BBA - Molecular Basis of Disease 1864 (2018) 1410–1414

Contents lists available at ScienceDirect

BBA - Molecular Basis of Disease

journal homepage: www.elsevier.com/locate/bbadis

Review

Endpoints in the design of clinical trials for primary sclerosing cholangitis☆

Cyriel Y. Ponsioen

Department of Gastroenterology & Hepatology, Academic Medical Center, Amsterdam, The Netherlands

ARTICLE INFO

Keywords:
Primary sclerosing cholangitis
Biomarker
Surrogate endpoints
Clinical trial

ABSTRACT

Primary sclerosing cholangitis is an enigmatic disease affecting the bile ducts, eventually leading to liver failure necessitating liver transplantation in many cases. There is currently no therapy that has proven to halt disease progression. One of the reasons for this is the lack of proper endpoints to measure the effect of medical intervention on the course of the disease. Relevant clinical endpoints such as death or liver transplantation occur too infrequently in this orphan disease to be used as endpoints in phase 2 or 3 trials. It is therefore of utmost importance to identify appropriate surrogate endpoints that are reasonably likely to measure true clinical benefit. This article will discuss a number of surrogate endpoints that are likely candidates to serve this role. This article is part of a Special Issue entitled: Cholangiocytes in Health and Diseaseedited by Jesus Banales, Marco Marzioni, Nicholas LaRusso and Peter Jansen.

1. Introduction

In the past three decades, many existing and a few experimental drugs have been studied in primary sclerosing cholangitis (PSC), the majority of which only in pilot trials [1]. So far, no drug has shown convincing evidence of halting disease progression or alleviating symptoms. Major reason for the lack of progression in drug development is our poor understanding of the pathophysiology of this enigmatic disease. Another factor is the lack of appropriate outcome measures. To obtain regulatory approval for a certain indication requires substantial evidence of benefit and evaluation of the risk versus benefit balance of the drug under study. Benefit is generally interpreted as showing improvement in clinical endpoints within the framework of adequately conducted clinical trials. Clinical endpoints measure how a patient feels, functions or survives. For conditions that are serious or life threatening and have an unmet medical need, the accelerated approval pathway in the US, or the conditional approval pathway in the European Union are available. These pathways allow drugs to be approved based on surrogate or intermediate clinical biomarkers that are *reasonably likely to predict clinical benefit.*

Clinical endpoints that measure improvement in disease course in PSC would be death, liver transplantation, occurrence of cholangiocarcinoma (CCA), and, possibly, development of cirrhosis. The first three qualify as level 1 endpoints according to the hierarchy as proposed by Fleming (Table 1), development of cirrhosis as level 2 [2]. Regulators commonly accept level 1 and 2 endpoints for registration purposes.

A big drawback for clinical trials in PSC is that, although it is a serious condition with a dismal prognosis, population-based median transplant-free survival is now estimated at around 21 years [3].This translates in an annual rate of clinically relevant events of < 4%. In the two largest clinical trials to date with a follow up of 5 years, both investigating ursodeoxycholic acid (UDCA) the annual event rate in the placebo groups was 3–6% [4,5]. One might expect that inclusion of hepatic decompensation, such as development of ascites, variceal hemorrhage, and/or encephalopathy, which would qualify as level 1 endpoint, into a composite endpoint may enhance the rate of events in clinical trials measuring the effect of a compound on disease progression in PSC. However, the high-dose UDC trial included in its composite endpoint also meeting minimal listing criteria for OLT, development of varices, and progression to cirrhosis, but despite this the annual event rate in the placebo group was approximately 6%. An annual event rate of 4% for placebo implies that for a clinical trial, even with an expected improvement in hazard rate of 50% in a 2:1 ratio almost 2800 patient-years of follow-up are necessary to attain a power of 80%. In an orphan disease as PSC this is simply not feasible. Therefore, there is an urgent need to identify surrogate endpoints that are reasonably likely to predict clinical benefit, (level 3).

Since 1995, there have been only four trials, all investigator initiated, that had a follow-up of two years or more and measured both clinical endpoints as well as surrogate parameters such as alkaline phosphatase (ALP), bilirubin, aspartate aminotransferase (AST), albumin, and histology [4–7]. In three trials there was no change in clinical endpoints and one trial showed worsening. In none of these

☆ This article is part of a Special Issue entitled: Cholangiocytes in Health and Diseaseedited by Jesus Banales, Marco Marzioni, Nicholas LaRusso and Peter Jansen.

E-mail address: c.y.ponsioen@amc.uva.nl.

http://dx.doi.org/10.1016/j.bbadis.2017.08.015
Received 21 May 2017; Received in revised form 13 August 2017; Accepted 15 August 2017
Available online 25 August 2017

Table 1
Hierarchy of (surrogate) endpoints [2].

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

studies a correlation of clinical outcome and surrogate endpoints was observed. Hence, these trials did not yield data to upgrade any of the level 3 surrogate endpoints to level 2. Of note, none of these trials had sufficient patient-years of follow-up. The longest trial, which was not able to attain its intended sample size, recruited 219 patients and had a follow-up of approximately 1000 patient-years, while the authors had estimated to require a total number of patient-years of 1730 to detect a rather bold treatment effect of 0.5. However, post-hoc analysis of the two longest trials with a follow-up of 5 years reveals a couple of interesting data regarding ALP, which will be discussed below.

When searching for surrogate endpoints that adequately reflect clinical outcome one is in fact trying to ascertain their what is called in psychometrics criterion validity, i.e. the extent to which the measure is truly related to the outcome. Two other concepts play an important role here: construct validity and content validity. The former pertains to the degree to which a test is truly measuring what it claims to be measuring; in other words, there is a biological rationale behind the test. For this, homogeneity in the outcome of interest may also be important. For instance, when assessing the criterion validity of a surrogate that is supposed to gauge the progression of PSC to endstage disease requiring OLT, one may not want to include cases that have received a transplant for intractable pruritus. However, given the low frequency of OLT, it is probably not feasible to be selective here. Content validity refers to the extent to which a measure captures all facets of a construct. Here, the problem of intercurrent cholangiocarcinoma may complicate the assessment of a biomarker for disease progression to endstage disease, where death and OLT would obviously coint as events. Presuming that the development of CCA is a chance event not related to disease stage would lead to treating death from CCA, or OLT for CCA in a competing risk analysis. The converse assumption would treat the occurrence of CCA as a relevant clinical endpoint of disease progression.

Here, a number of biomarkers that have been proposed by the International PSC Study Group as likely surrogate endpoint candidates will be discussed [1].

1.1. Biochemical biomarkers

ALP is the biochemical hallmark of cholestasis, and is invariably elevated in untreated PSC patients. In primary biliary cholangitis it has been demonstrated that the reduced canalicular secretion and hepatocellular retention of bile acids results in increased liver-ALP synthesis and/or release into the bloodstream rather than into bile [8]. As such it has construct validity when supposed to reflect cholestasis. It is still unclear to what extent ALP merely reflects impaired bile flow or also in part the degree of active bile duct inflammation. In PSC, studies have shown that after successful treatment of dominant strictures, cholestatic complaints are alleviated in parallel with decrease in ALP [9]. ALP has been used as an endpoint in all 26 clinical trials on PSC with published data in the past 22 years, either as (part of) the primary endpoint or as a secondary endpoint [1].

Post-hoc analysis of the high-dose Scandinavian UDCA trials with a follow-up of 5 years revealed that irrespective of the use of UDCA patients that during follow-up had a decrease of > 40% in their ALP level had a much better survival [10]. Likewise, post-hoc analysis of the high dose UDCA trial by Lindor et al. showed that both patients on UDCA as well as placebo treated individuals who showed normalisation of ALP during follow-up had a much better prognosis [11]. These data are corroborated by recent observational studies that all fuel the notion that ALP is a measure of prognosis [12–14].

Recently, data from the phase 2b simtuzumab trial have been presented at EASL 2017 in Amsterdam [15]. In this 96 month trial, 234 subjects were randomized in a 1:1:1 ratio to receive two different doses of weekly subcutaneous simtuzumab or placebo. Simtuzumab is a humanized IgG1 antibody against lysyl oxidase like-2, which is thought to play a central role in fibrosis formation. Primary outcome measure was the mean change in hepatic collagen content. Although the trial was negative, it yielded a lot of interesting data. With regard to ALP, baseline values were higher in patients who by week 96 had developed a PSC-related clinical event ($p < 0.001$). 30/191 biopsied subjects developed cirrhosis, which by itself also correlated with baseline ALP (Odds ratio 1.001, 95%CI 1.000–1.003). All these data point in the direction of ALP being a biomarker for disease progression, indicating at least a role for stratification into clinical trials and potentially as a measurement of treatment effect.

Recently, the Enhanced Liver Fibrosis panel (ELF) has been tested and validated in PSC. ELF contains a panel of three serum markers of fibrosis: hyaluronic acid, procollagen III aminoterminal peptide, and tissue inhibitor of metalloproteinase 1. In a Norwegian study comprising 308 patients, ELF was able to distinguish between mild and severe disease as reflected by transplant-free survival with an area under the curve of 0.81 (95%CI: 0.73–0.87) and optimal cutoff of 10.6 (sensitivity 70.2%, specificity 79.1%) [16]. An external cohort of 534 PSC patients from 7 countries validated the predictive utility of the ELF test for clinical outcomes. In the abovementioned simtuzumab trial the risk of a clinical event increased with higher baseline ELF score (hazard ratio 1.80; 95% CI: 1.43–2.26) and greater change n ELF at 12 weeks (hazard ratio 4.89; 95%CI: 2.76–8.96) [17]. Similar findings were observed for the prediction of progression to cirrhosis by baselie ELF (odds ratio 3.00; 95% CI: 1.97–4.57) and change at week 12 (odds ratio 3.04; 95% CI: 1.27–7.28). Interestingly, ELF can be elevated when there is acute inflammation or biliary obstruction. Therefore, it does not only measure fibrosis (construct validity) but also other biological processes that may affect outcome in liver disease, thereby possibly extending its content validity.

Trivedi et al. have recently published a study on the role of vascular adhesion protein1 (VAP1) in predicting clinical outcome [18].VAP1 is constitutively expressed in hepatic sinusoids and can promote the expression of mucosal addressin cell adhesion molecule-1 (MAdCAM-1). The latter is the ligand for the integrin $\alpha4\beta7$ expressed on gut homing mature T-cells, and as such is thought to play a keyrole in the aberrant gut-homing hypothesis, linking the gut to the biliary tree in PSC. In this paper, the authors found an association between elevated serum soluble VAP1 levels and transplant free survival (hazard rate 3.85; $p = 0.003$). This finding warrants further study.

1.2. Histology

Traditionally, liver biopsy has been the gold standard for clinical trials in many liver diseases, only to be supplanted recently by other biomarkers such as sustained viral remission in viral hepatitis trials, etc. Progression to cirrhosis on sequential liver biopsies has been accepted by regulatory authorities as a surrogate outcome measure for clinical trials in several liver diseases. Recently, the Ludwig, Ishak, and Nakanuma staging sytems were evaluated in PSC with regard to their prognostic value in PSC [19]. All three showed strong associations with transplant-free survival, as well as time to liver transplantation alone. The Nakanuma staging system showed the highest prognostic value. These results were subsequently validated in an international cohort study [20]. Interobserver variability proved adequate. A survey among expert liver pathologists involved in this study showed that all favored the Nakanuma staging system as the most appropriate scoring for PSC.