HEPATOLOGY

AASLD AMERICAN ASSOCIATION FOR THE STUDY OF LIVER DISEASES

SPECIAL ARTICLES | HEPATOLOGY, VOL. 63, NO. 4, 2016

Surrogate Endpoints for Clinical Trials in Primary Sclerosing Cholangitis: Review and Results From an International PSC Study Group Consensus Process

Cyriel Y. Ponsioen,[1] Roger W. Chapman,[2] Olivier Chazouillères,[3] Gideon M. Hirschfield,[4] Tom H. Karlsen,[5] Ansgar W. Lohse,[6] Massimo Pinzani,[7] Erik Schrumpf,[8] Michael Trauner,[9] and Gregory J. Gores[10]

Primary sclerosing cholangitis (PSC) is a rare, but serious, cholestatic disease for which, to date, no effective therapy exists to halt disease progression toward end-stage liver disease. Clinical trial design to study drugs that improve prognosis is hampered by the relatively low event rate of clinically relevant endpoints. To overcome this shortcoming, there is an urgent need to identify appropriate surrogate endpoints. At present, there are no established surrogate endpoints. This article provides a critical review and describes the results of a consensus process initiated by the International PSC Study Group to delineate appropriate candidate surrogate endpoints at present for clinical trials in this frequently dismal disease. The consensus process resulted in a shortlist of five candidates as surrogate endpoints for measuring disease progression: alkaline phosphatase (ALP); transient elastography (TE); histology; combination of ALP+ histology; and bilirubin. Of these, histology, ALP, and TE came out as the most promising. However, the expert panel concluded that no biomarker currently exceeds level 3 validation. Combining multiple endpoints is advisable. *Conclusion*: At present, there are insufficient data to support level 2 validation for any surrogate endpoint in PSC. Concerted efforts by all stakeholders are highly needed. Novel, promising noninvasive biomarkers are under study and should be incorporated as exploratory endpoints in clinical trials. (HEPATOLOGY 2016;63:1357-1367)

Primary sclerosing cholangitis (PSC) is a progressive cholestatic disease with inflammation and stricturing of the biliary tree. Liver transplantation (LT)-free survival ranges from 13 to 21 years depending on the patient population under study.[(1)] It is a rare disease with prevalence ranging from 0 to 16.2 per 100,000 inhabitants, whereby it qualifies for the status of orphan disease. Many immunomodulatory and antifibrotic agents have been studied in PSC, the majority of which did not progress beyond the pilot phase. The only drug that may have a role in slowing disease progression is the naturally occurring bile acid, ursodeoxycholic acid (UDCA). However, despite two decades of clinical research,

Abbreviations: ALP, alkaline phosphatase; AST, aspartate aminotransaminase; CPA, collagen proportionate area; EASL, European Association for the Study of the Liver; ELF, enhanced liver fibrosis panel; EMA, European Medicines Agency; FDA, U.S. Food and Drug Administration; IPSCSG, International Primary Sclerosing Cholangitis Study Group; LOXL2, lysyloxidase-like 2; LSM, liver stiffness measurement; LT, liver transplantation; MRC, magnetic resonance cholangiography; MRI, magnetic resonance imaging; OCEBM, Oxford Center for Evidence Based Medicine; PSC, primary sclerosing cholangitis; RCT, randomized, clinical trial; TE, transient elastography; UDCA, ursodeoxycholic acid.

Received May 16, 2015; accepted September 24, 2015.

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep.28256/suppinfo.

View this article online at wileyonlinelibrary.com.

DOI 10.1002/hep.28256

Potential conflict of interest: Dr. Hirschfield consults for Intercept and advises Falk. Dr. Trauner consults, advises, is on the speakers' bureau for, and received grants from Falk. He consults and advises Phenex. He is on the speakers' bureau for and received grants from Gilead and Roche. He is on the speakers' bureau for MSD and received grants from Intercept and Albireo. Dr. Ponsioen is on the speakers' bureau for and received grants from Dr. Falk. He is on the speakers' bureau for MSD and received grants from Takeda and AbbVie.

there is still no evidence from randomized trials that UDCA is truly beneficial in improving transplant-free survival. The 2011 guidelines from the American Association for Study of Liver Diseases now recommend against the use of UDCA in PSC. The Clinical Practice Guidelines from the European Association for the Study of the Liver (EASL) state that the "limited" database does not yet allow a specific recommendation for the general use of UDCA in PSC.[2] In 2011, during the Berlin annual EASL meeting, PSC was declared one of the biggest unmet needs in hepatology. Since then, not much has changed, which may, in part, be owing to the low prevalence of the disease. The largest randomized trial in PSC enrolled 219 patients, but failed to attain its predefined sample size.[3] This exemplifies another important hurdle for clinical trials testing drugs that are supposed to halt disease progression. The occurrence of a sufficient number of solid clinical endpoints, such as death, end-stage liver failure, or development of cholangiocarcinoma in PSC, takes too long to achieve in a disease in which the median period of time to death or transplantation is 12-21 years. The Scandinavian trial employed a follow-up of 5 years, but despite that, only 11 of 101 patients in the placebo group reached the predefined coprimary endpoint of death or LT, and the difference with 7 of 97 UDCA-treated patients was not statistically significant ($P = 0.37$).[3] Likewise, the event rates in terms of solid clinical endpoints such as death, LT, and complications of cirrhosis in the placebo arm of the only other large clinical trial as well as in the largest population-based cohort were also less than 4%.[1,3,4] Of note, 41% of placebo-treated subjects in the Lindor trial had baseline Ludwig's stage of III of IV, meaning that advanced stages were well represented.[4] An event rate of 4% implies, for example, that for a clinical trial in order to have an 80% power to pick up a hazard rate of 0.50 in a 2:1 active substance versus placebo-randomized design, one needs at least 2,775 patient-years of follow-up.

TABLE 1. Hierarchy of (Surrogate) Endpoints[5]

Level 1:	a true clinical-efficacy measure
Level 2:	a validated surrogate endpoint (for a specific disease setting and class of interventions)
Level 3:	a nonvalidated surrogate endpoint, yet one established to be "reasonably likely to predict clinical benefit" (for a specific disease setting and class of interventions)
Level 4:	a correlate that is a measure of biological activity, but that has not been established to be a higher level

The performance of phase II—but similarly also phase III—trials that need treatment duration and follow-up of at least 6 years are not feasible in an orphan disease such as PSC. Therefore, there is a great need for robust surrogate endpoints that do reliably reflect that the therapy under study will have a beneficial impact on disease progression.

In 2005, Fleming proposed a hierarchy of (surrogate) endpoints (see Table 1).[5] Regulatory authorities usually accept level 2 endpoints for registration purposes. Occasionally, they may grant provisional or accelerated approval for an investigational product on the basis of level 3 surrogate endpoint(s), but will likely require future confirmation of the treatment effect in long-term follow-up or open-label extension studies. The International PSC Study Group (IPSCSG) recognizes that given the notion that there is currently no effective medical therapy for PSC and there is a current resurgence of interest from the pharma industry for this orphan disease, there is an urgent need to identify meaningful surrogate endpoints for clinical trials in PSC.

ARTICLE INFORMATION:

From the [1]Department of Gastroenterology & Hepatology, Academic Medical Center, Amsterdam, The Netherlands; [2]Department of Gastroenterology & Hepatology, John Radcliffe Hospital, Headington, Oxford, United Kingdom; [3]AP-HP, Hôpital Saint-Antoine, Service d'Hépatologie and Sorbonne Universités, UPMC Univ Paris, France; [4]Center for Liver Research and NIHR Biomedical Research Unit University of Birmingham, Birmingham, United Kingdom; [5]Norwegian PSC Research Center, Oslo University Hospital, Oslo, Norway; [6]Department of Medicine, University Medical Center Hamburg, Eppendorf, Germany; [7]Division of Medicine, University College London, UCL Institute for Liver and Digestive Health, Royal Free Hospital, London, United Kingdom; [8]Norwegian PSC Research Center, Oslo University Hospital, Oslo, Norway; [9]Division of Gastroenterology and Hepatology, Department of Internal Medicine III, Medical University of Vienna, Vienna, Austria; [10]Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN

ADDRESS CORRESPONDENCE AND REPRINT REQUESTS TO:

Cyriel Y. Ponsioen, M.D., Ph.D.
Department of Gastroenterology and Hepatology
Academic Medical Center
P.O. Box 22700, 1100 DE Amsterdam, The Netherlands.
E-mail: c.y.ponsioen@amc.uva.nl
Fax: 31 20 5669158

In 1984, Arlene Fink et al. stated that "in this less than perfect world, consensus studies are intended to correct for the lack of conclusive data by putting the knowledge and experience of practitioners and other experts in touch with the available information".[6] The purpose of this article is to critically review currently and previously used endpoints in clinical as well as observational studies, identify the most appropriate candidates at present for clinical trials, and provide recommendations for further validation of these as well as of other potential future endpoints.

Overview of Available Literature

Table 2 provides an overview of clinical studies performed in adult PSC since 1995. As can be observed, most studies were open-label series, there are only 4 studies with more than 100 study subjects, and treatment duration exceeded 1 year in only one third. Main observations from this table are: (1) alkaline phosphatase (ALP) was used in all 26 studies and showed significant improvement in 17 of 26 and in 10 of 26 it was (part of) the primary outcome measure; (2) histology was used in 12 of 26 studies, significant improvement was observed in 3 of 12, in one study it was part of the coprimary endpoint; (3) clinical outcome was used as primary outcome measure in 4 of 26 studies with a median follow-up of 2.2-5.0 years, no difference was observed in three, and worsening in one trial; and (4) the Mayo PSC Risk Score was used a secondary outcome measure in nine studies, with (partial) improvement in two.

At present, there are four industry trials ongoing in PSC. Nor-UDCA is the C23 homolog of 3α,7β dihydroxy-C24 UDCA. In contrast to UDCA, nor-UDCA is barely amidated in liver cells and induces a bicarbonate-rich hypercholeresis mostly because of cholehepatic shunting of the unconjugated form.[31] It is currently being investigated in a randomized, placebo-controlled phase II trial that employs change in ALP as the primary endpoint. Simtuzumab is a monoclonal antibody against lysyloxidase-like 2 (LOXL2). The mechanism of LOXL2 is the induction of collagen (I and III) and elastin cross-linking making the scar less degradable and ultimately inducing tissue tension, which is an additional profibrogenic.[32] Results from a randomized, placebo-controlled phase IIb study are awaited. Primary endpoint in this trial is change from baseline in morphometric quantitative collagen on liver biopsy. LUM001 is an apical sodium-dependent bile acid transporter inhibitor, which is currently being tested in an open-label phase II study, using safety and tolerability as well as laboratory parameters, such as serum bile acids, as outcome measures. The bile acid derivative, obeticholic acid, is a potent farnesoid X nucelar receptor activator, which is currently under study in a phase II placebo-controlled design employing decrease in ALP as primary outcome measure.

Consensus Process

A consensus process was initiated by the IPSCSG applying a hybrid between a Delphi consensus and a Nominal Group process[6] in analogy to the European Crohn's and Colitis Organization guideline development for ulcerative colitis.[33] Panelists were selected among leading PSC scientists within the IPSCSG, based on their experience with clinical trials in PSC and to ensure widespread representation of nationalities and research centers. Three rounds of questionnaires were sent out together with anonymous feedback of the response to all panelists in order to identify a shortlist of candidate surrogate endpoints (Delphi process). All panelists gathered in November 2014 in Boston for further plenary discussion (Nominal Group process) of this shortlist. A review of all available evidence was presented by the chair, upon which statements were formulated and revised until consensus was reached. Consensus was defined as 100% agreement. In the case of disagreement, an item was deferred for further literature search or agreed to discard. Levels of evidence and grades of recommendation were based on the 2009 version of the Oxford Center for Evidence Based Medicine (OECBM; www.cebm.net) (see Table 3 and Supporting Table 1). Outcomes were discussed at two plenary IPSCSG meetings in April 2015 in Vienna.

CANDIDATE BIOMARKERS

At the completion of the Delphi rounds out of a list of 19 suggested biomarkers (Supporting Fig. 1), the following shortlist of candidate surrogate endpoints was agreed upon: ALP; transient elastography (TE); histology; combination of ALP+histology; and bilirubin (Fig. 1). These five biomarkers were further discussed in the plenary Nominal Group meeting.

TABLE 2. Overview of Clinical Studies in Adult PSC Since 1995

First Author	Year of Publication	No. of Subjects	Type	Duration (months)	Drug(s)	Endpoints*	Results
van Thiel[2]	1995	10	OL	12	Tacrolimus	ALP, ALT, AST, bilihistology	Improvement†
							NS?
Olsson[3]	1995	84	RCT	36	Colchicine	1st: clinical	No changes
						2nd: ALP, bili, ALT, albumin	No difference
						Symptoms	No changes
						Histology	No change
Lindor[4]	1997	105	RCT	24	UDCA	1st: clinical	No difference
						Histology	NS
						2nd: ALP, bili, AST	Improvement†
van Hoogstraten[5]	1998	48	RCT	24	UDCA TID vs. OD	1st: ALP, bili, AST	No differences
						2nd: histology	NS
						Cholangiography	NS
						Mayo risk score	No difference
Angulo[6]	1999	8	OL	12	Nicotine	ALP, AST, ALT	NS
van Milligen de Wit[7]	1999	17	OL	12	UDCA	ALP, AST, ALT, GGT, Ig, PTT	Improvement†
						Cytokines	NS
						Histology	No change
Schramm[8]	1999	15	OL	41	UDCA/prednisone/azathioprine	ALP, AST, ALP, bili,	Improvement†
						Cholangiography	No change
						Histology	Slight Improvement
Angulo[9]	2000	21	OL	12	Budesonide	ALP, bili, AST	Varying†
						Histology	No change
						Mayo risk score	NS
van Hoogstraten[10]	2000	18	RCT	2	Budesonide vs. prednisolone	ALP, AST, ALT, bili	Varying†
						Symptoms	Improvement†
Bharucha[11]	2000	20	OL	12	Pentoxyfylline	ALP	No change
						Symptoms	No change
Vleggaar[12]	2001	12	RCO	2	Nicotine	ALP, AST, bili	NS
						Symptoms	NS
Mitchell[13]	2001	26	RCT	24	UDCA vs. placebo	ALP, bili, AST, GGT	Varying†
						Histology	Improvement†
						Cholangiography	Improvement†
Harnois[14]	2001	30	OL	12	High-dose UDCA	ALP, ASAT, bili, albumin	Improvement†
Färkkilä[15]	2004	80	RCT	36	UDCA±metronidazole	1st: ALP	Improvement†
						2nd: histology	Improvement
						Symptoms	No difference
						Cholangiography	No difference
						Mayo risk score	Improvement†
Talwalkar[16]	2005	30	OL	12	Mycophenolate	1st: ALP, AST, bili	Varying†
						Mayo risk score	No change
Olsson[17]	2005	219	RCT	60	UDCA vs. placebo	1st: clinical	NS
						2nd: ALP, ALT, bili	NS
Talwalkar[18]	2007	16	OL	12	Tacrolimus	ALP	Improvement†
Vleggaar[19]	2008	14	RCO	3	Probiotics	ALP, AST, ALT, bili	No differences
						Symptoms	No change
Hommes[20]	2008	10	RCT	6	Infliximab vs. placebo	1st: ALP	NS
						2nd: histology	No differences
						Symptoms	No changes
Cullen[21]	2008	31	RCT	30	UDCA 10 vs. 20 vs. 30 mg/kg	ALP, AST, ALT, gGT, bili	Varying†
						Histology, Mayo risk score	
Angulo[22]	2008	30	OL	12	Silymarin	1st: ALP, AST, bili	Varying†
						2nd: Mayo risk score	No change
Silveira[23]	2009	16	OL	12	Minocycline	1st: ALP, bili, AST	Improvement†
						2nd: Mayo risk score	Improvement†
Lindor[24]	2009	150	RCT	60	UDCA vs. placebo	1st: clinical	Worsening†
						2nd: ALP, AST, bili	Improvement†
						Histology	NS‡
Martin[25]	2012	23	OL	12	Docosahexaenoic acid	1st: ALP	Improvement†
						2nd: AST, ALT, bili	Varying
						ELF	No change
Tabibian[26]	2013	35	RCT	3	Vancomycin vs. metronidazole	1st: ALP	Varying†

TABLE 2. *Continued*

First Author	Year of Publication	No. of Subjects	Type	Duration (months)	Drug(s)	Endpoints*	Results
						2nd: bili	Varying†
						Symptoms	Varying†
						Mayo risk score	Varying†
Tabibian[27]	2014	16	OL	3	Rifaximine	1st: ALP	No change
						2nd: bili GGT, Mayo riskscore, symptoms	No change

*1st: primary endpoint(s); 2nd: secondary endpoint(s); in case the primary endpoint was not defined, then the parameter on which sample size calculation was based if mentioned in the article was designated as such.
†Statistically significant meaning $P < 0.05$.
‡Follow-up biopsies available for only 31 patients owing to early termination of study.
Abbreviations: OL, open label; RCO, randomized crossover; ALT, alanine aminotransferase; bili, serum bilirubin; NS, not statistically significant; GGT, gamma-glutamyl transpeptidase; Ig, immunoglobulin; PTT, partial thromboplastin time.

ALP

ALP has been employed in all clinical trials in the past two decades and used as primary endpoint in over 40% of studies (see Table 3). Most trials assessed continuous ALP changes on group level, whereas a few studies used a predefined response in ALP (e.g., a decline of at least 50%). Seventeen of twenty-six trials reported significant improvement of ALP post-treatment. When attempting to correlate changes in ALP with clinical improvement, only four randomized, clinical trials (RCTs) with a follow-up of 2-5 years are available.[3,4,8,9] In the Scandinavian high-dose UDCA study with a follow-up of 5 years, no significant changes in either ALP or clinical improvement were observed.[3] In the very-high-dose UDCA study with the same length of follow-up, clinical worsening in the treated group, despite significantly more improvement in ALP compared to placebo, was observed.[4] These outcomes allow, as of yet, no conclusions with regard to the usefulness of ALP as a surrogate parameter for clinical efficacy of UDCA.

Recently, several observational studies have fueled the notion that ALP is a surrogate marker for transplant-free survival.[34-38] A post-hoc analysis of the very-high-dose UDCA trial as reported in the Stanich article on ALP dynamics revealed that those patients on UDCA as well as the placebo-treated ones that did normalize their ALP during follow-up had a better prognosis.[34] Of note, ALP was not a primary endpoint in this very-high-dose UDCA trial. The post-hoc analysis of the large Scandinavian UDCA trial also showed that irrespective of UDCA use, those patients with a 40% reduction or normalization of their ALP had a much better survival.[36] These observations suggest that although improvement of ALP is associated with better outcome, it is not mediated by UDCA, which, by itself, also lowers ALP levels. It is apparent from all these studies that ALP is, at the very least, a useful parameter for stratification of patients in clinical trials, although thresholds still need to be clarified. The panel concluded that ALP is widely recognized as a clinical measure of cholestasis. Currently, albeit not formally validated, it is regarded as a potential surrogate outcome parameter *[EL4, RGD]*.

TABLE 3. OCEBM Grades of Recommendation

A	Consistent level 1 studies
B	Consistent level 2 *or* 3 studies or extrapolations from level 1 studies
C	Level 4 studies *or* extrapolations from level 2 or 3 studies
D	Level 5 evidence *or* troublingly inconsistent or inconclusive studies of any level

Reprinted with permission from http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009.

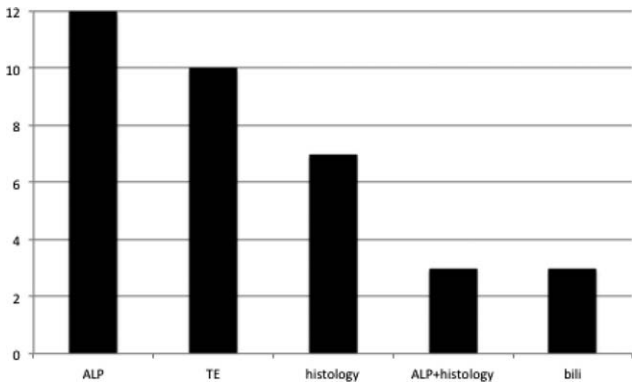

FIG. 1. Ranking of shortlist candidate surrogate (co-)primary endpoints at completion of Delphi consensus rounds. Shortlist was determined by any parameter that was chosen as (co-)primary endpoint at least twice by the panelists. Ranking was determined by top three preference of each panelist among all original 19 endpoints.

TE

Liver stiffness measurement (LSM) by TE has proven to be a powerful predictor of prognosis in chronic hepatitis C.[39] Recently, Corpechot et al. showed that baseline measurements and rate of LSM progression were strongly and independently linked with patients' outcomes, suggesting that TE may be an attractive surrogate endpoint.[40] Demonstration of progression to cirrhosis may be an acceptable intermediate endpoint for regulatory authorities. An international Prospective Study of the Prognostic Value of Transient Elastography (FibroScan) in PSC Patients (FICUS Study) within the IPSCSG will not only identify its use as a stratifier for trials, but will likely also shed more light on the correlation of changes in LSM with clinical events. The panel agreed that TE as a noninvasive measurement of liver stiffness is a potential surrogate endpoint of prognosis *[EL2b, RGC]*.

HISTOLOGY/LIVER BIOPSY

Histological assessment of the grade of inflammation and the amount of injury to the liver as measured by the level of fibrosis has been regarded as the most direct way to gauge disease severity in many liver diseases, and as such offers face validity. Histology has been used as an outcome parameter in 12 of 26 studies in the past 20 years (Table 3). The Ludwig's score of fibrosis, which has been the reference standard in PSC, was used in eight studies with the following stage descriptors: stage I = portal hepatitis; stage II = periportal hepatitis; stage III = bridging fibrosis; and stage IV = cirrhosis. In their original article on developing the Mayo PSC Risk Score, Wiesner et al. showed that increasing Ludwig stage was associated with poorer survival.[41] Recent data from Oxford in 123 PSC patients showed that patents with Ludwig stage 1/2 fared much better than those with stage 3/4 (personal communication, Prof. Chapman). Histology is regarded as less undulating than serum liver tests, but can be subject to considerable sampling variability.[42] As an appropriate surrogate endpoint, the Ludwig stage should either be stable or progress within a reasonable time frame in placebo-treated patients. Angulo et al. observed the progression in Ludwig stage in 107 PSC patients with paired liver biopsies, 84% of which had less than 24 months in between.[43] They found that at the second biopsy, 53% of stage I-III PSC patients had progressed, but14% of stage III-IV patients now had stage I-II. A Markov model based on the assumption that histological stage does not spontaneously regress in PSC estimated that 66% of stage II patients would progress after 2 years. A recent study from Amsterdam showed that staging by the novel Nakanuma system developed for primary biliary cirrhosis, by Ishak scoring, as well as by the classic Ludwig scoring, was strongly associated with transplant-free survival by all three scoring systems in descending order.[44] A multicenter validation study is currently underway, and studies looking into changes in biopsies over time are welcomed. These findings suggest that sampling error might not be a major drawback of liver biopsy. In hepatitis C, Ishak stage is clearly associated with long-term clinical outcome, but not in fragmented biopsies.[45] Increasing the amount of tissue by taking two passes through the same introduction needle may reduce variability and is not associated with increased risk. Given the regional differences in PSC, it may also help to sample the follow-up biopsy from the same area. Recently, more objective continuous measures of the amount of fibrosis have been developed, such as collagen proportionate area (CPA). Morphometric analysis using the CPA system has demonstrated an excellent positive correlation between the amount of fibrosis in a cirrhotic liver and the relative hepatic vein pressure gradient and liver tissue stiffness in liver diseases, such as steatohepatitis and chronic viral hepatitis.[46] This may substantially reduce interobserver variability and relative insensitivity to change, which are associated with ordinal histological scoring systems such as Ishak and Ludwig.

An additional advantage of histology as a surrogate parameter is that it may offer possibilities to study the mechanism of action of an investigational drug as well as providing tissue for safety assessment. Furthermore, liver biopsy has the advantage of face validity (i.e., assessing directly what happens in the tissue of interest). A disadvantage of liver biopsy is that it is an invasive procedure and carries some risk of adverse events. In a collective number of 1,152 liver biopsies in 541 PSC patients enrolled in the past 20 years in clinical studies, no mortality has been reported (Table 1). Bile leakage was reported in <0.2%. In general, mortality from liver biopsy ranges from 0.09 from 0.3 in 1,000 in 68,276 respective 98,445 biopsies.[47,48] The serious adverse event rate is reported to be around 0.5% when biopsy is performed under ultrasound guidance.[49,50] In addition to its value as an outcome parameter, the available data indicate that histology seems a proper stratification tool for clinical trials. The panel concluded that liver biopsy has the potential to be a robust

surrogate endpoint for clinical trials in PSC *[EL2b, RG B]*.

ALP+HISTOLOGY

In the absence of a convincing single-surrogate endpoint, combining multiple endpoints is considered advisable and should be explored further *[EL5, RGD]*.

When there is no validated and accepted biomarker to gauge the beneficial effects of an investigational therapy on symptoms or disease course, the combination of two or more endpoints can strengthen evidence for the treatment effect when used collectively as a coprimary endpoint. Vice versa, it is generally believed that if a treatment is ineffective, the chance to show that the treatment is effective in all coprimary endpoints should be small.[51] As such, a coprimary endpoint can offer more reassurance for a true treatment effect.

This is opposed to a composite endpoint, which is a combination of various clinical events, where any one of those events would count as satisfying the composite endpoint. Applying a composite endpoint can be very useful when a particular therapy is deemed to have similar effects on several related events to achieve adequate power in diseases where clinical events do not occur very often. For validation of specific surrogate endpoints, it is regarded as less suitable.

BILIRUBIN

Serum bilirubin is part of several prognostic scoring systems including the Child-Pugh-Turcott, the PSC Mayo Risk Score, and the Model for End-Stage Liver Disease score. These models generally have a horizon of only 2-4 years and perform best in predicting impending liver failure in end-stage liver disease. A relatively large cohort study from Germany incorporated bilirubin as a dichotomous variable in a prognostic model.[52] The researchers found that a persistently elevated bilirubin for more than 3 months was significantly associated with poorer survival both in uni- and multivariate analysis. However, median transplant-free survival of these patients was only 30 months, indicating that these likely were relatively late-stage patients. Bilirubin can also acutely rise in the case of acute cholangitis. Treating the suppurative cholangits with appropriate antibiotics and dilatation of the—often accompanying—dominant stricture usually results in lowering of bilirubin level. As such, treating dominant strictures may have a favorable effect on prognosis, although sound evidence for this is still pending. The temporary rise in bilirubin may likely not reflect long-term outcome in this situation. Hence, although bilirubin has been shown to be a marker of prognosis, the panel felt that it was unlikely to be suitable for clinical trials, because it only rises permanently in late-stage disease *[EL2b, RGC]*.

OTHER EXPLORATORY BIOMARKERS

The IPSCSG consensus panel recommends to incorporate serum liver tests (in particular, ALP), enhanced liver fibrosis (ELF) panel, or other serum fibrosis markers, imaging, as well as noninvasive measures of liver tissue stiffness as exploratory endpoints in clinical trials with the purpose of investigating their future potential as a surrogate endpoint *[EL4, RGC]*.

SEROLOGICAL MARKERS

The ELF panel is a set of three serological markers of liver fibrosis, that is, hyaluronic acid, tissue inhibitor of metalloproteinases-1, and propeptide of type III procollagen. This has been shown not only to correlate accurately with degree of liver fibrosis in chronic hepatitis C and nonalcoholic fatty liver disease, but also with prognosis.[53-55] The first promising experience with ELF in PSC was recently published in this journal.[56] PSC patients stratified by the ELF score tertiles exhibited significantly different transplant-free survival ($P < 0.001$), with higher scores associated with shorter survival, further confirmed in the validation set stratified by ELF Test tertiles ($P = 0.003$). In addition, the ELF Test distinguished between mild and severe disease as defined by clinical outcome (transplantation or death) with an area under the curve of 0.81.

IMAGING

The Amsterdam cholangiographic prognostic model has adequate predictive power also in earlier-stage disease, but requires an invasive procedure associated with a much higher risk of adverse events than liver biopsy, which is undesirable for clinical trials.[57] Recently, Ruiz et al. reported on their experience with sequential magnetic resonance imaging (MRI) and magnetic resonance cholangiography (MRC) in PSC.[58] They showed that 58% of patients showed radiographic aggravation over a period of up to 4 years. No correlation with clinical outcome was reported. A drawback

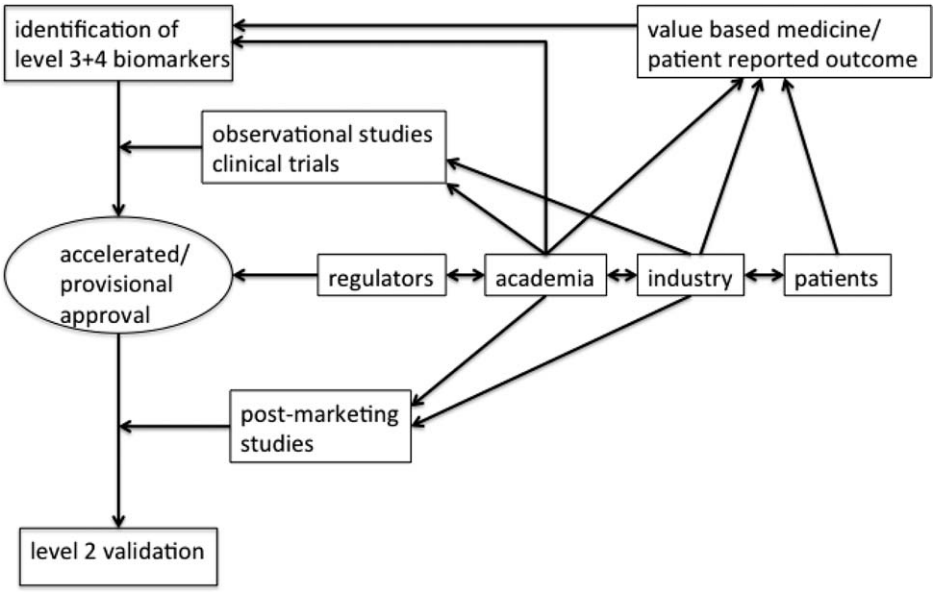

FIG. 2. Proposed roadmap for developing validated surrogate biomarkers.

of using MRC for staging PSC is that there is no standardization of scanning protocols and, consequently, quite some variation in image quality. Furthermore, the resolution and therefore the sensitivity to detect subtle caliber changes in the intrahepatic biliary tree are too low. Nilsson et al. recently reported on dynamic gadoxetate-enhanced MRI for the assessment of total and segmental liver function and volume in PSC.[59] They found that liver function was significantly more heterogeneously disturbed in PSC patients as compared to controls, and segmental function correlated with biliary obstruction. Although still experimental, this approach holds promise because it provides a functional estimate of the state of the liver in PSC and is not dependent on high resolution and observer interpretation.

Discussion

For clinical trials that target disease progression in PSC, the relative frequency of robust level 1 clinical endpoints, such as death, LT, and even development of complications of cirrhosis, is too low. Investigators will be reluctant to embark on phase III trials that need >2,500 patient years of follow-up in an orphan disease indication. Therefore, the field is in great need of proper level 2 endpoints. Thus far, none of the surrogate endpoints in PSC have exceeded level 3 or 4. In 1989, Prentice stated that any proper surrogate endpoint should satisfy the requirement that the biological marker must be correlated with the clinical endpoint.[60] There have been four clinical trials that have evaluated clinical outcome, with no difference between active drug and placebo in three, and worsening in one.[3,4,8,9] It is likely that all four trials were underpowered with regard to clinical events, so it is not possible to draw conclusions regarding correlations between clinical outcome and change in secondary outcome parameters, such as ALP, aspartate aminotransferase (AST), bilirubin, or histology. Hence, there are no data to support a level 2 claim for any of the endpoints used in clinical trials to date.

Acceptability of histology as a primary surrogate endpoint for clinical trials may be hampered by the fact that it is no longer part of the American and European guidelines for diagnosing large-duct PSC, as well as its invasive nature. However, the phase IIb simtuzumab trial, which assesses change in fibrosis on liver biopsy, apparently completed recruitment of 235 study subjects within 1.5 years (ClinicalTrials.gov: NCT01672853).

This consensus process yielded histology, ALP, TE, and combinations thereof as the most likely candidate surrogate parameters for clinical trial design at the moment, as well as a few exploratory parameters that may prove quite useful in the future (e.g., to supplant histology). In this respect, TE, ELF, and MRI seem promising. The urgent task ahead is to upgrade these to level 2 validation, such that one or more

surrogate parameters will become available that is/are acceptable for regular approval. For this, the IPSCSG proposes a roadmap as depicted in Fig. 2. A good example of an observational study is the aforementioned FICUS study. Foremost, close collaboration of all stakeholders—academia, patients, Pharma, as well as regulators—is key. As long as there is no cure for PSC, patient expectations in terms of what any therapy should offer is of paramount importance. In this respect, the emerging of concept of value-based medicine should play a more prominent role in determining what aspect(s) of impact of a novel therapy on the disease should be captured by a biomarker. With regard to endpoints that measure symptoms or quality of life, it should be mentioned that there are currently no validated questionnaires to assess patient-reported outcome measures in PSC. Collaboration with patient interest groups will be important to identify and validate the most relevant ones in this respect. Validating the above-mentioned biomarkers will need large, rigorously conducted clinical trials, which can only be afforded by industry. Studying the impact of novel therapies on disease behavior and correlating these to currently available and exploratory biomarkers will require postmarketing confirmation of treatment effect in long-term follow-up studies to accrue sufficient solid clinical events. Academia and Pharma should closely work together to achieve this. Moreover, to adequately analyze the impact of a novel therapy on a biomarker, the natural course and variation in the comparator group should be assessed. To this end, collaboration to share information on placebo-treated patients for example, baseline and change in follow-up histological staging, ALP dynamics, variation, and change in TE over time, is vital to accelerate biomarker validation. In the absence of level 2 biomarkers, Pharma companies will seek accelerated (U.S. Food and Drug Administration; FDA) or provisional (European Medicines Agency; EMA) approval. The FDA, for instance, may grant approval based on effects of a surrogate endpoint that is "reasonably likely, based on epidemiologic, therapeutic, pathophysiologic, or other evidence to predict benefit" on irreversible morbidity or mortality.[61] "Reasonably likely" implies that there are no fixed definitions and approval is currently made on a case-by-case basis. The same applies to validation by definitive studies such that a surrogate can be used for traditional approval. Hence, involvement of regulatory bodies early in the design of drug development programs, as well as in level 2 validation efforts is of great importance. Both FDA and EMA offer several tools for biomarker development.*†

Last, biomarkers should also be evaluated on their stratification properties, given that depending on the mechanism of action of an investigational compound, it can be important to recruit early-stage patients with predominantly inflammation or later-stage patients with a more advanced fibrosis.

Summary

This IPSCSG consensus process has produced a number of recommendations that are intended to offer guidance for drug development programs in PSC. Based on the available literature, there are only a handful of surrogate endpoints that seem reasonably likely to predict clinical benefit with regard to preventing disease progression and could serve as stratifier. Histology appears to have the most robust record currently, although it does not exceed level 3 validation. The foremost recommendation is that any drug development program should incorporate assessment of the above-mentioned (exploratory) surrogate endpoints in order to yield data that may upgrade one or more of these to level 2. In the event of provisional/accelerated approval, postmarketing confirmation of treatment effect in long-term follow-up studies is highly likely to be needed. Such studies offer an important opportunity to validating surrogate endpoints.

REFERENCES

1) Boonstra K, Weersma RK, van Erpecum KJ, Rauws EA, Spanier BW, Poen AC, et al. Population-based epidemiology, malignancy risk, and outcome of primary sclerosing cholangitis. HEPATOLOGY 2013;58:2045-2055.
2) European Association for the Study of the Liver. EASL Clinical Practice Guidelines: management of cholestatic liver diseases. J Hepatol 2009;51:237-267.
3) Olsson R, Boberg KM, Schaffalitsky de Muckadell O, Lindgren S, Hultcrantz R, Folvik G, et al. High-dose ursodeoxycholic acid in primary sclerosing cholangitis: a 5-year multicenter, randomized, controlled study. Gastroenterology 2005;129:1464-1472.
4) Lindor KD, Kowdley KV, Luketic VA, Harrison ME, McCashland T, Befeler AS, et al. High-dose ursodeoxycholic acid for the treatment of primary sclerosing cholangitis. HEPATOLOGY 2009;50:808-814.

*www.fda.gov/Drugs/Guidance ComplianceRegulatoryInformation/Guidances/default.

†www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/document_listing/document_listing_000319.jsp&mid=WC0b01ac0580022bb0.

5) Fleming TR. Surrogate endpoints and FDA's accelerated approval process. Health Aff (Millwood) 2005;24:67-78.
6) Fink A, Kosecoff J, Chassin M, Brook RH. Consensus methods: characteristics and guidelines for use. Am J Public Health 1984; 74:979-983.
7) van Thiel DH, Carroll P, Abu-Elmagd K, Rodriguez-Rilo H, Irish W, McMichael J, et al. Tacrolimus (FK 506), a treatment for primary sclerosing cholangitis: results of an open-label preliminary trial. Am J Gastroenterol 1995;90:455-459.
8) Olsson R, Broomé U, Danielsson Å, Hagerstrand I, Järnerot G, Lööf L, et al. Colchicine treatment of primary sclerosing cholangitis. Gastroenterology. 1995;108:1199-1203.
9) Lindor KD. Ursodiol for primary sclerosing cholangitis. N Engl J Med 1997;336:691-695.
10) Van Hoogstraten HJ, Wolfhagen FH, van de Meeberg PC, Kuiper H, Nix GA, Becx MC, et al. Ursodeoxycholic acid therapy for primary sclerosing cholangitis: results of a 2-year randomized controlled trial to evaluate single versus multiple daily doses. J Hepatol 1998;29:417-423.
11) Angulo P, Bharucha AE, Jorgensen RA, Desotel CK, Sandborn WJ, Larusso NF, et al. Oral nicotine in treatment of primary sclerosing cholangitis (a pilot study). Dig Dis Sci 1999;44:602-607.
12) van Milligen de Wit M, Kuiper H, Camoglio L, van Bracht J, Jones EA, Tytgat GN, van Deventer SJ. Does ursodeoxycholic acid mediate immunomodulatory and anti-inflammatory effects in patients with primary sclerosing cholangitis? Eur J Gastroenterol Hepatol 1999;11:129-136.
13) Schramm C, Schirmacher P, Helmreich-Becker I, Gerken G, Meyer zum Buschenfelde KH, Lohse A. Combined therapy with azathioprine, prednisolone, and ursodiol in patients with primarysclerosing cholangitis. Ann Intern Med 1999;131: 943-946.
14) Angulo P, Batts K, Jorgensen RA, Larusso N, Lindor KD. Oral budesonide in the treatment of primary sclerosing cholangitis. Am J Gastroenterol 2000;95:2333-2337.
15) Van Hoogstraten HJ, Boland GJ, van Steenbergen W, Griffioen P, Hop WC, van Hattum J, et al. Budesonide or prednisone in combination with ursodeoxycholic acid in primary sclerosing cholangitis: a randomized double-blind pilot study. Am J Gastroenterol 2000;95:2015-2022.
16) Bharucha AE, Jorgensen R, JLichtman SN, LaRusso NF, Lindor KD. A pilot study of pentoxifylline for thetreatment of primary sclerosing cholangitis. Am J Gastroenterol 2000;95: 2338-2342.
17) Vleggaar FP, van Buuren HR, van Berge Henegouwen GP, Hop WC, van Erpecum KJ. No beneficial effects of transdermal nicotine in patients with primary sclerosing cholangitis: results of a randomized double-blind placebo-controlled cross-over study. Eur J Gastroenterol Hepatol 2001;13:171-175.
18) Mitchell SA, Bansi DS, Hunt N, Bergmann KV, Fleming KA, Chapman RW. A preliminary trial of high-dose ursodeoxycholic acid in primary sclerosing cholangitis. Gastroenterology 2001; 121:900-907.
19) Harnois DM, Angulo P, Jorgensen RA, Larusso NF, Lindor KD. High-dose ursodeoxycholic acid as a therapy for patients with primary sclerosing cholangitis. Am J Gastroenterol 2001;96: 1558-1562.
20) Farkkila M, Karvonen AL, Nurmi H, Nuutinen H, Taavitsainen M, Pikkarainen P, et al. Metronidazole and ursodeoxycholic acid for primary sclerosing cholangitis: a randomized placebo-controlled trial. HEPATOLOGY 2004;40:1379-1386.
21) Talwalkar JA, Angulo P, Keach JC, Petz JL, Jorgensen RA, Lindor KD. Mycophenolate mofetil for the treatment of primary sclerosing cholangitis. Am J Gastroenterol 2005;100:308-312.
22) Talwalkar JA, Gossard AA, Keach JC, Jorgensen RA, Petz JL, Lindor RN. Tacrolimus for the treatment of primary sclerosing cholangitis. Liver Int 2007;27:451-453.
23) Vleggaar FP, Monkelbaan J, van Erpecum KJ. Probiotics in primary sclerosing cholangitis: a randomized placebo-controlled crossover pilot study. Eur J Gastroenterol Hepatol 2008;20:688-692.
24) Hommes DW, Erkelens W, Ponsioen C, Stokkers P, Rauws E, van der Spek M, et al. A double-blind, placebo-controlled, randomized study of infliximab in primary sclerosing cholangitis. J Clin Gastroenterol 2008;42:522-526.
25) Cullen SN, Rust C, Fleming K, Edwards C, Beuers U, Chapman RW. High dose ursodeoxycholic acid for the treatment of primary sclerosing cholangitis is safe and effective. J Hepatol 2008;48:792-800.
26) Angulo P, Jorgensen RA, Kowdley KV, Lindor KD. Silymarin in the treatment of patients with primary sclerosing cholangitis: an open-label pilot study. Dig Dis Sci 2007;53:1716-1720.
27) Silveira MG, Torok NJ, Gossard AA, Keach JC, Jorgensen RA, Petz JL, Lindor KD. Minocycline in the treatment of patients with primary sclerosing cholangitis: results of a pilot study. Am J Gastroenterol 2009;104:83-88.
28) Martin CR, Blanco PG, Keach JC, Petz JL, Zaman MM, Bhaskar KR, et al. The safety and efficacy of oral docosahexaenoic acid supplementation for the treatment of primary sclerosing cholangitis - a pilot study. Aliment Pharmacol Ther 2011;35: 255-265.
29) **Tabibian JH, Weeding E**, Jorgensen RA, Petz JL, Keach JC, Talwalkar JA, et al. Randomised clinical trial: vancomycin or metronidazole in patients with primary sclerosing cholangitis—a pilot study. Aliment Pharmacol Ther 2013;37:604-612.
30) Tabibian JH, Gossard AA, El-Youssef M, Eaton JE, Petz JL, Jorgensen RA, et al. Prospective clinical trial of rifaximin therapy for patients with primary sclerosing cholangitis. Am J Ther 2014 Jun 9. doi: [Epub ahead of print]
31) Fickert P, Wagner M, Marschall HU, Fuchsbichler A, Zollner G, Tsybrovskyy O, et al. 24-norursodeoxycholic acid is superior to ursodeoxycholic acid in the treatment of sclerosing cholangitis in Mdr2 (Abcb4) knockout mice. Gastroenterology 2006;130:465-481.
32) Barry-Hamilton V, Spangler R, Marshall D, McCauley S, Rodriguez HM, Oyasu M, et al. Allosteric inhibition of lysyl oxidase–like-2 impedes the development of a pathologic microenvironment. Nat Med 2010;16:1005-1013.
33) Dignass A, Eliakim R, Magro F, Maaser C, Chowers Y, Geboes K, et al. Second European evidence-based consensus on the diagnosis and management of ulcerative colitis. Part 1: definitions and diagnosis. J Crohns Colitis 2012;6:965-990.
34) Stanich PP, Björnsson E, Gossard AA, Enders F, Jorgensen R, Lindor KD. Alkaline phosphatase normalization is associated with better prognosis in primary sclerosing cholangitis. Dig Liv Dis 2011;43:309-313.
35) Al Mamari S, Djordjevic J, Halliday JS, Chapman RW. Improvement of serum alkaline phosphatase to <1.5 upper limit of normal predicts better outcome and reduced risk of cholangiocarcinoma in primary sclerosing cholangitis. J Hepatol 2013;58: 329-334.
36) Lindström L, Hultcrantz R, Boberg KM, Friis-Liby I, Bergquist A. Association between reduced levels of alkaline phosphatase and survival times of patients with primary sclerosing cholangitis. Clin Gastroenterol Hepatol 2013;11:841-846.
37) Rupp C, Rössler A, Halibasic E, Sauer P, Weiss KH, Friedrich K, et al. Reduction in alkaline phosphatase is associated with

longer survival in primary sclerosing cholangitis, independent of dominant stenosis. Aliment Pharmacol Ther 2014;40:1292-1301.
38) de Vries EM, Boonstra K, Leeflang M, Beuers U, Ponsioen CY. Normalization of alkaline phosphatase within the first year after diagnosis is of prognostic value in primary sclerosing cholangitis. Gastroenterology 2014;146(Suppl 1):S–907.
39) Vergniol J, Foucher J, Terrebonne E, Bernard PH, le Bail B, Merrouche W, et al. Noninvasive tests for fibrosis and liver stiffness predict 5-year outcomes of patients with chronic hepatitis C. Gastroenterology 2011;140:1970-1979.e3.
40) Corpechot C, Gaouar F, El Naggar A, Kemgang A, Wendum D, Poupon R, et al. Elastography are associated with severity of fibrosis and outcomes. Gastroenterology. 2014;146:970-979.
41) Wiesner RH, Grambsch PM, Rolland Dickson E. Primary sclerosing cholangitis: natural history, prognostic factors, and survival analysis. HEPATOLOGY 1989;10:430-436.
42) Olsson R, Hagerstrand I, Broomé U, Danielsson A. Sampling variability of percutaneous liver biopsy in primary sclerosing cholangitis. J Clin Pathol 1995;48:933-935.
43) Angulo P, Larson D, Therneau T. Time course of histological progression in PSC. Am. J. Gastroenterol 1999;94:3310-3313.
44) de Vries EM, Verheij J, Hubscher SG, Leeflang MM, Boonstra K, Beuers U, Ponsioen CY. Applicability and prognostic value of histologic scoring systems in primary sclerosing cholangitis. J Hepatol 2015;13:1212-1219.
45) Everhart JE, Wright EC, Goodman ZD, Dienstag JL, Hoefs JC, Kleiner DE, et al. Prognostic value of Ishak fibrosis stage: findings from the hepatitis C antiviral long-term treatment against cirrhosis trial. HEPATOLOGY 2009;51:585-594.
46) Roselli M, MacNaughtan J, Jalan R, Pinzani M. Beyond scoring: a modern interpretation of disease progression in chronic liver disease. Gut 2013;62:1-9.
47) Piccinino F, Sagnelli E, Pasquali G, Giusti G. Complications following percutaneous liver biopsy. J Hepatol 1986;2:165-173.
48) Poynard T, Ratziu V, Bedossa P. Appropriateness of liver biopsy. Can J Gastroenterol 2000;14:1-6.
49) Cadranel JF, Rufat P, Degos F; for the Group of Epidemiology of the French Association for the Study of the Liver. Practices of liver biopsy in France: results of a prospective nationwide survey. HEPATOLOGY 2000;32:477-481.
50) Lindor KD, Bru C, Jorgensen RA, Rakela J, Bordas JM, Gross JB, et al. The role of ultrasonography and automatic-needle biopsy in outpatient percutaneous liver biopsy. HEPATOLOGY 1996;23:1079-1083.
51) Li QH. Evaluating co-primary endpoints collectively in clinical trials. Biomed J 2009;51:137-145.
52) Tischendorf JJ, Hecker H, Krüger M, Manns MP, Meier PN. Characterization, outcome, and prognosis in 273 patients with primary sclerosing cholangitis: a single center study. Am J Gastroenterol 2007;102:107-114.
53) Parkes J, Guha IN, Roderick P, Harris S, Cross R, Manos MM, et al. Enhanced Liver Fibrosis (ELF) test accurately identifies liver fibrosis in patients with chronic hepatitis C. J Viral Hepat 2010;18:23-31.
54) Nobili V, Parkes J, Bottazzo G, Marcellini M, Cross R, Newman D, et al. Performance of ELF serum markers in predicting fibrosis stage in pediatric non-alcoholic fatty liver disease. Gastroenetrology 2009;136:160-167.
55) Parkes J, Roderick P, Harris S, Day C, Mutimer D, Collier J, et al. Enhanced liver fibrosis test can predict clinical outcomes in patients with chronic liver disease. Gut 2010;59:1245-1251.
56) **Vesterhus M, Hov JR**, Holm A, Schrumpf E, Nygård S, Godang K, et al. Enhanced liver fibrosis score predicts transplant-free survival in primary sclerosing cholangitis. HEPATOLOGY 2015;62:188-197.
57) Ponsioen C, Reitsma J, Boberg K, Aabakken L, Rauws E, Schrumpf E. Validation of a cholangiographic prognostic model in primary sclerosing cholangitis. Endoscopy 2010;42:742-747.
58) Ruiz A, Lemoinne S, Carrat F, Corpechot C, Chazouillères O, Arrivé L. Radiologic course of primary sclerosing cholangitis: assessment by three-dimensional magnetic resonance cholangiography and predictive features of progression. HEPATOLOGY 2013;59:242-250.
59) Nilsson H, Blomqvist L, Douglas L. Dynamic gadoxetate-enhanced MRI for the assessment of total and segmental liver function and volume in primary sclerosing cholangitis. J Magn Res Imaging 2014;39:879-886.
60) Prentice RL. Surrogate endpoints in clinical trials: definitions and operational criteria. Stat Med 1989;8:431-440.
61) Sanyal AJ, Friedman SL, McCullough AJ, Dimick L. Challenges and opportunities in drug and biomarker development for nonalcoholic steatohepatitis: findings and recommendations from an American Association for the Study of Liver Diseases (AASLD)–U.S. Food and Drug Administration (FDA) joint workshop. HEPATOLOGY 2014;61:1392-1405.

Author names in bold designate shared co-first authorship.