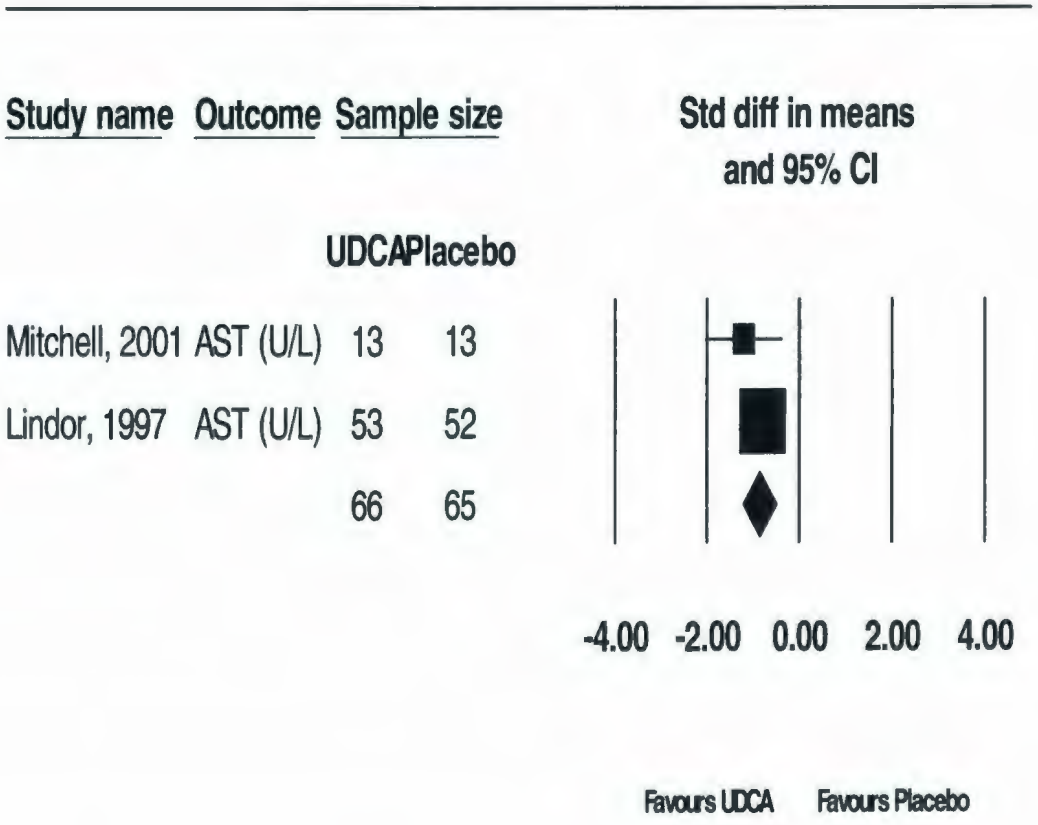
Three of the six fully published trials listed the results for the liver enzyme AST to be analyzed in this meta-analysis (F2,F3,F6); however, one trial (F6) used normalized laboratory values for their analysis and was therefore not suitable for pooling in this meta-analysis. Examining figure 3.7, we can see that the trials show a treatment effect favoring UDCA. The formal test of heterogeneity (appendix III) for the outcome of these two trials yields a Q-value of 0.750, which with one degree of freedom (df=1), yields a p-value of 0.386. This would indicate that the included trials are statistically homogenous and are able to be combined. The pooled estimate of AST using the SMD of the UDCA group compared to the placebo group was -0.868 using a fixed effects model with a 95% CI (-1.227, -0.509) and a statistically significant p-value of 0.000 (table 3.12).

Table 3.12

Outcome Measurement AST (U/L)

RCT	UDCA (n)	Placebo (n)	Model	SMD	Lower 95% CI	Upper 95% CI	P-value
F2	53	52		-0.792	-1.190	-0.395	0.000
F3	13	13		-1.201	-2.036	-0.366	0.005
Overall	**66**	**65**	**Fixed**	**-0.868**	**-1.227**	**-0.509**	**0.000**

Note: AST = Aspartate aminotransferase; RCT = Randomized controlled trial; UDCA = Ursodeoxycholic acid; n= The total number of patients in the group; SMD = Standardized mean difference.



Note: UDCA= Ursodeoxycholic Acid; U/L = units per liter; std diff = standard difference

Figure 3.7: A Forest Plot; Using Standardized Mean Difference and 95% CI, Comparing AST (U/L) levels in PSC Patients Receiving UDCA or Placebo.

3.6.8 ALP.

Five out of the six fully published trials listed the results for the liver enzyme ALP to be analyzed in this meta-analysis (F1,F2,F3,F4,F6); however, one trial (F6) used normalized laboratory values for their analysis and was therefore not suitable for pooling in this meta-analysis. Examining figure 3.8, we can see that two of the four trials (F1,F4) show a treatment effect favoring UDCA. The other two trials (F2,F3) show a trend towards favoring UDCA; however, do not reach statistical significance. This would imply at least some degree of heterogeneity between the groups. In fact, the formal test of heterogeneity (appendix III) for the outcome of these four trials yields a Q-value of 25.192, which with three degrees of freedom (df=3), yields a p-value of 0.000 indicating heterogeneity amongst the trials.

The pooled estimate of ALP using the SMD of the UDCA group compared to the placebo group was -0.822 using a fixed effects model with a 95% CI (-1.153, -0.491) and a statistically significant p-value of 0.000 (table 3.13).

The trials were examined more carefully to try and explain the potential heterogeneity. One potential cause of heterogeneity for this outcome might be the wide range of UDCA (750mg to 20mg/kg/day) used in each study. However, looking at figure 3.8, we can see that all treatment effects show a positive effect with variation in the size of that treatment effect favoring UDCA. This type of heterogeneity is called quantitative heterogeneity and is acceptable for analysis.

The non-significant χ^2 test for heterogeneity was interpreted with caution for the outcome measure ALP (U/L). It has been reported that the statistical test for

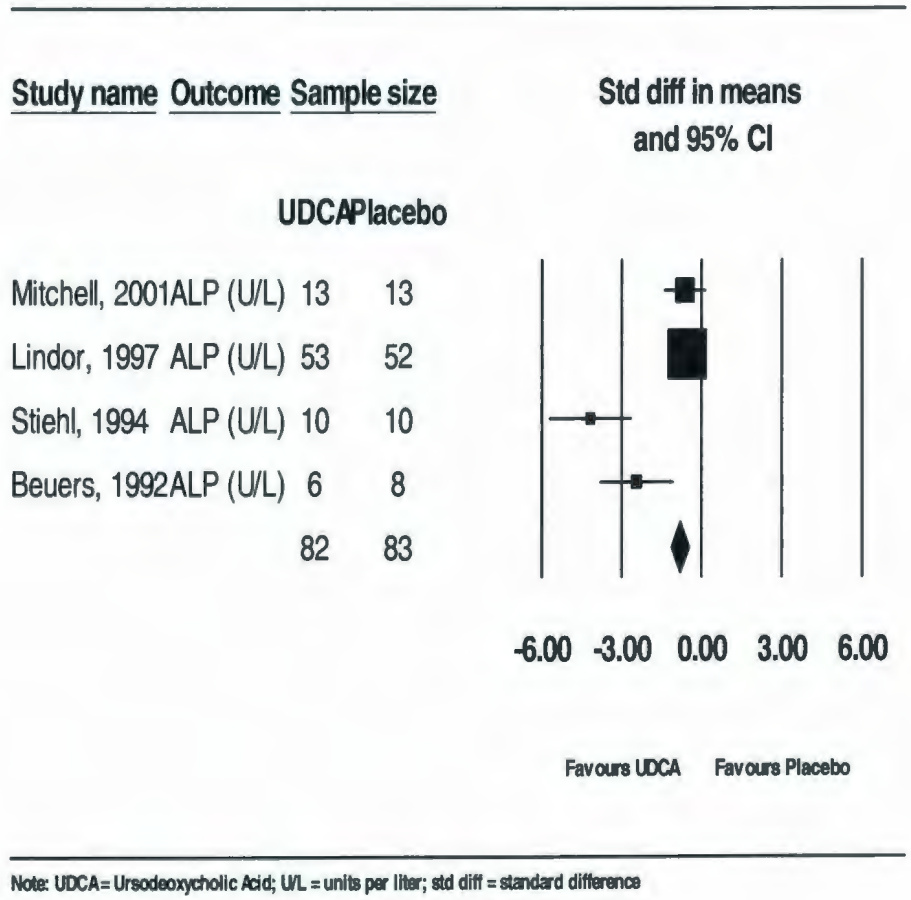
heterogeneity should be interpreted with caution when used in a meta-analysis that has a small number of trials as it may over estimate the overall heterogeneity of the trials (Spector & Thompson, 1991).

Table 3.13

Outcome Measurement - ALP (U/L)

RCT	UDCA (n)	Placebo (n)	Model	SMD	Lower 95% CI	Upper 95% CI	P-value
F1	6	8		-2.441	-3.832	-1.049	0.001
F2	53	52		-0.536	-0.926	-0.147	0.007
F3	13	13		-0.624	-1.411	0.163	0.120
F4	10	10		-4.175	-5.738	-2.612	0.000
Overall	**82**	**83**	**Fixed**	**-0.822**	**-1.153**	**-0.491**	**0.000**

Note: ALP = Alkaline phosphatase; RCT = Randomized controlled trial; UDCA = Ursodeoxycholic acid; n= The total number of patients in the group; SMD = Standardized mean difference.



Note: UDCA = Ursodeoxycholic Acid; U/L = units per liter; std diff = standard difference

Figure 3.8: A Forest Plot; Using Standardized Mean Difference and 95% CI, Comparing ALP (U/L) levels in PSC Patients Receiving UDCA or Placebo.

3.7 Sensitivity Analysis – Publications Bias:

Several outcomes have been analyzed in this meta-analysis. These outcomes included surrogate markers such as liver enzymes, liver function and histology and hard endpoints such as overall mortality and need for a liver transplant. Although we measured the overall mortality rates in this meta-analysis, the results were not statistically significant. This may have been because a difference in mortality does not exist between the UDCA and placebo groups or because the meta-analysis was underpowered to detect a statistically significant difference. A meta-analysis seeks to pool similar studies in order to identify a more accurate treatment effect. In the search for studies, there is always a possibility that non-significant trials have been carried out but remain unpublished. An extensive literature search should help to minimize this problem, commonly referred to as publication bias.

Funnel plots can be used to look for evidence of publication bias. These are graphical representations comparing each trials effect size against a measure of its size such as sample size, standard deviation or standard error (Tonellli, Hackam, & Garg, 2009). Trials with larger sample sizes should better approximate the true treatment effect, whereas small studies have more variation in their estimates of effect size as random variation plays a larger role (Tonellli, Hackam, & Garg, 2009). If a funnel plot is asymmetric towards small positive studies, there is a higher probability that small unpublished negative studies exist which may change the overall effect size of UDCA.

Funnel plots were performed on the outcomes in this meta-analysis that achieved statistical significance. Figure 3.9 demonstrates a funnel plot for the outcome of

bilirubin. The plot demonstrates that the included trials are symmetrical within the funnel plot; however, a smaller positive trial is present which is seen as being skewed towards the bottom left of the plot. In this particular case, there are only three studies used for analysis and therefore it is difficult to interpret the results of publication bias with any certainty.

The outcome measurement ALP was also statistically significant in favor of a treatment effect for UDCA. Figure 3.10, a funnel plot looking for publication bias shows that two smaller studies are skewed to the bottom left of the funnel plot making it possible that small negative unpublished trials have been missed. There are only four studies used for this analysis, three of which contain very few patients and therefore it is difficult to interpret the results of publication bias with any certainty.

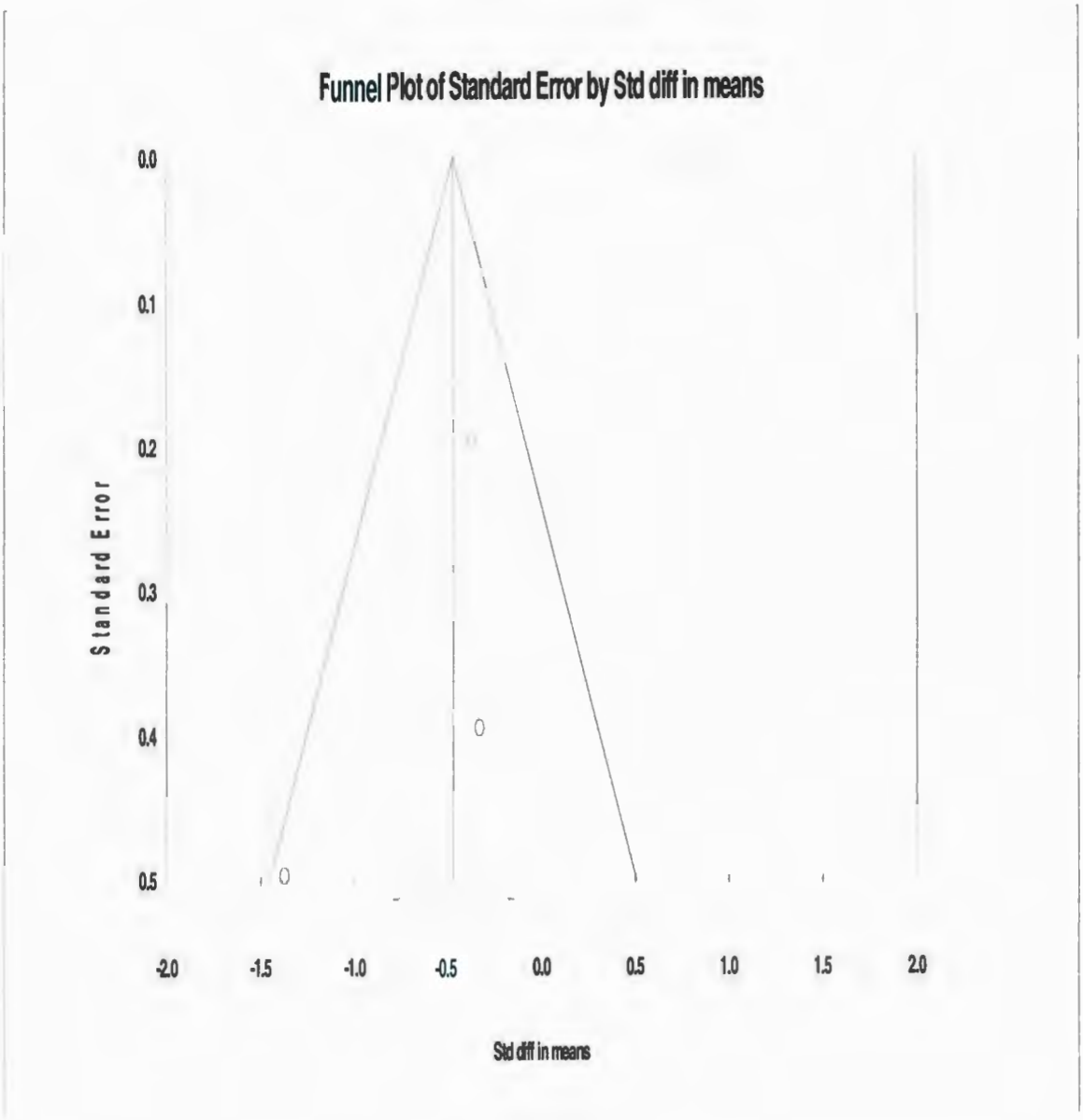


Figure 3.9: Funnel Plot of Standard Error by Standardized Mean Difference, for the Outcome Bilirubin (umol/L) in PSC Patients Receiving UDCA or Placebo.

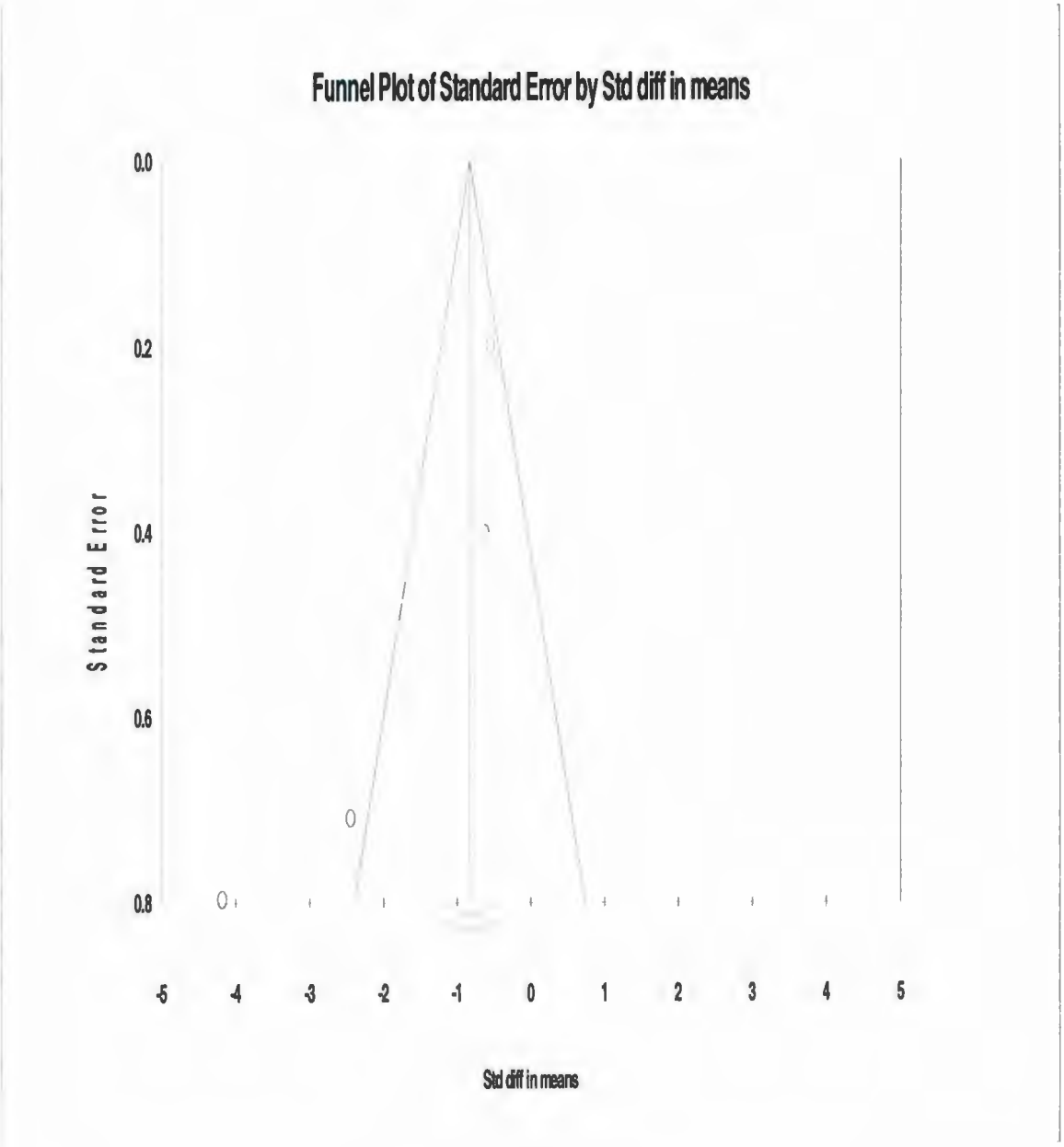


Figure 3.10: Funnel Plot of Standard Error by Standardized Mean Difference, for the Outcome ALP (U/L) in PSC Patients Receiving UDCA or Placebo.

Chapter 4

Discussion

This chapter provides an interpretation for the results of this meta-analysis including an assessment of the usefulness of the Jadad scoring system to determine a study's quality. A discussion of the use of surrogate markers in clinical research will be performed. This chapter will discuss the statistical methods used in meta-analysis with a focus on the use of MD and SMD to measure an outcomes effect size. Literature that was not included in this study but may have been useful and data that was missing from included studies in this meta-analysis will be reviewed.

4.1 Quality of Studies – The Jadad Scoring System

The Jadad scoring system (appendix I) was used to assess each trials quality. This method relies on explicit detail in the methodology section of each publication. This scoring system has some inherent advantages. It is convenient and easy to use as the system only incorporates five questions. An article can be assessed for quality usually in less than ten minutes. However, the Jadad five point scoring system is not without critique. An over simplified scoring system may give the impression that a score of four to five translates into a strong RCT. Generally speaking this may be true, however key methodology may still be lacking. There are no points awarded for power calculations to determine sample size or for describing how the withdrawals and dropouts are treated. Despite these shortcomings, the Jadad system is an accepted standard for quality scoring of RCTs and is widely used.

Quality scoring systems, although helpful, may underestimate the overall quality of an RCT. Due to publication restrictions and limited space in some journals, the authors may be limited in the amount of detail they can include regarding trial design and analysis. Therefore, some trials may be penalized simply because they could not provide the necessary details to obtain a high quality score.

4.2 Interpretation of Meta-Analysis

PSC is a devastating disease affecting young people. It is often fatal without an eventual liver transplantation. Prior to liver transplant, patients are at higher risk for liver decompensation and an increased risk of malignancy. It is not surprising that there has been a lot of interest in pursuing medical treatments that might halt the progression of the disease and delay the time to death or the time to transplant.

This meta-analysis set out to answer the question:

"Does UDCA prolong survival in patients with PSC?"

When research questions are initially formulated, it is very important to be as specific as possible in order to accurately answer the research question being asked.

Immediate concerns about the above question may be:

1. Prolong survival: By how much? What is clinically relevant for a delay in time to death? Some may suggest that any statistically significant difference between treatment groups may be relevant with an outcome such as mortality.

2. Patients with PSC: An argument can be made that using this group as a whole is too broad. All patients with PSC are likely not equal. UDCA may have a different outcome on the disease depending on the patients' disease stage or the severity of liver

function at the time of treatment. Patients may respond differently if they are symptomatic at the time of diagnosis or if they have a more indolent course.

However, in this situation, PSC is such a rare condition that restricting the research question any further would eliminate the ability to identify pertinent trials. It was felt that a more appropriate plan would be to keep the question broad to seek all the available literature on this topic. In this way the results may be more generalizable to a wider group of patients. This would also increase the likelihood of being able to answer a research question on this topic.

This question turned out to be more difficult to answer than expected. PSC is a rare disease and many of the trials that have been published have very small numbers and are underpowered to answer the above question. Whenever a rare disease is studied, there are also fewer studies available in the literature given the small numbers of patients available. There are also potential differences between trials in terms of methodology.

There were differences in the RCT's used for this meta-analysis. The six trials differed in patient recruitment. Three of the trials had under thirty patients (F1,F3,F4); whereas three trials recruited over one hundred patients each (F2,F5,F6). It is obvious that the results of this meta-analysis would be weighted towards the larger trials. However, looking at the point estimate of effect for each trial, except for F6, we can see that although the CI's are large, the point estimate for each trial favors UDCA. Even though F6 showed an effect in favor of placebo, the effect was very small and not statistically significant. The 95% CI's for all six trials overlap. This would indicate that

for our primary outcome, although the trials are not weighted equally, they would appear at first glance to be fairly homogenous.

The RCTs used all cause mortality as an outcome measurement; this could be criticized as being too broad; however, in this particular disease the patients are young and other than IBD, they do not have many co-morbidities. Indeed, when the results are looked at more carefully, all deaths apart from one patient can be directly attributed to PSC and its complications.

Although the initial results for this meta-analysis could not detect a survival benefit for patients with PSC taking UDCA, perhaps a subgroup of patients taking high dose (>15mg/kg/day) UDCA may experience a greater treatment benefit resulting in a prolonged survival. Larger doses of UDCA might be needed to provide sufficient enrichment of the bile acid pool in cholestasis and may enhance the immunomodulatory effects of UDCA (Beuers, et al., 2009). Three trials (F3,F5,F6) measuring mortality as an outcome used high dose UDCA and two trials (F1,F2) used low/standard dosing of UDCA. There was no treatment benefit in either subgroup for the outcome overall mortality. In fact, the one trial (F6) showing a higher mortality in the UDCA group, although not statistically significant, used high dose (>15mg/kg/day) UDCA. Given the very small number of deaths observed, many more patients would need to be recruited to answer the question of survival benefit with UDCA.

Three out of the four trials (F3,F5,F6) listing liver transplant as an outcome measurement used high dose (>15mg/kg/day) UDCA. Although the results were not statistically significant, three of the four trials (F2,F3,F6) display a trend toward an

increased need for liver transplant in the UDCA group. The numbers of patients requiring liver transplant are too small to make any meaningful conclusion for this outcome measurement, however, it is interesting that three of the four trials suggest UDCA may increase a patients need for liver transplantation. Using high dose UDCA for this outcome measurement doesn't appear to add any treatment benefit. Examining these trials more carefully, it becomes apparent that patients were older than the mean age of diagnosis for PSC in all four trials (F2,F3,F5,F6) measuring this outcome. In addition, more than 50% of patients in two of these trials (F2,F3) had advanced histological staging. One trial (F5) did not include histological staging as an outcome measure. Certainly older age and advanced histological staging may signify a poorer prognosis in PSC. Theses reasons may have contributed to the increased need for liver transplant in these studies irrespective of the treatment group and may be contributing factors for the apparent lack of efficacy of UDCA for this outcome measurement.

We suspected that our primary outcome would be difficult to demonstrate with the available literature. A priori, we decided to examine a number of surrogate outcome measurements to help answer our primary research question.

It has been shown that before patients die from liver disease or require a transplant, they deteriorate in terms of liver function (albumin, bilirubin), liver biochemistries (AST, ALP) and histology. Many of the trials measured these variables before and after treatment with UDCA, as surrogate markers for improvement in the underlying liver disease. We therefore decided to use these surrogate markers as outcome measurements in our meta-analysis. Our hypothesis was that if we could detect

a statistically significant improvement in these outcome measurements following treatment with UDCA, then we could infer a treatment benefit in favor of UDCA. The following section will discuss the use of surrogate markers in clinical research.

A change in liver histology was recorded for four of the six trials. Our outcome measurement was worsening of liver histology at the end of each study. The meta-analysis shows no significant difference between UDCA and placebo groups. The largest study measuring this outcome (F2) actually recorded a trend toward worsening of liver histology for the UDCA group, although the results were not statistically significant. The heterogeneity of these trials is likely explained on the basis of sampling error in the histology for PSC. As previously mentioned, this is a patchy disease and the same patient may demonstrate completely different histological stages of disease depending on the area of the liver biopsied (Cullen & Chapman, 2005). Any study that uses histology alone as an end point for this disease should be interpreted with caution.

In order to determine when patients with PSC have advanced disease, there are several validated scoring systems that are based on clinical and biochemical parameters which can be used. The mayo-risk score for liver disease (Kim W. , et al., 2000) and the model for end stage liver disease (MELD) score (Kamath, et al., 2001) are two of the most widely used scoring systems to predict survival. The MELD scores are helpful for prioritizing patients for transplant and are often used as the basis for listing patients for liver transplant. Wiesner, et al; demonstrated the use of the MELD score in liver transplant allocation based on three month survival rates for 3437 patients awaiting liver transplant. Patients having a MELD score less than nine experienced a three-month

survival of 98%, whereas a score of forty experienced a three month survival of only 30% (Wiesner, et al., 2003).

The Mayo-risk score for PSC is as follows:

$R = 0.03(\text{age[y]}) + 0.54 \log_e (\text{bilirubin [mg/dl]}) + 0.54 \log_e (\text{AST [U/L]}) + 1.24(\text{variceal bleeding [0/1]}) - 0.84(\text{albumin[g/dL]})$.

R= risk score

Variceal bleeding: 0= no prior bleed, 1 = prior bleed

Probability of survival at time t years is calculated as $S(t) = S_0(t)^{\exp(R-1.00)}$

Survival function coefficient [$S_0(t)$]

1 year = 0.963, 2 years = 0.919, 3 years = 0.873, 4 years = 0.833

The MELD score is as follows:

MELD = 3.8[Ln serum bilirubin (mg/dL)] + 11.2[Ln INR] + 9.6[Ln serum creatinine (mg/dL)] + 6.4

These validated scoring systems use surrogate markers to try and predict hard outcomes such as mortality and need for liver transplant in patients with PSC. It is therefore appropriate that we chose bilirubin, AST and albumin as surrogate outcome measurements for this meta-analysis. Although not included in the above scoring systems, we have also chosen ALP as an outcome measurement because PSC is a cholestatic disease and ALP is the most affected liver enzyme in cholestasis. This liver

enzyme may not have been included in the above scoring system as the value can be influenced by extra-hepatic biliary obstruction that can occur in PSC, through benign and malignant strictures, and may not be reflective of the severity for the underlying liver disease.

The analysis of albumin in this meta-analysis identifies a trend towards improvement in the UDCA group in patients with PSC; however, statistical significance was not achieved. The trials were determined to be homogeneous for this outcome using the χ^2 test for homogeneity. It is possible that if more patients were recruited for these studies then a statistically significant difference may have been observed given that the treatment effect for most trials was in favor of the UDCA group.

The outcomes for bilirubin, AST and ALP all achieved statistical significance in favor of the UDCA group.

The surrogate markers for our primary outcome showed either a statistically significant benefit in favor of UDCA or a trend towards significance. A larger recruitment of patients or further RCTs may have determined statistical significance for these non-significant outcomes. Whether these surrogate outcomes can reliably predict survival or time to transplant needs to be addressed.

4.3 The Use of Surrogate Markers in Clinical Research

It is becoming more common for RCTs to use surrogate markers for measuring the effect of medical therapy on disease. Ideally outcome measures are clinically firm endpoints such as death or time to an event (eg. transplant, dialysis). Unfortunately, it has become increasingly more difficult to measure treatment efficacy with these "hard"

outcome measurements, especially when a disease is rare (eg. PSC) or the time to an event (death, transplant) is long. In this situation "hard" outcomes are challenging to use because of the large sample size needed and the length of the trial that would need to be conducted. Investigators are always looking for ways to answer their research questions more efficiently and effectively. If it was possible to use a "surrogate" outcome that was easily measured and occurred at an earlier time point than the "hard" outcome, the study might take a shorter time to conduct and might need fewer patients to be recruited.

There has been a lot of research performed using biomarkers as surrogates for hard outcomes. Ideally there should be a strong independent and consistent association between the surrogate end point and the clinical end point (Bucher, Guyatt, Cook, Holbrook, & McAlister, 1999). The use of proteinuria as a surrogate marker for time to dialysis is one example. It has been well validated that continued worsening proteinuria serves as an indicator of impending renal failure and the requirement of dialysis. There should also be RCT evidence that improvement in the surrogate end point is consistently associated with improvement in the target outcome (Bucher, Guyatt, Cook, Holbrook, & McAlister, 1999). Using the same example of proteinuria as a surrogate marker for renal failure, there is evidence that both angiotensin converting enzyme (ACE)-inhibitors and angiotensin receptor blockers (ARBs) are effective at lowering proteinuria and delaying progression to renal failure (Rigatto & Barrett, 2009).

It would appear that using surrogate markers instead of "hard" outcomes for clinical trials is a useful alternative when investigating a disease that is rare or when the outcome of interest takes a long time to develop. Unfortunately, there are many pitfalls

that the researcher needs to be aware of before choosing a surrogate marker. Although surrogate markers may be associated with the disease of interest, this does not imply that they can be used in place of clinically relevant outcomes for a disease. For instance, low levels of high-density lipoprotein (HDL) are associated with worsening atherosclerotic disease. However, in a trial using the drug torcetrapib, HDL levels were increased but this did not translate into a change in progression of coronary artery disease (Rigatto & Barrett, 2009).

Another consideration when using surrogate markers instead of "hard" outcomes is the fact that validation of a surrogate marker for one intervention does not imply that the same surrogate is valid for another intervention. For instance, when statins are investigated, low-density lipoprotein (LDL) cholesterol levels serve as a useful surrogate for cardiovascular disease. However, when a medication called sevelamer was studied, decreased levels of LDL did not translate into lower levels of cardiovascular disease (Rigatto & Barrett, 2009).

It is also possible that not all treatment effects can be accounted for by a single biomarker (Atkinson, et al., 2001). Often multiple markers are needed in combination to account for treatment effects and outcome measurements. An example of this would be the MELD score that uses a combination of markers (bilirubin, INR, creatinine) to approximate liver disease severity in order to help allocate liver transplants. This meta-analysis demonstrates a non-statistically significant trend in favor of an increased need for liver transplant in the UDCA group. The trials did not mention measuring INR or creatinine. Either these were measured and not reported or an alternative model was used

to determine allocation for liver transplant. The mayo risk score has been validated to determine survival for patients with PSC, perhaps this scoring system was used to determine the need for liver transplant in each trial.

Another consideration that must be observed when using surrogate markers is the potential for treatments being studied to cause harmful effects on the outcomes that were not measured. The dopaminergic agent ibopamine had been shown to positively influence such surrogate outcomes for heart failure as ejection fraction and heart rate variability; however, an RCT then demonstrated that the drug actually increased mortality in patients with heart failure (Bucher, Guyatt, Cook, Holbrook, & McAlister, 1999).

In hepatology, a number of surrogate markers used in research and clinical practice have not been properly validated (Gluud, Brok, Gong, & Koretz, 2007). The authors suggest that there are two steps necessary to validate a surrogate marker. The first step is to demonstrate a correlation between the surrogate and the clinical outcome. For instance, there is a strong correlation between serum bilirubin and mortality, which has been validated as one of the strongest surrogate markers in Mayo models (Gluud, Brok, Gong, & Koretz, 2007). The second step is to prove that the interventions effect on the surrogate marker predicts the interventions effect on the hard outcome. The authors suggest that in the case of PBC, despite UDCA showing an improvement in bilirubin levels, this does not translate into an improvement in mortality (Gluud, Brok, Gong, & Koretz, 2007).

Our current meta-analysis attempted to answer the question of whether UDCA prolongs survival in patients with PSC. Surrogate markers of liver histology, liver

biochemistries and liver function were used for clinical outcomes such as mortality and time to transplant. PSC is a patchy disease and therefore it may not be appropriate to use worsening liver histology as a surrogate marker. Liver biopsies have the potential for sampling error and may not accurately reflect the true stage of the disease.

Critically appraising these surrogate markers with the information stated above would suggest that no one individual marker would be sufficient to predict mortality and time to transplant. However, looking at a combination of markers may be more appropriate. The mayo risk score for PSC outlined above, uses surrogate markers (bilirubin, albumin, AST) in addition to the static markers age and history of variceal bleeding to predict survival in patients with PSC. A similar model is used for PBC. One study showed that this model for PSC was validated in 124 patients. The results showed good correlation between the estimated survival using this model and actual survival (Kim W. , et al., 2000). It seems reasonable that the trials evaluating UDCA used AST, bilirubin and albumin as surrogate outcomes for survival in patients with PSC. ALP levels may be influenced by external biliary strictures and would therefore not always be associated with liver dysfunction. There is no conclusive evidence that ALP alone may be used as a reliable surrogate marker when evaluating treatments and predicting disease outcome in patients with PSC.

Using Gluud's review on surrogate markers, the first step in surrogate outcome validation would be fulfilled in that there is a correlation between the surrogate marker and the clinical outcome, as demonstrated by the Mayo score. However, as seen in this current meta-analysis, the second step of surrogate marker validation was not fulfilled.

UDCA's effect on the surrogate markers did not accurately predict UDCA's effect on the hard outcomes. Perhaps there were too few hard outcomes of mortality or need for transplantation to detect a statistically significant difference in treatment groups. Alternatively, treatment may need to occur much earlier in the course of the disease to modify hard outcomes.

4.4 The Use of Mean Difference and Standardized Mean Difference in Meta Analysis

Meta analyses have traditionally used binary outcomes to create pooled effect sizes and CI's. The results are usually reported as pooled odds ratios with 95% CI's. This meta-analysis analyzed data using both binary outcomes (mortality, worsening of liver histology and requirement for liver transplant) and continuous data (liver biochemistries and liver function).

When continuous data are used for outcome measurements, either mean differences (MD) or SMD's can be calculated for each trial and these results can be combined to produce a pooled MD or SMD with 95%CI. However, this type of meta-analysis appears to be more complicated in its statistical pooling of the included trials (Gotzsche, Hrobjartsson, Marie, & Tendal, 2007). According to the Cochrane collaboration, there are several ways to calculate the pooled mean difference using continuous data. The simplest way to pool continuous data is to record the sample size, mean and standard deviation in each group at the beginning and end of the study. The mean difference and standard error can then be calculated. Each trial can be assigned a weight and the inverse variance method of meta-analysis can be used to create a pooled

mean difference along with its CI. In order to accurately combine the data amongst different trials, it is important to ensure that the continuous outcomes from different studies use the same units of measurement. When units of measurement cannot be converted to the same unit, SMD's can be used instead of the MD. The SMD is the difference in means divided by the standard deviation (SD). The SD is the pooled SD of patient outcomes in the whole trial.

There are potential pitfalls to using the SMD in meta-analysis. It is important to ensure that the measurement scales used in each trial are measuring the same outcome. Problems can also occur when inclusion criteria differ between studies, as tighter inclusion criteria may create populations that are more similar resulting in smaller SD. If two or more studies show an equal treatment effect, the SMD might be different amongst these equal studies as the SD might be smaller in those studies using tighter inclusion criteria (Alderson & Green, 2002). Another potential problem with using MD and SMD in meta-analyses is the interpretation of skewed data. Outliers have the ability to skew the mean, ultimately leading to an effect size that may not be truly representative of the data (Alderson & Green, 2002).

Gotzsche, et al; looked at data extraction errors in meta-analyses that use SMD and found many errors occurred during the data extraction process. Errors that this group found included the extraction of standard error (SE) instead of SD. The authors concluded that this would inflate the overall effect estimates.

Data abstractors need to know the direction of the effect size in order for the analysis of continuous data to be accurate. This can become difficult in some settings

where a high score can signify a negative outcome such as certain depression scores (Gotzsche, Hrobjartsson, Marie, & Tendal, 2007). Of the 21 meta-analyses reviewed by Gotzshe, et al; there were errors affecting the outcome of the results in 63% of the studies. One meta-analysis was subsequently retracted and in two studies, a significant difference in results disappeared or appeared. The authors concluded that meta-analyses using MD or SMD for reporting pooled effect sizes should be interpreted with caution. Indeed, the statistics can be more challenging for continuous measurements than for binary outcomes in meta-analysis. However, if more than one person carries out data abstraction and differences are discussed and consensus is reached, one would suspect the meta-analysis would be less likely to have significant error.

It is difficult to interpret the results of meta-analysis using SMD as the effect sizes and CI's are reported in standardized values. As each trial has been weighted differently, there is no set SD that can be used to convert the values back into clinically meaningful MD. Perhaps this type of analysis is better served as a qualitative measure of strength for the treatment effect and can serve as an indicator for treatment effect direction.

4.5 Potential Pitfalls With The Test For Heterogeneity

A meta-analysis cannot be performed until the investigator determines if the studies being examined are similar enough to combine the results. It is the statistical pooling of results that forms the meta-analysis. However, as previously outlined, there are several steps in between that add to the overall quality of a meta-analysis. One essential part of the process is determining whether the results of each study are similar enough to be combined. This is termed "homogeneity" and should be accomplished first

qualitatively by observing the trials similarities (eg. study populations, treatments received, trial duration) and then quantitatively by looking at effect directions and CI for each outcome. Once the trials appear homogenous through the examination of each trial (qualitative analysis), a statistical test for heterogeneity can be performed (quantitative analysis). If the p-value for the χ^2 test of heterogeneity is >0.1, the trials are felt to be similar enough to be combined.

Unfortunately, the overall power for the χ^2 test of heterogeneity can be low under a few different circumstances that may apply to this current meta-analysis. If the overall amount of data that each trial provides is low or when the meta-analysis is heavily weighted by studies with many more included patients than the other studies; such as the RCTs (F2,F5,F6), the test of heterogeneity may have low power and should be interpreted with caution (Hardy & Thompson, 1998).

4.6 The Fixed and Random Effects Models in Meta-Analysis

The fixed effect and random effect methods are two statistical models that can be used to combine data in a meta-analysis. The fixed effects model assumes that the same underlying treatment effect is observed between studies. In this situation there is within study variance to consider but between study variance is felt to be due to random error alone. If the test of heterogeneity is non significant (meaning the outcomes of the studies are statistically similar), then using a fixed effects model for combining studies will give a more precise estimate of treatment effect (Spector & Thompson, 1991). However, if studies are not felt to be homogeneous then a random effects model can be used. A

random effects model takes into account both between study variance and the within study variance (Spector & Thompson, 1991).

One solution to determining the ability of the test of heterogeneity to detect variability between trials is to first perform an analysis with the fixed effects model. If the test of heterogeneity is significant then perform a random effects analysis. If the fixed effects and random effects model provide the same results then the individual studies are more likely to be homogenous and there is adequate power in the χ^2 test for heterogeneity (Alderson & Green, 2002).

This meta-analysis reported fixed effects results when the combined analysis was homogenous. When there was quantitative heterogeneity present then a random effects model was viewed and compared to the results of the fixed effects model; however, ultimately a fixed effects model was chosen in this situation because the small number of trials included in this meta-analysis would limit the ability of a random effects model to detect between trial variation (Borenstein, Hedges, & Rothstein, 2007).

4.7 Publication Bias

Funnel plots were used to graphically demonstrate potential publication bias in this meta-analysis. There has been considerable research in developing statistical methods to help quantify publication bias as well. One method called the failsafe N approach is based on Rosenthal's theory called the "file drawer" theory, meaning that there may be several small non-published studies that may not be accounted for as they remain in someone's file drawer (Rosenthal, 1979). A formula was created that determines how many unpublished studies would be required to make the effect no longer statistically

significant. The problems with this method of assessing publication bias is that the formula assumes that the mean effect size in the missing studies is nil, when in fact the effect size may be positive or negative, thus altering the number of missing studies needed to render the effect size non significant. The Rosenthal method is also based on combining p-values across studies; however, meta-analysis computes p-values for the combined effect, thus decreasing the value of the failsafe N method for detecting publication bias in meta-analysis (Borenstein M. , 2005).

Egger's regression is another method for statistically assessing publication bias. This method employs linear regression to quantify bias captured by the funnel plot. The standard normal deviate is regressed on precision (inverse of the standard error). The size of the treatment effect is captured by the slope of the regression line and bias is captured by the intercept (Borenstein, Hedges, Higgins, & Rothstein, 2005).

These statistical methods for assessing publication bias should only be used when there is a range of studies with different volumes of patients (Borenstein M. , 2005). Although funnel plots can be used to graphically demonstrate publication bias for this meta-analysis, statistical methods to assess publication bias may be misleading.

4.8 Literature That May Have Added To This Meta-Analysis

The literature search identified one trial that may have been useful for this meta-analysis (De Maria, Colantoni, Rosenbloom, & Van Thiel, 1996). This study was reviewed and data extracted by two independent abstractors. Unfortunately all the data needed for entry into the meta-analysis was not available for abstraction. The authors for this study were contacted but didn't answer our request for more data. This section

presents a summary of this trial outlining the potential relevance for our current meta-analysis.

In 1996, DeMaria, et al; published an article titled "UDCA does not improve the clinical course of PSC over a 2 year period" (De Maria, Colantoni, Rosenbloom, & Van Thiel, 1996). This was a RCT including 59 patients with PSC. There were three groups including 20 patients in the UDCA group (300mg/day), 20 patients in the placebo group and 19 patients in a third arm using colchicine. This study was conducted over 24 months with regular three month follow up to assess disease status. At the end of the study period, it was determined that no group was different in regards to liver function or liver injury. The authors concluded that UDCA was no better than colchicine or placebo for PSC. In terms of combinability with our other studies, the average age in the UDCA group was 32 (+/- 5.1) and 31.2 (+/- 5.0) in the placebo group. There were more males present with a ratio of 14/6. Approximately 45% of patients in the study had UC. The average histologic stage of disease at the beginning of the study was 2.2 (+/0.4) in the UDCA group and 2.3 (+/- 0.2) in the placebo group.

The study participants were slightly younger and at an earlier stage of disease than some of the other studies used in this meta-analysis. There are also fewer patients with IBD (UC or crohn's) than the other studies included in this meta-analysis. Further heterogeneity may be explained by the dosage of UDCA used in this study. Based on an average 70 kg male, the dosage used would be 4.2 mg/kg/day, which was well below the dosing used in our current meta-analysis (10-30mg/kg/day). The Jadad scoring system would generate a quality score of three out of five. Points were lost for not describing

how blinding was carried out as well as determining whether all patients were accounted for at the end of the trial. Unfortunately pre and post values were not included for liver biochemistries and function. There was no data available for overall mortality. This study may not have added to our overall results as the study likely used an inadequate dose of UDCA to achieve a clinical difference in study endpoints.

4.9 Incomplete Data From Studies Included In This Meta-Analysis

Two trials (F5 and F6) were included in this study because mortality and requirement for liver transplant data were available, however, no interpretable data was available for the surrogate outcomes measured. The authors for the Olsson study (F5) were contacted but didn't answer our request for more data. This section presents a summary of these trials outlining their potential relevance for our current meta-analysis.

In 2005, Olsson, et al; published an article titled "High dose UDCA in PSC: A 5-year multicenter RCT" (Olsson, et al., 2005). This was an RCT including 219 patients with PSC. There were two groups including 110 patients in the UDCA group (17-23mg/kg/day) and 109 patients in the placebo group. This study was conducted over five years with follow up at six-month intervals. At the end of the study period, it was determined that there was no statistically significant difference between groups in the primary outcome measurement of death or liver transplant. There was a trend for significance in liver enzymes and function. The authors concluded that there was no benefit to using high dose UDCA in patients with PSC. In terms of assessing qualitative homogeneity with our other studies, the average age in the UDCA group was 43.6 (+/- 12.7) and 43.1 (+/- 11.2) in the placebo group. This study included 70% males and 85%

of patients had IBD. There was no histology data for this study. These patient characteristics were similar to the patients included in the other studies for this meta-analysis and would therefore appear appropriate to be used for analysis. The Jadad scoring system generated a quality score of four out of five. One point was lost because details of the randomization process were not explained. Although pre UDCA/placebo values were given for liver biochemistry and liver function, the post results were not explicit and were demonstrated in graph form only. This study's complete results would have been useful for our meta-analysis as there was good methodology used and patients appeared similar to the other studies used for this meta- analysis. This study's data on liver biochemistry and function would have helped to balance the other large study by Lindor, et al; 1997, that was included in the meta-analysis. A subgroup analysis for the surrogate outcomes may have been possible comparing high dose UDCA (>15mg/kg/day) to regular dosing (10-15mg/kg/day). This study did include mortality and liver transplant data that could be used for the meta-analysis. To date, this study is the largest one published with the longest follow up period comparing UDCA to placebo in patients with PSC. A complete meta-analysis has to be interpreted with caution if this trials data on liver enzymes and function have not been included in its results. Despite attempts at contacting the authors, no response was received.

In 2009, Lindor, et al; published an article titled "High-Dose Ursodeoxycholic Acid for the Treatment of Primary Sclerosing Cholangitis" (Lindor, et al., 2009). This was an RCT including 150 patients with PSC. There were two groups including 76 patients in the UDCA group (28-30mg/kg/day) and 74 patients in the placebo group.

This study was designed to follow patients with PSC for five years; however, ended after six years of patient recruitment secondary to futility. At the end of the study period, it was determined that 39% of the patients in the UDCA group versus 26% of patients in the placebo group had reached one of the pre-established primary endpoints (table 3.5). After adjusting for baseline characteristics, the authors determined that the risk of death, liver transplant or minimal listing criteria was two times greater for the patients in the UDCA group. Although there was a trend for increased mortality and requirement for liver transplant in the UDCA group, these results did not reach statistical significance. At three years, there was a statistically significant difference between liver biochemistry and function favoring UDCA. The authors concluded that although long term, high dose UDCA improved serum liver enzymes and function, this did not translate into an improved survival for patients with PSC. Trial similarities were assessed with our other included studies. The average age in the UDCA group was 46.6 and 45.3 in the placebo group. This study included 57% males and 77% of patients with IBD. The dosage of UDCA used in this study was 28-30 mg/kg/day, which is the highest dosage used of all included trials in this meta-analysis. This study's mean age was higher than the majority of our included trials and higher than the mean age of diagnosis for patients with PSC. This study included more females with PSC than the majority of our included trials and a higher percentage of females than what would be expected from the general population with PSC. These baseline characteristics would suggest that this trial showed a degree of heterogeneity compared to the other included trials in this meta-analysis and perhaps this trial may not be generalizable to a typical population with PSC.

Liver biopsies were taken before randomization and after five years in 31 patients. Six patients in the UDCA group and four patients in the placebo group developed cirrhosis. Unfortunately, the authors did not provide any further data on worsening of liver histological stage for both treatment groups. These results could not be used in our meta-analysis as they may underestimate the number of patients who developed histologic progression in liver disease as patients can develop worsening in histological stage without necessarily developing cirrhosis.

Although there was information provided on pre/post UDCA and placebo values for liver biochemistry and liver function, the results were displayed in normalized values and were therefore not suitable for pooling with the other included RCTs. Interestingly, the results show that liver biochemistry and function improved in the placebo group from baseline. This observation would not be expected from the natural history of disease in PSC and was not observed in other trials using high dose UDCA.

This study's data on liver histology and liver biochemistry/function would have helped to balance the other large study by Lindor; et al from 1997, that was included in this meta-analysis. A subgroup analysis may have been possible comparing high dose UDCA (>15mg/kg/day) to regular dosing (10-15mg/kg/day). This study did include mortality and liver transplantation data that could be used for the meta-analysis. Interestingly, this study was the only one that showed a trend toward increased mortality for the UDCA group. The study authors suggested that UDCA may modulate apoptosis. Perhaps a higher dose of UDCA prevented apoptosis of activated stellate cells, which continued to be active in fibrogenesis ultimately leading to deterioration in liver disease

(Lindor, et al., 2009). The authors also suggested that higher doses of UDCA may cause unabsorbed medication to enter the colon and be modified into hepatotoxic bile acids. These theories may explain the results of this particular study, although the other studies using high dose UDCA (F3,F5) did not demonstrate similar findings.

4.10 Ursodeoxycholic Acid in Primary Sclerosing Cholangitis: A Comparison of Two Meta-Analyses

In 2009, Shi et al; published a meta-analysis titled "Ursodeoxycholic Acid in Primary Sclerosing Cholangitis: Meta-Analysis of Randomized Controlled Trials (Shi, Li, Zeng, Lin, & Xie, 2009). This was a meta-analysis that included eight studies. Six of the included studies (Beuers, et al., 1992), (Stiehl, Walker, Stiehl, Rudolph, Hofmann, & Theilmann, 1994), (De Maria, Colantoni, Rosenbloom, & Van Thiel, 1996), (Lindor, 1997), (Mitchell, Bansi, Hunt, Von Bergmann, Fleming, & Chapman, 2001), and (Olsson, et al., 2005), were fully published and two were in abstract form (Lo, et al., 1992), (Bansi, Christie, Fleming, & Chapman, 1996).

In comparison, our current meta-analysis included five of the eight trials used in the Shi, et al., 2009 study and one RCT that was not included in the meta-analysis by Shi. The trials by (Beuers, et al., 1992), (Stiehl, Walker, Stiehl, Rudolph, Hofmann, & Theilmann, 1994), (Lindor, 1997), (Mitchell, Bansi, Hunt, Von Bergmann, Fleming, & Chapman, 2001), (Olsson, et al., 2005) were included in this current meta-analysis, however, the study by (De Maria, Colantoni, Rosenbloom, & Van Thiel, 1996) was excluded as there was no placebo group. We chose only placebo controlled trials for our meta-analysis in order to assess the true effect of Ursodeoxycholic acid. Combining

placebo controlled trials and non-placebo controlled trials could lead to unwanted heterogeneity.

Our current meta-analysis decided a priori to pool published abstracts separately and include them in a sensitivity analysis if adequate data were available for abstraction. The abstract by Lo, et al., 1992 was reviewed in our current meta-analysis, however, there was insufficient information included in the abstract for analysis. The abstract by Bansi, et al., 1996 was also reviewed and excluded from our current meta-analysis as it was felt to have been published later as a full RCT. This would have implications if patients included in the meta-analysis were recorded twice.

Our current meta-analysis also included the latest RCT by Lindor, et al., 2009, which was not available for the meta-analysis by Shi, et al., 2009. The Lindor, et al., 2009 study was of high methodological quality and included a large number of patients. A meta-analysis that does not include this RCT would have to be interpreted with caution.

The primary and secondary outcomes were not clearly defined in the study by Shi, et al., however, they did include the outcomes assessed in our current meta-analysis. Forrest plots were not included for graphical demonstration of results in the study by Shi, et al., however, OR and 95% CI along with p-values were shown. In our current study, statistical analysis was displayed using OR along with 95% CI for binomial outcomes and with SMD along with 95% CI for continuous outcome measurements. Forest plots were used for graphical demonstration of the results.

The meta-analysis by Shi, while it included more RCTS than our meta-analysis, must be interpreted with some caution. The trials included in a meta-analysis should be homogeneous in terms of design and the results which are available for analysis. Combining trials with differing designs and with limited data available for analysis may lead to spurious results. A failure to include the most up to date trials may lead to out of date conclusions.

Chapter 5

Conclusion

This chapter is divided into two sections. The first part presents the educational lessons that have been observed during this research endeavor. The second part focuses on the particular conclusions that are relevant to this meta-analysis.

At the time this study was initially conducted, there were few meta-analyses using continuous variables. Performing the meta-analysis using SMD and investigating the potential complications and challenges when such an analysis is performed was educational.

Reported results using continuous variables need to be interpreted with caution given the difficulty in such a statistical analysis and the many possible errors encountered in combining such data.

Statistical tests for heterogeneity are not the only way to assess homogeneity amongst trials and the test itself needs to be interpreted with caution especially when the trials combined have few patients.

However, meta-analysis is still very useful in research. This type of study can help answer questions on medical management when there are several well-conducted trials with non-significant results, usually because each trial is underpowered. The results can be pooled to achieve an overall effect size for the outcome being observed in the hopes of achieving statistical significance.

This meta-analysis would suggest that there are significant differences between the UDCA and placebo group for the surrogate outcomes bilirubin, ALP and AST. If

further data were available from the trials, we may have been able to calculate average Mayo risk scores for each trial and then pool these scores to get a better assessment of the usefulness of UDCA in PSC. Combining surrogate markers in this fashion (i.e. using a previously validated scoring system) may have been more appropriate to help answer our research question.

Alternatively, if individual patient data were available, the results may have been combined for each outcome measurement to increase the strength of this study; as opposed to combining individual study results for each outcome. The acquisition of individual patient level data may have increased the ability to infer a treatment effect. However, when combining individual patient results across studies, continuous outcome measurements such as laboratory values would have to be standardized in each study to ensure the validity of combining the data.

When all cause mortality was analyzed on its own and by the subgroup of UDCA dosage, there was no significant difference between treatment groups. A conclusion cannot be made whether UDCA can prolong survival in patients with PSC. Perhaps with more patients a treatment effect may have been seen. PSC affects young people and has devastating complications including liver failure and malignancy. It will be important for research in this area to continue. Further study may need to be performed to accurately answer our current research question. Perhaps if UDCA is used earlier in the course of the disease, more favorable outcomes may be observed. After the 2009 Lindor study, it is less likely that further research will be done using high dose UDCA in PSC. Three high dose UDCA trials (F3,F5,F6) have failed to demonstrate an improvement in overall

survival and time to transplant. Although the latest study (F6) may not appear to be generalizable to a typical population of patients with PSC, the study was well performed with good methodology. The study demonstrated worse outcomes in the UDCA group, although these results were not statistically significant. Future research may focus on alternative treatment strategies to slow the progression of disease in patients with PSC.

After a careful review of the literature, it would appear that further research is needed to explore the role of surrogate markers for clinical outcomes in the field of hepatology. Although scoring systems such as the Mayo risk score have shown a correlation with clinical outcomes in PSC, this has not been confirmed in clinical trials. Perhaps, these clinical trials had too few outcomes to detect a statistically significant difference in clinical outcome measurements. The other possibility is that current surrogate markers being used in PSC such as liver biochemistry and function may not be effective for detecting hard clinical outcomes such as death and time to transplant.

Bibliography

References marked with an asterisk indicate studies included in the meta-analysis.

Alderson, P., & Green, S. (2002, November). Cochrane Collaboration: Open Learning Material for Reviewers. Retrieved November 25, 2009, from cochrane collaboration: http://cochrane-net.org/openlearning/HTML/modA1.htm

Angulo, P. (2002). Use of Ursodeoxycholic Acid in Patients with Liver Disease. Current Gastroenterology Reports , 4, 37-44.

Atkinson, J. M., Colburn, W. D., DeMets, D., Downing, G., Hoth, D., Oates, J., et al. (2001). Biomarkers and Surrogate Endpoints: Preferred Definitions and Conceptual Framework. Clinical Pharmacology and Therapeutics , 69 (3), 89-95.

Bambha, K., Kim, W., Talwalkar, J., Torgerson, H., Benson, J., Therneau, T., et al. (2003). Incidence, Clinical Spectrum, and outcomes of Primary Sclerosing Cholangitis in A United States Community. Gastroenterology , 125 (5), 1364-9.

Bansi, D., Chapman, R., & Fleming, K. (1996). Antineutrophil cytoplasmic antibodies in chronic liver disease: Prevalence, titre, specificity and IgG subclass. Journal of Hepatology , 24, 581-86.

Bansi, D., Christie, J., Fleming, K., & Chapman, R. (1996). High-Dose Ursodeoxycholic Acid in Primary Sclerosing Cholangitis: A randomized double blind placebo controlled trial. Gastroenterology , A1146.

Berger, V. (2006). Is the Jadad Score the Proper Evaluation of Trials. Journal of Rheumatology , 33 (8), 1710-12.

Bergquist, A., Montgomery, S., Bahmanyar, S., Olsson, R., Danielsson, A., Lindren, S., et al. (2008). Clinical Gastroenterology and Hepatology. 6 (8), 939-43.

Beuers, U., Boberg, K., Chapman, R., Chazouilleres, O., Invernizzi, P., Jones, D., et al. (2009). EASL Clinical Practice Guidelines: Management of Cholestatic Liver Diseases. *Journal of Hepatology , 51*, 237-267.

*Beuers, U., Spengler, U., Kruis, W., Ademir, U., Wiebecker, B., Heldwein, W., et al. (1992, April). Ursodeoxycholic Acid For Treatment of Primary Sclerosing Cholangitis: A Placebo-controlled Trial. Hepatology , 707-13.

Borenstein, M. (2005). Softwared for Publication Bias. In H. Rothstein, A. Sutton, & M. Borenstein, Publication Bias in Meta-Analysis - Prevention, Assessment and Adjustments (pp. 193-220). John Wiley and Son, Ltd.

Borenstein, M., Hedges, L., & Rothstein, H. (2007). *Meta-Analysis: Fixed Effect vs Random Effects.* (c. m. analysis, Producer) Retrieved 2010, from www.meta-analysis.com:www.sciencedownload.net/demodownload/fixed%20effect%20vs.%20random%20effects.pdf

Borenstein M, Hedges L, Higgins J, Rothstein H. Comprehensive Meta-Analysis Version2. Engelwood, NJ, Biostat; 2005.

Bucher, H., Guyatt, G., Cook, D., Holbrook, A., & McAlister, F. (1999). Users Guides to the Medical Literature: Applying Clinical Trial Results: How to use an article measuring the effect of an intervention on surrogate end points. Journal of the American Medical Association , 282 (8), 771-778.

Chapman, R. (2003). The Management of Primary Sclerosing Cholangitis. Current Gastroenterology Reports , 5, 9-17.

Chen, W., & Gluud, C. (2003). Bile Acids For Primary Sclerosing Cholangitis . The Cochrane Library (2), 1-36.

Cullen, S., & Chapman, R. (2005). Review Article: Current Management of Primary Sclerosing Cholangitis. Alimentary Pharmacology and Therapeutics , 21, 933-948.

Cullen, S., Rust, C., Flemming, K., Edwards, C., Beuers, U., & Chapman, R. (2008). High Dose Ursodeoxycholic Acid For the Treatment of Primary Sclerosing Cholangitis is Safe and Effective. Journal of Hepatology , 48 (5), 792-800.

De Maria, N., Colantoni, A., Rosenbloom, E., & Van Thiel, D. (1996). Ursodeoxycholic Acid Does Not Improve the Clinical Course of Primary Sclerosing Cholangitis Over a Two Year Period. Hepato-Gastroenterology , 43 (12), 1472-1479.

Farges, O., Malassagne, B., Sebagh, M., & Bismuth, H. (1995). Primary Sclerosing Cholangitis: Liver Transplantation or Biliary Surgery. Surgery , 117 (2), 146-55.

Farrant, J., Hayllar, K., Wilkinson, M., Karani, J., Portmann, B., Westaby, D., et al. (1991). Natural History and Prognostic Variables in Primary Sclerosing Cholangitis. Gastroenterology , 100 (6), 1710-7.

Gautam, M., Cheruvattath, R., & Balan, V. (2006). Recurrence of Autoimmune Liver Disease After Liver Transplantation: A Systematic Review. Liver Transplantation, 12, 1813-24.

Gluud, C., Brok, J., Gong, Y., & Koretz, R. (2007). Hepatology may have problems with putative surrogate outcome measures. *Journal of Hepatology , 46,* 734-742.

Gong, Y., Huang, Z., Christensen, E., & Gluud, C. (2007). Ursodeoxycholic Acid For Patients With Primary Bilary Cirrhosis: An updated systematic review and meta analysis of randomized clinical trials using bayesian approach as sensitivity analyses. American Journal of Gastroenterology , 102 (8), 1799-1807.

Gotzsche, P., Hrobjartsson, A., Marie, K., & Tendal, B. (2007). Data Extraction Errors in Meta Analyses That Use Standardized Mean Differences. Journal of the American Medical Association , 298 (4), 430-437.

Goulis, J., Leandro, G., & Burroughs, A. (1999). Randomised Controlled Trials Of Ursodeoxycholic Acid Therapy for Primary Biliary Cirrhosis: A meta analysis. Lancet , 354 (9184), 1053-60.

Grant, A., Lalor, P., Hubscher, S., Briskin, M., & Adams, D. (2001). MAdCAM-1 expressed in Chronic Inflammatory Liver Disease Supports Mucosal Lymphocyte Adhesion to Hepatic Endothelium. Hepatology , 33, 1065-72.

Gruchy, S., & Fardy, J. (2003). A Meta-Analysis of Ursodeoxycholic Acid For the Treatment of Primary Sclerosing Cholangitis. 68th Annual Scientific Meeting and Postgraduate Course (p. 230). Baltimore: American College of Gastroenterology.

Hardy, R., & Thompson, S. (1998). Detecting and Describing Heterogeneity in Meta-Analysis. Statistics in Medicine , 17, 841-856.

Harnois, D., Angulo, P., Jorgensen, R., LaRusso, N., & Lindor, K. (2001). High-Dose Ursodeoxycholic Acid as a Therapy for Patients with Primary Sclerosing Cholangitis. The American Journal of Gastroenterology , 96 (5), 1558-1562.

Haynes, R., Sackett, D., Guyatt, G., & Tugwell, P. (2008). Clinical Epidemiology (Vol. 88). Hamilton: Lippincott Williams and Wilkins.

Jadad, A., Moore, R., Carroll, D., Jenkinson, C., Reynolds, D., Gavaghan, D., et al. (1996). Assessing the Quality of Reports of Randomized Clinical Trials: Is Blinding Necessary? Controlled Clinical Trials , 17 (1), 1-12.

Janes, C., Dickson, E., Okazaki, R., Bonde, S., McDonagh, A., & Riggs, B. (1995). Role of hyperbilirubinemia in the impairement of osteoblast proliferation associated with cholestatic jaundice. Journal of Clinical Investigation , 95, 2581-2586.

Kamath, P., Wiesner, R., Malinchoc, M., Kremers, W., Therneau, T., Kosberg, C., et al. (2001). A model to predict survival in patients with end-stage liver disease. Hepatology , 33 (2), 464-70.

Kaplan, G., Laupland, K., Butzner, D., Urbanski, S., & Lee, S. (2007). The Burden of Large and Small Duct Primary Sclerosing Cholangitis in Adults and Children: A Population-Based Analysis. American Journal of Gastroenterology , 102 (5), 1042-9.

Kim, W., Jorgemsen, R., Malinchoc, M., Benson, J., Dickson, E., & Lindor, K. (1997). Are there Patients with Primary Sclerosing Cholangitis in Whom Ursodeoxycholic Acid May Be Beneficial? Gastroenterology , A1301.

Kim, W., Therneau, T., Wiesner, R., Poterucha, J., Benson, J., Malinchoc, M., et al. (2000). A revised natural history model for primary sclerosing cholangitis. Mayo Clinic Proceedings , 75 (7), 688-94.

Knox, T., & Kaplan, M. (1994). A double blind controlled trial of oral pulse methotrexate therapy in the treatment of primary sclerosing cholangitis. Gastroenterology , 106 (2), 494-499.

L'abbe, K., Detsky, A., & O'Rourke, K. (1987). Meta-Analysis in Clinical Research. Annals of Internal Medicine , 107, 224-233.

LaRusso, N., Wiesner, R., & Ludwig, J. (1988). Prospective Trial of Penicillamine in Primary Sclerosing Cholangitis. Gastroenterology , 95, 1036-1042.

Lazaridis, K., Gores, J., & Lindor, K. (2001). Ursodeoxycholic Acid Mechanisms of Action and Clinical Use in Hepatobiliary Disorders. Journal of Hepatology , 35, 134-146.

Lee, Y., & Kaplan, M. (1995). Primary Sclerosing Cholangitis. New England Journal of Medicine , 332 (14), 924-33.

Lindblad, A., Glaumann, H., & Strandvik, B. (1998). A Two-Year Prospective Study of the Effect of Ursodeoxycholic Acid on Urinary Bile Acid Excretion and Liver Morphology in Cystic Fibrosis-Associated Liver Disease. Hepatology , 27 (1), 166-174.

*Lindor, K. (1997). Ursodiol For Primary Sclerosing Cholangitis. New England Journal of Medicine, 336 (10), 691-5.

Lindor, K., Jorgensen, R., & Anderson, M. (1996). Ursodeoxycholic Acid and Methotrexate for Primary Sclerosing Cholangitis: A Pilot Study. American Journal of Gastroenterology , 91 (3), 511-515.

*Lindor, K., Kowdley, K., Luketic, V., Harrison, M., McCashland, T., Befeler, A., et al. (2009). High-Dose Ursodeoxycholic Acid for the Treatment of Primary Sclerosing Cholangitis. *Hepatology , 50* (3), 808-814.

Lo, S., Hermann, R., Chapman, R., Fleming, K., Shearman, J., Cusick, P., et al. (1992). Ursodeoxycholic Acid in Primary Sclerosing Cholangitis: A double blind placebo controlled trial. Hepatology , 16:92A.

Ludwig, J., Barham, S., LaRusso, N., Elveback, L., Wiesner, R., & McCall, J. (1981). Morphologic Features of Chronic Hepatitis Associated With Primary Sclerosing Cholangitis and Chronic Ulcerative Colitis. Hepatology , 1 (6), 632-40.

MacCarty, R., LaRusso, N., Wiesner, R., & Ludwig, J. (1983). Primary Sclerosing Cholangitis: Findings on Cholangiography and Pancreatography. Radiology , 149, 39-44.

Michaels, A., & Levy, C. (2008, March 12). The Medical Management of Primary Sclerosing Cholangitis. Retrieved March 12, 2008, from Medscape: www.pubmedcentral.nih.gov

Mitchell, S., & Chapman, R. (1997). Review Article: The Management of Primary Sclerosing Cholangitis. Alimentary Pharmacology and Therapeutics , 11, 33-43.

Mitchell, S., & Chapman, R. (1998). The Management of Primary Sclerosing Cholangitis. Clinics in Liver Disease , 2 (2), 353-372.

*Mitchell, S., Bansi, D., Hunt, N., Von Bergmann, K., Fleming, K., & Chapman, R. (2001). A Preliminary Trial of High-Dose Ursodeoxycholic Acid in Primary Sclerosing Cholangitis. Gastroenterology , 121, 900-907.

Myburgh, J. (1994). Surgical Biliary Drainage in Primary Sclerosing Cholangitis: The role of the Hepp Couinaud Approach. Archies of Surgery , 129 (10), 1057-62.

O'Brien, C., Senior, J., Arora-Mirchandani, R., Batta, A., & Salen, G. (1991). Ursodeoxycholic Acid For the Treatment of Primary Sclerosing Cholangitis: A 30-month pilot study. Hepatology , 14 (5), 838-847.

O'Brien, S., Craig, P., & Hatfield, A. (1993). The Effect of Ursodeoxycholic (UDCA) Treatment in Primary Sclerosing Cholangitis (PSC) - Results of a Pilot Study. Gastroenterology , 104: A966.

*Olsson, R., Boberg, K., De Muckadell, O., Lindgren, S., Hultcrantz, R., Folvik, G., et al. (2005). High-Dose Ursodeoxycholic Acid in Primary Sclerosing Cholangitis: A 5-Year Multicenter, Randomized, Controlled Study. Gastroenterology , 129, 1464-1472.

Olsson, R., Broome, U., Danielsson, A., Haderstrand, I., Jarnerot, G., Loot, L., et al. (1995). Colchicine treatment of Primary Sclerosing cholangitis. Gastroenterology, 108, 1199-1203.

Palma, J., Reyes, H., Ribalta, J., Hernandex, I., Sandoval, l., Almuna, R., et al. (1997). Ursodeoxycholic Acid in the Treatment of Cholestasis of Pregnancy: A Randomized, Double Blind Study Controlled with Placebo. Journal of Hepatology, 27 (6), 1022-8.

Paumgartner, G., & Beuers, U. (2002, September). Ursodeoxycholic Acid in Cholestatic Liver Disease: Mechanisms of Action and Therapeutic Use Revisited. Hepatology, 525-531.

Podda, M., Ghezzi, C., Battezzati, P., Bertolini, E., Crosignani, A., Petroni, M., et al. (1989). Effect of Different Doses of Ursodeoxycholic Acid in Chronic Liver Disease. Difestive Diseases and Sciences , 34 (12), 59S-65S.

Portincasa, P., Vacca, M., Moschetta, A., Petruzzelli, M., Palasciano, G., Van Erpecum, K., et al. (2005). Primary Sclerosing Cholangitis: Updates in Diagnosis and Therapy. World Journal of Gastroenterology , 11 (1), 7-16.

Poupon, R., Balkau, B., Eschwege, E., & Poupon, R. (1991). A multicenter controlled trial of ursodiol for the treatment of primary biliary cirrhosis. New England Journal of Medicine (324), 1548-1554.

Poupon, R., Lindor, K., Cauch-Dudek, K., Dickson, E., Poupon, R., & Heathcote, E. (1997). Combined Analysis of Randomized Controlled Trials of Ursodeoxycholic Acid in Primary Biliary Cirrhosis. Gastroenterology , 113, 884-890.

Rigatto, C., & Barrett, B. (2009). Biomarkers and Surrogates in Clinical Studies. In P. Parfrey, & B. Barrett, Methods in Molecular Biology: Clinical Epidemiology: Practice and Methods (pp. 137-154). St. John's, NL: Humana Press.

Rodriques, C., Fan, G., & Wong, P. (1998). Ursodeoxycholic Acid May Inhibit Deoxycholic Acid-Induced Apoptosis by Modulating Mitochondrial Transmembrane Potential and Reactive Oxygen Species Production. Molecular Medicine , 4, 165-178.

Rosen, C., Nagorney, D., Wiesner, R., Coffey, R., & LaRusso, N. (1991). Cholangiocarcinoma Complicating Primary Sclerosing Cholangitis. Annals of Surgery , 213 (1), 21-5.

Rosenthal, R. (1979). The File-drawer problem and tolerance for null results. *Psychological Bulletin , 86,* 638-641.

Sacks, H., Berrier, J., Reitman, D., Ancona-Berk, V., & Chalmers, T. (1987). Meta-analysis of Randomized Controlled Trials. The New England Journal of Medicine, 316, 450-5.

Shi, J., Li, Z., Zeng, X., Lin, Y., & Xie, W. (2009). Ursodeoxycholic Acid in Primary Sclerosing Cholangitis: Meta-Analysis of Randomized Controlled Trials. Hepatology Research , 39 (9), 865-73.

Shi, J., Wu, C., Lin, Y., Chen, Y., Zhu, L., & Xie, W. (2006). Long Term Effects of Mid Dose Ursodeoxycholic Acid In Primary Biliary Cirrhosis: A Meta-Analysis of Randomized Controlled Trials. American Journal of Gastroenterology , 101 (7), 1529-38.

Silveira, M., & Lindor, K. (2008). Clinical Features and Management of Primary Sclerosing Cholangitis. World Journal of Gastroenterology , 14 (21), 3338-3349.

Silveira, M., & Lindor, K. (2008). Primary Sclerosing Cholangitis. Canadian Journal of Gastroenterology , 22 (8), 689-698.

Siqueira, E., Schoen, R., Silverman, W., Martin, J., Rabinovitz, M., Weissfeld, J., et al. (2002). Detecting Cholangiocarcinoma in Patients with Primary Sclerosing Cholangitis. Gastrointestinal Endoscopy , 56 (1), 40-7.

Spector, T., & Thompson, S. (1991). The potential and limitations of meta-analysis. Journal of Epidemiology and Community Health , 45 (2), 89-92.

*Stiehl, A., Walker, S., Stiehl, L., Rudolph, G., Hofmann, W., & Theilmann, L. (1994). Effect of Ursodeoxycholic Acid on Liver and Bile Duct Disease in Primary Sclerosing Cholangitis. A 3-year pilot study with a placebo controlled study period. Journal of Hepatology , 20, 57-64.

Stielh, A., Rudolph, G., Kloters-Plachky, P., Sauer, P., & Walker, S. (2002). Development of Domiant Bile Duct Stenoses in Patients with Primary Sclerosing Cholangitis Treated with Ursodeoxycholic Acid: Outcome After Endoscopic Treatment. Journal of Hepatology , 36 (2), 151-6.

Tint, G., Salen, G., & Colalillo, A. (1982). Ursodeoxycholic Acid: A safe and Effective Agent for Dissolving Cholesterol Gallstones. Annals of Internal Medicine , 97, 351-731.

Tonelli, M., Hackam, D., & Garg, A. (2009). Primer on Systematic Review and Meta-Analysis. In B. Parfrey, & B. Barrett, Clinical Epidemiology: Practice and Methods (Vol. 473, pp. 217-233). St. John's: Humana Press.

Van De Meeberg, P., Houwen, R., Sinaasappel, M., Heijerman, H., Bijleveld, C., & Vanberge-Henegouwen, G. (1997). Low Dose versus High Dose Ursodeoxycholic Acid in Cystic Fibrosis-Related Cholestatic Liver Disease. Scandinavian Journal of Gastroenterology , 32 (4), 369-373.

Van De Meeberg, P., Wolfhagen, F., Van Berge-Henegouwen, G., Salemans, J., Tangerman, A., Van Buuren, H., et al. (1996). Single or Multiple Dose Ursodeoxycholic Acid For Cholestatic Liver Disease: Biliary Enrichment and Biochemical Response. Journal of Hepatology , 887-893.

Van Hoogstraten, H., Wolfhagen, F., Van de Meeberg, P., Kuiper, H., Nix, G., Becx, M., et al. (1998). Ursodeoxycholic Acid Therapy For Primary Sclerosing Cholangitis: Results of a 2-year Randomized Controlled Trial To Evaluate Single Versus Mutiple Daily Doses. Journal of Hepatology , 417-423.

Van Milligen de Wit, A., Van Deventer, S., & Tytgat, G. (1995). Immunogenetic Aspects of Primary Sclerosing Cholangitis: Implications for Therapeutic Strategies. American Journal of Gastroenterology , 90 (6), 893-900.

Van Thiel, D., Wright, H., & Gavaler, J. (1992). Ursodeoxycholic Acid (UDCA) Therapy For Primary Sclerosing Cholangitis (PSC): Preliminary Report of A Randomized Controlled Trial. Hepatology , 16:62A.

Vera, A., Gunson, B., Ussatoff, V., Nightingale, P., Candinas, D., Radley, S., et al. (2003). Colorectal Cancer in Patients with Inflammatory Bowel Disease After Liver Transplantation For Primary Sclerosing Cholangitis. Transplantation , 75 (12), 1983-8.

Whiteside, T., Lasky, S., Si, l., & Van Thiel, D. (1985). Immunologic analysis of mononuclear cells in liver tissues and blood of patients with primary sclerosing cholangitis. Hepatology , 5, 468-74.

Wiesner, R., & LaRusso, N. (1980). Clinicopathologic Features of the Syndrome of Primary Sclerosing Cholangitis. Gastroenterology , 79, 200-6.

Wiesner, R., Edwards, E., Freeman, R., Harper, A., Kim, R., Kamath, P., et al. (2003). Model for end-stage liver disease (MELD) and allocation of donor livers. *Gastroenterology , 124* (1), 91-96.

Wiesner, R., LaRusso, N., Dozois, R., & Beaver, S. (1986). Peristomal Varices After Proctocolectomy in Patients with Primary Sclerosing Cholangitis. Gastroenterology , 90 (2), 316-322.

Appendix I

Jadad Quality Score Calculation

Jadad Score Calculation

1) Was the study described as randomized (this includes words such as randomly, random, and randomization)? 0/1

2) Was the method used to generate the sequence of randomization described and appropriate (table of random numbers, computer-generated, etc)? 0/1

3) Was the study described as double blind? 0/1

4) Was the method of double blinding described and appropriate (identical placebo, active placebo, dummy, etc)? 0/1

5) Was there a description of withdrawals and dropouts? 0/1

6) Deduct one point if the method used to generate the sequence of randomization was described and it was inappropriate (patients were allocated alternately, or according to date of birth, hospital number, etc). 0/-1

7) Deduct one point if the study was described as double blind but the method of blinding was inappropriate (e.g., comparison of tablet vs. injection with no double dummy). 0/-1

Appendix II
Data Extraction Sheet

Data Extraction Sheet for Meta-Analysis:
Is UDCA Effective at Preventing Disease Progression in Patients with PSC?

Title: _____

Author(s): _____

Journal: _____ **Abstract:** _____

Name of Journal: _____

Conference (if applicable): _____

Country of Publication: _____

Total # of Pts: _____

Tx group: _____**pts** **placebo grp:** _____**pts** **Other grp (**_____**) :** _____**pts**

OR

 # in single UDCA grp: _____ , **# in mult UDCA grp:** _____

Dosage of UDCA: _____ **mg/kg/day**

Single or multidose: _____

Randomization complete: _____ **(yes/no)**

Blinding: _____**(none),** _____**(single),** _____ **(double),** _____ **(triple)**

Follow up: _____ **Month intervals,** _____ **Total follow up in yrs, Duration of Trial:** ____ **(yrs)**

Inclusion Criteria Listed: _____ **(yes/no), Exclusion Criteria Listed:** ____ **(yes/no)**

Are all pts accounted for at the end of the trial: _____ **(yes/no)**

Intention to tx analysis on all pts: _____ **(yes/no)**

Primary Outcomes listed a priori: _____ **(yes/no)**

Baseline Characteristics

	UDCA	Placebo	Other (_____)
Age			
Male: Female ratio			
# IBD pts			
Bilirubin level (mg/dl)			
AST (U/L)			
ALP (U/L)			
Albumin (g/dl)			
Varices (%)			

Primary Outcome Measurement: (Tick all that apply & * the Primary Outcome Measurements)

- Histology: _____
- ERCP: _____
- Liver Serum Tests: _____
- Death: _____
- Time to Transplant: _____
- Other (specify): _____

Results of Liver Biochemistries at Onset of Study and at the End of Study

	At Entry		End of Study		Total Δ in value		p-value	
	UDCA	Placebo	UDCA	Placebo	UDCA	Placebo	UDCA	Placebo
AST (U/L) (+/- SD)								
ALP (U/L) (+/- SD)								
Bilirubin (mg/dl) (+/- SD)								
Albumin (g/dl) (+/- SD)								

End Results of Liver Histology as compared to the beginning of the study

Improvement/ No change		Worsening	
UDCA	Placebo	UDCA	Placebo

End Results of ERCP as compared to the beginning of the study

Improvement / No change		Worsening	
UDCA	Placebo	UDCA	Placebo

Tolerability of UDCA: (Table indicates # of pts)

	UDCA	Placebo
Side Effects Reported		
S/E Requiring Discontinuation of UDCA		

Deaths Reported During The Trial:

UDCA	Placebo

Pts Requiring Transplant During The Trial:

UDCA	Placebo

Applicable Trial for this Meta-Analysis: _____ (yes/no)

Quality Score: _____

Comments:

Appendix III
Statistical Formulae

Statistical Formulae :

Standard 2X2 table

a	b	a+b
c	d	c+d
a+c	b+d	n

Odds Ratio (OR) :

$OR = (ad)/(bc)$

95% Confidence interval (CI) for OR:

$95\%CI = OR^{1\pm(1.96/\sqrt{x^2})}$

$x^2 = [n(ad-bc)^2]/[(a+c)(b+d)(a+b)(c+d)]$

Test of Heterogeneity:

$S_h = \Sigma(\ln(OR))^2/v_1 - [(\Sigma\ln(OR)/ v_1)^2/(\Sigma 1/ v_1)]$

$v_1 = (s.e.(\ln(OR))^2$

$s.e.(\ln(OR)) = \sqrt{[(1/a_i)+(1/b_i)+(1/c_i)+(1/d_i)]}$

where S_h has a chi-square distribution with k-1 degrees of freedom
k=number of trials

Mantel- Hanenszel Estimator of the Pooled Odds Ratio

$ORmh = \Sigma(weight_i \text{ X } OR_i)/\Sigma weight_i$

$OR_i = (a_i X d_i)/(b_i X c_i)$

$Weight_i = 1/variance_i$

$variance_i = N_i/(b_i X c_i)$

95% Confidence Interval for Mantel Haenszel Equation

95% CI = e ln(ORmh) +/- 1.96 x sqrt(var ORmh)

var ORmh =(∑F / 2 x ∑R2) + [∑G / (2 x ∑R x ∑S)] + (∑H/(2 x ∑S2)

where:

F = [ai x di x (ai + di)]/ni2
G = [ai x di x (bi+ci)] + (bi x ci x (ai + di))] / ni2
H = (bi x ci x (bi+ci)) / ni2
R = (ai x di) / ni
S = (bi x ci) / ni

Calculations for Using Standardized Mean Difference in Meta Analysis

Starting with
Means, SD pre and post, N, in each group

Raw difference in means

MeanChange(1) = Group 1 mean difference

MeanChange(2) = Group 2 mean difference

RawDiff = MeanChange(1) - MeanChange(2)

SDChange(1) = Sqr(SDPre(1) ^ 2 + SDPost(1) ^ 2 - 2 * CorrPrePost * SDPre(1) * SDPost(1))

SDChange(2) = Sqr(SDPre(2) ^ 2 + SDPost(2) ^ 2 - 2 * CorrPrePost * SDPre(2) * SDPost(2))

SDChangePooled = Sqr((((n(1) - 1) * SDChange(1) ^ 2 + (n(2) - 1) * SDChange(2) ^ 2) / (n(1) + n(2) - 2)))

MeanChangeDiffSE = Sqr(1 / n(1) + 1 / n(2)) * SDChangePooled

LogOddsRatio = PI * StdDiff / Sqr(3)
LogOddsSE = Sqr(PI ^ 2 * StdDiffSE ^ 2 / 3)
LogOddsVariance = LogOddsSE ^ 2
Where PI = 3.14159265358979
OddsRatio = Exp(LogOddsRatio)