PROF. AFTAB ALA (Orcid ID : 0000-0001-6257-3160)

DR. THOMAS DAMGAARD SANDAHL (Orcid ID : 0000-0001-9807-6852)

PROF. PETER FERENCI (Orcid ID : 0000-0003-2306-1992)

Article type : Special Article

Title Page

HEP-21-0718

Designing clinical trials in Wilson disease.

By

Peter Ott[1*], Aftab Ala[2], Frederick K. Askari[3], Anna Czlonkowska[4], Ralf-Dieter Hilgers[5], Aurélia Poujois[6¤], Eve A. Roberts[7], Thomas Damgaard Sandahl[1*], Karl Heinz Weiss[8*], Peter Ferenci[9 #], Michael L. Schilsky[10,#]

[#]Shared last authorship.

*European Reference Network on Hepatological Diseases (ERN RARE-LIVER)

[¤]European Reference Network on Hereditary Metabolic Disorders (METABERN)

From

[1]Department of Hepatology and Gastroenterology, Aarhus University, Aarhus, Denmark

[2]Department of Gastroenterology and Hepatology, The Royal Surrey NHS Foundation Trust, Guildford, Surrey, UK; Department of Clinical and Experimental Medicine, University of Surrey, Guildford, Surrey, UK; Institute of Liver Studies, King's College Hospital, London, UK.

[3]Department of Medicine, University of Michigan, Ann Arbor, MI, USA

[4]Second Department of Neurology, Institute of Psychiatry and Neurology, Warsaw, Poland

[5]Department of Medical Statistics, RWTH Aachen University, Aachen, Germany

[6]Neurology Department and National Reference Centre for Wilson's Disease, Rothschild Foundation Hospital, Paris, France

[7]Departments of Paediatrics, Medicine, and Pharmacology and Toxicology, University of Toronto, Toronto, Ontario, Canada

[8]Salem Medical Center, Heidelberg, Germany and Department of Internal Medicine IV at University Hospital Heidelberg, Heidelberg, Germany,

[9]Department of Gastroenterology and Hepatology, Medical University of Vienna, Vienna, Austria

[10]Departments of Medicine and Surgery, Yale University Medical Center, New Haven, CT, USA

Authors e-mail addresses:

Peter Ott, Peterott@rm.dk
Aftab Ala, aftab.ala1@nhs.net
Frederick K. Askari, faskari@med.umich.edu
Anna Czlonkowska, czlonkow@ipin.edu.pl
Ralf-Dieter Hilgers, rhilgers@ukaachen.de
Aurélia Poujois, apoujois@for.paris
Eve A. Roberts, eve.roberts@Dal.Ca
Thomas Damgaard Sandahl, Thomsand@rm.dk
Karl Heinz Weiss, karlheinz.weiss@stadtmission-hd.de
Peter Ferenci, peter.ferenci@meduniwien.ac.at
Michael L. Schilsky michael.schilsky@yale.edu

Keywords:

RCT, endpoint, surrogate marker, Wilson severity score, copper metabolism

Word Count:

5998 words, including Main text and References

Foot note Page

Corresponding Author:

Peter Ott
Department of Hepatology and Gastroenterology,
Aarhus University,
Aarhus, Denmark
Phone: (+DK) 40 19 38 26
FAX:
E-mail: peterott@rm.dk

Abbreviations

APRI: AST/platelet ratio

CAP: Controlled Attenuation Parameter

CuEXC : Exchangeable copper

EMA: European Medicines Agency

FDA: Food and Drug administration

FIB-4: Fibrosis-4 index

KF: Kayser-Fleischer

MELD: Model of end stage liver disease

MRE: Magnetic resonance elastography

NCC: Non-ceruloplasmin bound copper

RCT: Randomized clinical Trial

REC: Relative exchangeable copper

SOC: Standard of care

UWDRS: Unified Wilson Disease rating scale

Conflict of interest statements,

Conflicts of interest: None directly related to the paper, but

PF: Has received funding from Alexion (former Wilson Therapeutics: copatent holder for TTM-bis choline), and is advisor for Univar, Vivet Therapeutics, Ambys, Ultragenix & Gilead.

AA: Is on the advisory boards of Alexion, Vivet, Univar, Ultragenyx. He has received unrestricted educational grants from Alexion, GMPO, Univar, and the Wilson Disease Association.

RDH: Has received research funding from Bayer & Regeneron.

FA: Is on the advisory boards of Alexion, Deep Genomics, Vivet Therapeutics, Ultragenyx, and has received research funding from Alexion, Ultragenyx and Intercept.

TDS: Is on the advisory boards of Alexion.

PO: Is on the speakers list for GMPO.

KHW: Advises for Alexion, Orphalan (GMPO) , Pfizer, Ultragenyx, Univar, Vivet therapeutics. KHW Has received research funding from Orphalan (GMPO), Univar, Alexion, MSD & Bayer.

EAR: None.

MLS: Grant support from GMPO, Alexion and the Wilson Disease Association

AP: Is on the advisory boards of Alexion, Vivet Therapeutics, Orphalan, Univar and has received research funding from Orphalan

AC: Advises for Alexion, Orphalan (GMPO), Pfizer, Ultragenyx, Univar, Vivet therapeutics. Has received research funding from Orphalan (GMPO), Univar, Alexion,

Financial support statement

The work was supported by an unrestricted grant from The Memorial Foundation of Manufacturer Vilhelm Pedersen & Wife. The foundation played no role in the planning or any other phase of the study.

Author contribution statement

PO, TDS, MLS, & PF conceived the paper. All authors contributed to the draft and critical revision of the paper. All authors approved the final version.

Acknowledgements:

This work was supported by The Memorial Foundation for Manufacturer Vilhelm Pedersen & Wife.

Ralf-Dieter Hilgers received funding from the European Union's 7th Framework Programme for research, technological development and demonstration under the IDEAL Grant Agreement no 602552, the European Union's Horizon 2020 research and innovation programme under the EJP RD COFUND-EJP no 825575 and ERICA under grant agreement No 964908.

ABSTRACT

Background and Aims: Wilson disease (WD) is an autosomal recessive disorder caused by *ATP7B* gene mutations leading to pathologic accumulation of copper in the liver and brain. Adoption of initial treatments for WD was based on empiric observations. These therapies are effective, but there are still are unmet needs for which new treatment modalities are being developed. Randomized controlled phase 3 studies are lacking for current WD treatments.

Approach and Results: The first Wilson Disease Aarhus Symposium (May 2019) included a workshop on randomized clinical trial (RCT) design. The authors of the paper were organizers or presented during this workshop and this article presents their consensus on the design of clinical trials for WD, addressing trial population, treatment comparators, inclusion and exclusion criteria and treatment endpoints. To achieve adequate recruitment of patients with this rare disorder, the study groups should include all clinical phenotypes and treatment-experienced as well as treatment-naive patients.

Conclusions: The primary study endpoint should be clinical or a composite endpoint until appropriate surrogate endpoints are validated. Standardization of clinical trials will permit pooling of data and allow for better treatment comparisons, as well as reduce the future numbers of patients needed per trial.

Wilson disease (WD) is an autosomal recessive disorder of reduced biliary copper excretion due to mutations in *ATP7B* leading to pathological copper accumulation in liver, brain and other tissues(1-3). Symptom onset is generally in adolescence to early adulthood but may occur at any age.

WD requires life-long therapy to prevent, reduce or stabilize symptoms(1,2). Current treatments were introduced without controlled studies. The chelators (d-penicillamine and trientine) that increase urinary copper excretion and zinc salts that decrease enteric copper absorption have raised WD patient survival to near-normal for age-matched populations with good adherence and initiation before severe organ damage(4-7). However, excellent outcomes are not universal. Up to 45% of patients have poor long-term medication adherence, with risk of disease progression(1,6, 8). Incomplete resolution of symptoms is common(4, 9,10). Medication side effects lead to cessation in many(11). Drug-induced paradoxical neurological deterioration may occur during initial treatment (4,12-14). Thus, developing new treatments for WD is necessary (Table 1).

The European Medicines Agency (EMA) and the US Food and Drug Administration (FDA) support development of drugs for rare diseases and published guidelines for trial design and analytical method development (16,17). Currently, only one controlled, partly blinded, short trial in neurologic WD was reported(14). The interpretation of other studies is limited by non-uniform definitions of outcomes(11). With more randomized studies expected in the future, utilizing uniform definitions of outcomes will facilitate study comparisons.

The first Wilson Aarhus Symposium (May 2019), included a workshop on randomized clinical trial (RCT) design. A diverse group of international experts contributed. In premeetings Drs. Ott, Ferenci, Weiss and Schilsky defined the most important issues and invited experts to address them at the meeting. This manuscript summarizes conclusions from the proceedings aimed at providing guidance for design and conduct of future WD Phase 2 and Phase 3 clinical trials. Study populations, outcome measures and needs for research are identified.

STUDY POPULATION

Clinical presentation

The *clinical phenotype of WD* patients at presentation is variable and includes acute and chronic symptoms (Figure 1). Since non-acute phenotypes are not clearly separated (3,15) studies should include all phenotypes except those with acute liver failure (ALF) or end-stage disease refractory to medical therapy (see Exclusion Criteria).

Clinically asymptomatic siblings of WD patients are effectively identified by genetic testing. Though appearing healthy, affected individuals have elevated hepatic copper and progress to overt disease without treatment. They can be included in clinical trials after clinical and metabolic characterization.

As *preconception*, *prenatal and newborn genetic testing* becomes more widespread, fetuses and newborn infants will be diagnosed before development of pathologic copper overload. Treatment will aim to prevent development of injury and disease(16). Drug selection and age for treatment initiation is unclear and require further study.

As disease course and underlying pathophysiology is similar, children ≥12 years-old and adults can be included in the same RCTs after appropriate bioethical considerations and dosing modification. Separate studies are needed for younger children.

Patient Genotype

Only approximately half of reported *ATP7B* mutations are considered pathogenic or likely pathogenic(17). Several studies failed to demonstrate clear genotype-phenotype correlations (1,15,18). Siblings and even monozygotic twins may have diverse phenotypes(19). Thus, stratification based on genotype is not reasonable except in future trials for gene repair.

Treatment status

Ideally, an RCT includes only treatment-naïve patients. However, the recruitment phase may be unacceptably long, even if patients treated for<28 days are included. Most current studies therefore included both treatment-naïve and treatment-experienced patients. The ratio should be balanced since clinical and biochemical improvement is more likely in the treatment-naïve. It is reasonable to stratify analysis of the treatment-experienced patients to <3 or ≥3-years because biochemically and clinical stability is more likely after 3 years (20). A run-in period on current treatment is recommended for treated patients to ensure baseline compliance, reduce study drop-out and help standardize data collection(21). The most common design is to randomize treatment-experienced patients to the trial drug or the patient's current treatment. While this pragmatic choice is supported by the author panel, certain possible biases must be taken into account. Because using a run-in period and requiring clinical stability (see Section Stability below) likely ensure that current treatment is optimized at inclusion, the trial drug may be held to a higher standard than if only treatment-naïve patients were studied. At the same time, double-blinding may be difficult and more costly; however, certain outcome measures can be obtained in a single-blinded fashion.

INCLUSION CRITERIA

Diagnosis

The diagnosis of WD should rest on standardized, validated diagnostic criteria such as the Leipzig score (22,23). A liver biopsy is not requisite for WD diagnosis or for inclusion in a clinical trial unless study endpoints include hepatic copper content or histology.

Stability

Inclusion criteria may require clinical and biochemical stability, however a generally accepted definition is lacking. After 3-years of uninterrupted treatment, further symptom regression is unlikely and clinical condition, treatment dosing and measures of copper metabolism are usually stable. For patients with <3-years treatment, the definition of stability should leave room for possible symptom regression. Some patients will never be stable despite treatment ('treatment failures') and RCTs with that specific focus are needed.

EXCLUSION CRITERIA SPECIFIC TO WD

Acute liver failure

Patients with ALF or at high risk of ALF should be excluded from pharmacologic trials. Use of the "New Wilson Index for Predicting Mortality" developed for WD children presenting with liver failure(24) can help identify these individuals. As a score of $\geq$11 predicted death, we recommend excluding WD patients with a score >10 despite limited data in adults(25).

End-stage liver disease

Patients with clinical instability due to refractory ascites, overt hepatic encephalopathy or gastroesophageal variceal bleeding within 6-months should be excluded unless treated and stabilized. Hepatocellular carcinoma and cholangiocarcinoma should also exclude enrolment. Patients with compensated cirrhosis may be included. Patients with a waiting time for transplant >1 year can be enrolled. Liver transplantation should be an exclusion criterion.

Neurological end-stage disease

Patients with marked disabilities may improve, and be included in an RCT. Those with severe neurological deficits (bedridden, fixed dystonia or parkinsonism, severe cognitive impairment) non-responsive to treatment for >12 months should be excluded from treatment trials.

WITHDRAWAL CRITERIA FROM TRIALS

Patients should be withdrawn from treatment trials if they experience drug injury (ALT increases >5-10-fold normal or hyperbilirubinemia >2-fold normal); worsening of cirrhosis (new onset of ascites, encephalopathy, variceal bleeding and/or jaundice); neurologic deterioration [i.e. by a predefined increase in United Wilson Disease Rating Scale (UWDRS)]; or significant psychiatric disease such as the onset of psychosis, severe depression or behavioral changes.

Paradoxical neurologic deterioration has been described as rapid neurological worsening within the 6 months of start of an initial or secondary treatment(13). If this is not defined as a treatment failure *per protocol*, the protocol should provide concise instructions about dose reduction and subsequent reduced rate of dose escalation

ENDPOINTS

Clinically important or surrogate endpoints

The primary endpoint should define the effectiveness of treatment. It is needed for power calculations to determine the number of patients needed for the trial. In a phase 3 trial, the EMA states that "*ideally a ´hard´ and clinically relevant endpoint"* is used as the primary endpoint variable(26). We define "*a clinically important*" endpoint as a clinical effect of treatment on how the patient feels, functions and survives (27,28). This endpoint should be objectively measurable, reflect important aspects of clinical disease progression and have a meaningful relation to the patients' quality of life (27,28).

Surrogate endpoints must be validated to ensure they adequately reflect the clinically important outcome. Their use as a primary endpoint may shorten the study duration and reduce the sample size. As discussed below, surrogate endpoints meeting these criteria are lacking for WD. Identifying surrogate endpoints should be prioritized in future work.

Surrogate markers are chosen because of their relation to the pathophysiology and disease natural history(26); however, they are insufficient to verify long-term patient benefit.

Exploratory endpoints are included to better estimate the efficacy and confirm the mechanism of action of treatments.

*Composite endpoint*s combining different endpoints are necessary when a single meaningful primary endpoint cannot be defined. Use of *multiple simultaneous endpoints*, clinical or biochemical, may be necessary despite a less clear interpretation (26).

Hepatic endpoints

Endpoints should relate to the goals of treatment. On treatment, patients with near-normal histology or minimal steatosis should remain stable, whereas those with inflammation, fibrosis or cirrhosis should improve or at least remain stable (Figure 1). Markers of treatment failure include fibrosis progression, cirrhotic decompensation or liver failure requiring transplantation or causing death.

Clinically important endpoints and potential surrogates

Routine laboratory parameters:

Alanine amino transferase (ALT) and aspartate aminotransferase (AST) are markers of hepatocellular necrosis and should be included as secondary or exploratory endpoints and measured for monitoring treatment safety. Biomarkers of liver protein synthesis (albumin, INR, pseudocholinesterase) and excretion (bilirubin) should be included as estimates for liver function. These parameters form part of scoring systems for those patients with cirrhosis such as the MELD score and Childs-Pugh score.

Development of fibrosis

Change in hepatic fibrosis is a potentially useful endpoint and can be included as a secondary or exploratory endpoint or as part of a composite endpoint.

The best way to assess hepatic fibrosis is uncertain, but includes histologic grading, elastography (sound wave or obtained by MRI) and biochemical methods. In WD liver biopsy may be less useful because histological findings did not clearly differentiate between progressors and non-progressors in past trials (29,30) and some patients hesitate to undergo biopsy. Transient elastography is a potential surrogate endpoint (Supplement S.1.1) but prospective studies of the rate of fibrosis progression/regression in WD are needed. Until then it is recommended as a surrogate marker. Noninvasive biochemical markers of fibrosis such as AST/platelet ratio (APRI), FIB-4 index and Enhanced Liver Fibrosis index, are less sensitive than elastography and may be included as exploratory endpoints (Supplement S1.1).

Further developments of MR methodology assessing hepatic fibro-inflammation, steatosis, and iron content may be of interest as exploratory endpoints (Supplement S.1.1)

Progression to cirrhosis and its complications

Development of complications of cirrhosis evolve slowly but are clinically important as endpoints in studies with long-term duration. Ideally, on treatment they may improve (i.e. disappearance of ascites or esophageal varices), but worsening may lead to study withdrawal (See Withdrawal section).

For WD patients with cirrhosis, validated prognostic information can be obtained using the Child-Pugh(31), MELD(32), and MELD-Na scores(33). These scores could be included as surrogate markers of liver disease progression or regression on treatment; however, there are no supportive data for their use in WD patients without cirrhosis.

Other possible surrogate hepatic markers

One treatment target is the reduction or prevention of inflammation. Biomarkers for hepatic inflammation need to be developed since ALT alone is insufficient (Supplement S.1.2).

The “New Wilson Index for Predicting Mortality” (24) discussed above (see Exclusion Criteria) may be a useful endpoint for safety as a rising score may portend severe liver injury since the score captures elements of SIRS and acute phase injury.

The potential use of quantitative dynamic liver function tests described in the Supplement (S.1.3) as surrogate endpoints should be evaluated.

Neurological endpoints

Neurological manifestations of WD can be classified into syndrome types based on predominant signs, such as tremor, ataxia, bradykinesia (parkinsonism-like) and dystonia. The choice of neurological endpoint should encompass this wide variability. This consideration led to the development of scoring systems for assessment of neurologic status in clinical trials (3,34-37).

The Unified Wilson Disease Rating Scale

The Unified Wilson's Disease Rating Scale (UWDRS) is a widely used scoring system for WD (35-37). As described in more detail in Supplement S.2.1., Part I of UWDRS assesses consciousness, Part II is a patient reported evaluation of disability and Part III a rater determined neurological examination. The use of UWDRS can be blinded if the assessor is unaware of the treatment. Video recordings may allow a centralized evaluation. The inter-observer agreement is sufficient to permit use of single observer assessments (See Supplement S.2.1). A possible limitation of Part III is if the total score is used to estimate disease severity, a positive change in one item (i.e. handwriting) can neutralize a negative change in another (i.e. speech) which may not be equivalent for the patient. Analysis of elements of the UWDRS is indicated to determine which are most relevant to patient functionality.

The use of a common neurological rating scale will facilitate comparison between studies, and presently we recommend the use of UWDRS. However, less complex and time-consuming measurements of the patients' neurological functional status are desirable. UWDRS Part II may be of interest since it is less time consuming (patient reported) and correlated with UWDRS III(37). The modified Rankin score (38) deserves further evaluation since it correlated with UWDRS after liver transplantation for neurological WD(39).

Surrogate neurological endpoints and markers

Magnetic resonance imaging (MRI)

To be a useful surrogate endpoint, MRI findings require objective and reproducible evaluation parameters and longitudinal studies demonstrating that MRI changes correlate with clinical findings in individual patients. Until such data are available, the use of MRI in clinical trials is exploratory. Validation of MRI is in progress (Supplemental S.2.2). Importantly, it would allow for blinded, centralized evaluation.

Other possible clinical neurological endpoints

Small interesting reports suggest a possible value of evoked potentials (Supplement S.2.3) but further studies are needed.

Psychiatric and other endpoints

Psychiatric manifestations of WD are relevant as study endpoint(s) since they affect quality of life (40). At the present time, with no validated instruments specific for WD available, we recommend the use of a simple standardized questionnaires, such as the Patient Health Questionnaire-9 (PHQ-9). Any treatment of psychiatric disorders should be monitored during trials. Psychometric

test batteries may be useful to detect subtle changes in cognition and/or psychomotor performance, but need validation in WD (Supplement S.2.3)

Other less common symptoms of WD such as arthropathy, female reproductive abnormalities, renal and skin disturbances may be considered as tertiary or exploratory endpoints.

COPPER METABOLISM AND STUDY ENDPOINTS

For treatments modifying copper metabolism, their impact on copper metabolism should be a focus of phase 1 and 2 trials while phase 3 trials should focus on the impact of the treatment on clinical outcomes. At present, none of the measures of copper metabolism discussed below is validated as surrogate endpoints since there still is a need to demonstrate that with treatment they have a positive correlation with good clinical outcome (20).

Meassurements of bioavailable copper.

Determination of the “free” bioavailable copper concentration has been proposed as a possible surrogate marker. This copper fraction is considered biologically active and is the target of treatment to prevent the extrahepatic uptake of copper. There are several approaches to measure “free” copper:

Non-ceruloplasmin copper (NCC) is estimated by subtracting ceruloplasmin-bound copper from the total serum copper concentration (1,2). A weakness of the methodology is biologically implausible negative values in some patients (Supplement S.3.1). Reports on the correlation between NCC normalization and clinical outcome are conflicting(41-43), but in a recent phase 2 study the NCC estimate correlated with clinical outcome during treatment with bis-choline tetrathiomolybdate(44).

Measurement of exchangeable copper (CuEXC) is obtained by incubation of serum with EDTA to remove loosely bound copper and subsequent removal of ceruloplasmin-bound copper by ultrafiltration(49). The method does not depend on the measurement of ceruloplasmin. Correlation between CuEXC and organ damage was observed in an animal study(45). CuEXC was related to patient compliance(46) but longitudinal data in patients have not been reported.

For further discussion of the measurement of bioavailable (“free”) copper, see Supplement (S.3.1). At the present stage neither NCC nor CuEXC has been validated as a surrogate endpoint. The data does not allow a conclusion as to which is more valuable for treatment monitoring. At least one of these should be included as an exploratory endpoint.

Newly reported mass spectrometry based methods that directly measure ceruloplasmin copper and total copper (47). This method suggest weaknesses with both the estimation of NCC and CuEXC methodologies.

24-h urinary copper excretion

The 24-h urinary Cu excretion is used for diagnosis and treatment monitoring of WD (1,2,23,43). Symptomatic patients treated with cupriuretic chelators have initial increases in copper excretion that decreases with time but remains above the normal range (20,43,48). The intra-individual variation is pronounced but lower than with NCC(43). Because 24-h urine Cu excretion is dose-dependent, reflects dietary intake and total body copper content(48), it is useful to monitor the treatment of a given patient assuming relatively consistent dietary copper intake.

With zinc therapy (no cupriuretic effect) the pattern is different, and in those with elevated urinary Cu excretion, there is a slow decrease in Cu excretion that takes months to years to reach the normal range(42).

Measurement of 24-h urinary Cu excretion after a 48-hours “drug holiday” might overcome the problems of interpretation during chelator therapy (1,48,49), and may reflect whole-body copper in these individuals. In compliant patients, and for an individual patient, urinary Cu excretion after 48-h “drug holiday” reflects whole-body copper content and not differences in dosing or drug absorption, facilitating treatment comparisons(43,50).

24-hour urinary Cu excretion should be included in a clinical study as a surrogate marker. In studies including chelating agents collection after a 48 hours “drug holiday” may be preferred to facilitate comparison between treatments.

Other possible exploratory endpoints related to copper metabolism

For long-term treatments changes in organ copper content may be ideal but very hard to obtain. One noninvasive approach is the quantification of Kayser-Fleischer (KF) rings intensity by orbital computed tomography (Supplement S.3.2). Measurement of copper in CSF is more invasive but may reflect cerebral copper burden (Supplement S.3.2). Hepatic copper concentration in liver biopsy samples is not useful for evaluating therapy because it may vary within the liver(51) and remains elevated despite clinical improvement (30,52).

Patient-reported outcomes

Quality of life and functional status are important efficacy measures of long-term therapy(28) but cannot be used as primary endpoints since their relation to long-term disease progression is unknown. The “minimal UWDRS” transformed into a patient reported outcome included 9 items

related to activities of daily living that correlated with UWDRS scores(37). It is recommended that a specific quality of life index for WD be developed and used as a secondary outcome measure until validation.

Choosing endpoints in clinical trials in WD

While copper parameters are useful primary endpoints in phase 2 studies, none has been validated to be used as primary surrogate endpoint in phase 3 studies. In these studies, which includes various clinical phenotypes, no single clinical endpoint would cover all situations (Fig 2). Therefore the primary endpoint must be a composite, including assessment of the most relevant clinical and biochemical features. The simplest form will include definitions of progression, regression or no change of disease. A more advanced composite endpoint would be a “WD severity score” including more parameters with weighting according to their impact on disease severity and patient functionality. Such a score may be a more sensitive composite primary endpoint in future RCTs in WD and would also be useful for validation of specific measures of copper metabolism for use as surrogate outcome measures.

TRIAL DESIGN CONSIDERATIONS

Choice of study drug and comparator for WD treatment in trials

With the currently available treatments for WD survival is near normal (4,6,7). Since clinical deterioration can develop within months after treatment discontinuation (53,54), placebo monotherapy as the comparator in an RCT is impossible. In current studies, standard of care (SOC) is used as comparator. New treatments can be directly compared to SOC or as add-on treatment to SOC alone. Ideally, SOC should be standardized for all study participants. In current trials SOC’s varies according to local traditions, economics, and differences in regulatory approval of medications.

Sample size

In any RCT the necessary sample size depends on the minimally relevant difference in treatment outcomes. Regulatory authorities(26) and the scientific community (55,56) recognize the need for innovative solutions to design and analyze clinical trials with as few participants as possible, especially for rare disorders. To increase the number of eligible patients, stratification should only be used if an impact on the outcome is expected (26,57). Also, the sample size may be reduced if

longitudinal evaluation of endpoint variables is applied using all available data rather than baseline to end-of-study comparisons(58).

Length of trial

The optimal trial duration is derived from knowledge of the natural history of the disease. When patients with WD were treated with chelators or zinc, partial normalization of ALT, albumin, and prothrombin time was seen after 6 months, and most patients reached values close to or within the normal range after 12–24 months (30,42). Histological normalization may take 6–10 years in adults (30,59), possibly shorter in children(60). Neurological symptoms will typically stabilize and start improving after 2 months of treatment, but improvements after 3 years are possible(14). Thus, studies with clinical endpoints may need trial durations of at least 1-2 years. The use of surrogate endpoints will help to shorten the trial length.

Specific designs

Depending on the specific aim of the study, the choice of study design will be influenced by the rarity of the disease and the availability of suitable study subjects.

Crossover trial designs reduce variability and thus the necessary sample size. The crossover design requires that the disease should not progress between periods and there should be no residual treatment effects. This may be difficult to fulfil in long-term studies in WD, but the design may be applicable to short-term studies.

Sequential designs have been developed for use in superiority studies and will often reduce sample size. The trial continues until superiority is demonstrated in one arm or until there is a certain number of included patients. Outcomes must be available shortly after the individual patient's trial termination.

Adaptive designs have specific advantages in rare diseases(55). With adaptive designs, the trial is separated into two or more independent phases in which an analysis described in the protocol can lead to protocol changes such as stop for futility or efficacy, or changes in sample size, endpoints, inclusion criteria or even removal or addition of new arms of active treatment(56, 61). With the *responsive-adaptive randomization* designs, the randomization ratios change during the trial according to the observed responses(26). A *flexible adaptive enrichment* design allows the trial to start with a large population of 'straightforward' patients. Based on the experience of the first part of the trial, more specific subpopulations are assessed in the second phase.

Even *more advanced solutions* are under development(56) including multi-arm sequential designs and the use of external data for the analysis of phase 3 trials(62). In the latter case it is important that the data obtained are of similar quality, i.e. according to the recommendations in this paper.

STATISTICAL METHODS

Small sample size studies require more complex statistical analysis than larger studies. Methods are being developed to deal with multiple endpoints, sensitivity analyses, adjustment for baseline variables and stratification and the evaluation of repeated measurements.

Bayesian methods may further increase the information that can be extracted from an RCT, although regulatory authorities will require validation of the 'prior beliefs' that will be included in the analysis. These methods may be most valuable for sample size estimations (63).

MONITORING

For monitoring the clinical and biological improvement, adherence and safety of a new treatment, the frequency of visits must take into account the disease phase and severity (20). During the initial phase of treatment after diagnosis, symptomatic patients should be assessed every 2-4 weeks for 2 months and then at 2-3 months intervals until the end of the first year. If a treatment naïve patient is randomized to receive SOC, dose modifications may be needed according to the drug selected. In the late phase of a trial, follow-up should be twice yearly, even in asymptomatic or stable patients (2). More frequent monitoring may be needed after treatment modifications or based on clinical indication.

Monitoring should also detect signs of overtreatment, such as neutropenia, sideroblastic anemia, hyperferritinemia and possibly hepatic iron accumulation. Overtreated patients are also expected to present with low serum copper, low NCC and CuEXC. They will also have low 24-h urinary copper relative to the treatment used. Urinary excretion after a 48-h drug holiday may be helpful in monitoring patients on d-penicillamine or trientine. In case of overtreatment, therapy should temporarily be discontinued, and, rarely, copper replaced.

CONCLUSIONS

With the ongoing development of new therapies for WD, we recommend that newly initiated clinical trials follow the above consensus guidance to improve the impact of the individual studies and facilitate their comparison. In order to achieve adequate recruitment, the trial population should include all clinical phenotypes and treatment-experienced and treatment-naive patients. The most

important clinical hepatic and neurologic endpoints were discussed. The primary study endpoint should be clinical or a composite endpoint, since no surrogate endpoints are validated. Centers around the world are urged to provide this validation since the use of surrogate endpoints would shorten trial duration and speed development of new therapies for WD.

Legends to Figures

Figure 1. Clinical course of WD.

After a subclinical period, WD presents with hepatic (mean age 17.6 years) and/or neurologic (mean age 23.4 years) symptoms. Approximately 60% have both. 3-5% presents with acute hepatic failure (ALF) which is fatal without liver transplantation. In the remaining patients, the medical treatment aims at preventing or stopping disease progression and if possible induce regression of symptoms.

Figure 2. : Proposed design for prospective, randomized phase 2 and phase 3 studies in Wilson disease.

Since currently there is no single endpoint describing all possible features of Wilson disease, we propose to develop a composite score (“severity score”) which includes and weights several clinical and laboratory parameters. Until then, a combination of changes of single parameters from baseline can be described as improved, unchanged or worse.

References

1.EASL. EASL Clinical Practice Guidelines: Wilson's disease. J Hepatol 2012;56:671-685.

2.Roberts EA, Schilsky ML. Diagnosis and treatment of Wilson disease: an update. Hepatology 2008;47:2089-2111.

3.Czlonkowska A, Litwin T, Dusek P, Ferenci P, Lutsenko S, Medici V, Rybakowski JK, et al. Wilson disease. Nat Rev Dis Primers 2018;4:21.

4.Beinhardt S, Leiss W, Stattermayer AF, Graziadei I, Zoller H, Stauber R, Maieron A, et al. Long-term outcomes of patients with Wilson disease in a large Austrian cohort. Clin Gastroenterol Hepatol 2014;12:683-689.

5.Bruha R, Marecek Z, Pospisilova L, Nevsimalova S, Vitek L, Martasek P, Nevoral J, et al. Long-term follow-up of Wilson disease: natural history, treatment, mutations analysis and phenotypic correlation. Liver Int 2011;31:83-91.

6.Dziezyc K, Karlinski M, Litwin T, Czlonkowska A. Compliant treatment with anti-copper agents prevents clinically overt Wilson's disease in pre-symptomatic patients. Eur J Neurol 2014;21:332-337.

7.Czlonkowska A, Tarnacka B, Litwin T, Gajda J, Rodo M. Wilson's disease-cause of mortality in 164 patients during 1992-2003 observation period. J Neurol 2005;252:698-703.

8.Maselbas W, Czlonkowska A, Litwin T, Niewada M. Persistence with treatment for Wilson disease: a retrospective study. BMC Neurol 2019;19:278.

9.Weiss KH, Thurik F, Gotthardt DN, Schafer M, Teufel U, Wiegand F, Merle U, et al. Efficacy and safety of oral chelators in treatment of patients with Wilson disease. Clin Gastroenterol Hepatol 2013;11:1028-1035 e1021-1022.

10.Litwin T, Dziezyc K, Czlonkowska A. Wilson disease-treatment perspectives. Ann Transl Med 2019;7:S68.

11.Appenzeller-Herzog C, Mathes T, Heeres MLS, Weiss KH, Houwen RHJ, Ewald H. Comparative effectiveness of common therapies for Wilson disease: A systematic review and meta-analysis of controlled studies. Liver Int 2019;39:2136-2152.

12.Czlonkowska A, Litwin T, Karlinski M, Dziezyc K, Chabik G, Czerska M. D-penicillamine versus zinc sulfate as first-line therapy for Wilson's disease. Eur J Neurol 2014;21:599-606.

13.Litwin T, Dziezyc K, Karlinski M, Chabik G, Czepiel W, Czlonkowska A. Early neurological worsening in patients with Wilson's disease. J Neurol Sci 2015;355:162-167.

14.Brewer GJ, Askari F, Lorincz MT, Carlson M, Schilsky M, Kluin KJ, Hedera P, et al. Treatment of Wilson disease with ammonium tetrathiomolybdate: IV. Comparison of tetrathiomolybdate and trientine in a double-blind study of treatment of the neurologic presentation of Wilson disease. Arch Neurol 2006;63:521-527.

15.Ferenci P, Stremmel W, Czlonkowska A, Szalay F, Viveiros A, Stattermayer AF, Bruha R, et al. Age and Sex but Not ATP7B Genotype Effectively Influence the Clinical Phenotype of Wilson Disease. Hepatology 2019;69:1464-1476.

16.Valentino PL, Roberts EA, Beer S, Miloh T, Arnon R, Vittorio JM, Schilsky ML. Management of Wilson Disease Diagnosed in Infancy: An Appraisal of Available Experience to Generate Discussion. J Pediatr Gastroenterol Nutr 2020;70:547-554.

17.HGMD. The Human Gene Mutation Database. At: http://www.hgmd.cf.ac.uk/ac/index.php 2020; (Accessed 31/10-2020)

18.Ferenci P, Roberts EA. Defining Wilson disease phenotypes: from the patient to the bench and back again. Gastroenterology 2012;142:692-696.

19.Czlonkowska A, Gromadzka G, Chabik G. Monozygotic female twins discordant for phenotype of Wilson's disease. Mov Disord 2009;24:1066-1069.

20.Woimant F, Poujois A: Monitoring of medical therapy and copper endpoints. In: Wilson Disease Pathogenesis, Molecular mechanisms, Diagnosis, Treatment and Monitoring. . Weiss, K.H. and Schilsky, M. ed. Amsterdam: Academic Press, 2019; 223-232.

21.Huo X, Armitage J. Use of Run-in Periods in Randomized Trials. JAMA 2020;324:188-189.

22.Ferenci P, Caca K, Loudianos G, Mieli-Vergani G, Tanner S, Sternlieb I, Schilsky M, et al. Diagnosis and phenotypic classification of Wilson disease. Liver Int 2003;23:139-142.

23.Nicastro E, Ranucci G, Vajro P, Vegnente A, Iorio R. Re-evaluation of the diagnostic criteria for Wilson disease in children with mild liver disease. Hepatology 2010;52:1948-1956.

24.Dhawan A, Taylor RM, Cheeseman P, De Silva P, Katsiyiannakis L, Mieli-Vergani G. Wilson's disease in children: 37-year experience and revised King's score for liver transplantation. Liver Transpl 2005;11:441-448.

25.Petrasek J, Jirsa M, Sperl J, Kozak L, Taimr P, Spicak J, Filip K, et al. Revised King's College score for liver transplantation in adult patients with Wilson's disease. Liver Transpl 2007;13:55-61.

26.EMEA. Guideline on clinical trials in small populations. At: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-trials-small-populations_en.pdf, 2006; (Accessed 31/10-2020).

27.Biomarkers Definitions Working G. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 2001;69:89-95.

28.Consortium IRD. Patient-Centered Outcome Measures. Initiatives in the Field of Rare Diseases. At: https://www.irdirc.org/wp-content/uploads/2017/12/PCOM_Post-Workshop_Report_Final.pdf 2016 (Accessed 31/10-2020).

29.Sini M, Sorbello O, Sanna F, Battolu F, Civolani A, Fanni D, Faa G, et al. Histologic evolution and long-term outcome of Wilson's disease: results of a single-center experience. Eur J Gastroenterol Hepatol 2013;25:111-117.

30.Schilsky ML, Scheinberg IH, Sternlieb I. Prognosis of Wilsonian chronic active hepatitis. Gastroenterology 1991;100:762-767.

31.Peng Y, Qi X, Guo X. Child-Pugh Versus MELD Score for the Assessment of Prognosis in Liver Cirrhosis: A Systematic Review and Meta-Analysis of Observational Studies. Medicine (Baltimore) 2016;95:1-29.

32.Kamath PS, Wiesner RH, Malinchoc M, Kremers W, Therneau TM, Kosberg CL, D'Amico G, et al. A model to predict survival in patients with end-stage liver disease. Hepatology 2001;33:464-470.

33.Kim WR, Biggins SW, Kremers WK, Wiesner RH, Kamath PS, Benson JT, Edwards E, et al. Hyponatremia and mortality among patients on the liver-transplant waiting list. N Engl J Med 2008;359:1018-1026.

34.Aggarwal A, Aggarwal N, Nagral A, Jankharia G, Bhatt M. A novel Global Assessment Scale for Wilson's Disease (GAS for WD). Mov Disord 2009;24:509-518.

35.Czlonkowska A, Tarnacka B, Moller JC, Leinweber B, Bandmann O, Woimant F, Oertel WH. Unified Wilson's Disease Rating Scale - a proposal for the neurological scoring of Wilson's disease patients. Neurol Neurochir Pol 2007;41:1-12.

36.Leinweber B, Moller JC, Scherag A, Reuner U, Gunther P, Lang CJ, Schmidt HH, et al. Evaluation of the Unified Wilson's Disease Rating Scale (UWDRS) in German patients with treated Wilson's disease. Mov Disord 2008;23:54-62.

37.Volpert HM, Pfeiffenberger J, Groner JB, Stremmel W, Gotthardt DN, Schafer M, Weiss KH, et al. Comparative assessment of clinical rating scales in Wilson's disease. BMC Neurol 2017;17:Article ID 140.

38.de Haan R, Limburg M, Bossuyt P, van der Meulen J, Aaronson N. The clinical meaning of Rankin 'handicap' grades after stroke. Stroke 1995;26:2027-2030.

39.Poujois A, Sobesky R, Meissner WG, Brunet AS, Broussolle E, Laurencin C, Lion-Francois L, et al. Liver transplantation as a rescue therapy for severe neurologic forms of Wilson disease. Neurology 2020;94:e2189-e2202.

40.Dening TR, Berrios GE. Wilson's disease. Psychiatric symptoms in 195 cases. Arch Gen Psychiatry 1989;46:1126-1134.

41.Brewer GJ, Askari F, Dick RB, Sitterly J, Fink JK, Carlson M, Kluin KJ, et al. Treatment of Wilson's disease with tetrathiomolybdate: V. Control of free copper by tetrathiomolybdate and a comparison with trientine. Transl Res 2009;154:70-77.

42.Brewer GJ, Dick RD, Johnson VD, Brunberg JA, Kluin KJ, Fink JK. Treatment of Wilson's disease with zinc: XV long-term follow-up studies. J Lab Clin Med 1998;132:264-278.

43.Pfeiffenberger J, Lohse CM, Gotthardt D, Rupp C, Weiler M, Teufel U, Weiss KH, et al. Long-term evaluation of urinary copper excretion and non-caeruloplasmin associated copper in Wilson disease patients under medical treatment. J Inherit Metab Dis 2019;42:371-380.

44.Weiss KH, Askari FK, Czlonkowska A, Ferenci P, Bronstein JM, Bega D, Ala A, et al. Bis-choline tetrathiomolybdate in patients with Wilson's disease: an open-label, multicentre, phase 2 study. Lancet Gastroenterol Hepatol 2017;2:869-876.

45.Schmitt F, Podevin G, Poupon J, Roux J, Legras P, Trocello JM, Woimant F, et al. Evolution of exchangeable copper and relative exchangeable copper through the course of Wilson's disease in the Long Evans Cinnamon rat. PLoS One 2013;8:e82323.
46.Guillaud O, Brunet AS, Mallet I, Dumortier J, Pelosse M, Heissat S, Rivet C, et al. Relative exchangeable copper: A valuable tool for the diagnosis of Wilson disease. Liver Int 2018;38:350-357.

47.Solovyev N, Ala A, Schilsky M, Mills C, Willis K, Harrington CF. Biomedical copper speciation in relation to Wilson's disease using strong anion exchange chromatography coupled to triple quadrupole inductively coupled plasma mass spectrometry. Anal Chim Acta 2020;1098:27-36.

48.Walshe JM. The pattern of urinary copper excretion and its response to treatment in patients with Wilson's disease. QJM 2011;104:775-778.

49.Walshe JM. Monitoring copper in Wilson's disease. Adv Clin Chem 2010;50:151-163.

50.Dziezyc K, Litwin T, Chabik G, Czlonkowska A. Measurement of urinary copper excretion after 48-h d-penicillamine cessation as a compliance assessment in Wilson's disease. Funct Neurol 2015;30:264-268.

51.Liggi M, Mais C, Demurtas M, Sorbello O, Demelia E, Civolani A, Demelia L. Uneven distribution of hepatic copper concentration and diagnostic value of double-sample biopsy in Wilson's disease. Scand J Gastroenterol 2013;48:1452-1458.

52.Ferenci P, Steindl-Munda P, Vogel W, Jessner W, Gschwantler M, Stauber R, Datz C, et al. Diagnostic value of quantitative hepatic copper determination in patients with Wilson's Disease. Clin Gastroenterol Hepatol 2005;3:811-818.

53.Ping CC, Hassan Y, Aziz NA, Ghazali R, Awaisu A. Discontinuation of penicillamine in the absence of alternative orphan drugs (trientine-zinc): a case of decompensated liver cirrhosis in Wilson's disease. J Clin Pharm Ther 2007;32:101-107.

54.Scheinberg IH, Jaffe ME, Sternlieb I. The use of trientine in preventing the effects of interrupting penicillamine therapy in Wilson's disease. N Engl J Med 1987;317:209-213.

55.Day S, Jonker AH, Lau LPL, Hilgers RD, Irony I, Larsson K, Roes KC, et al. Recommendations for the design of small population clinical trials. Orphanet J Rare Dis 2018;13:195.

56.Hilgers RD, Bogdan M, Burman CF, Dette H, Karlsson M, Konig F, Male C, et al. Lessons learned from IDeAl - 33 recommendations from the IDeAl-net about design and analysis of small population clinical trials. Orphanet J Rare Dis 2018;13:77.

57.Hilgers RD, Manolov M, Heussen N, Rosenberger WF. Design and analysis of stratified clinical trials in the presence of bias. Stat Methods Med Res 2019:962280219846146.

58.Mohlenberghs G, Verbeke G: An Introduction to Generalized (Non)linear Mixed Models. In: De Boeck PW, M., ed. Explanatory Item Response Models. Statistics for Social Science and Piublic Policy. New Youk, NY: Springer 2004.

59.Cope-Yokoyama S, Finegold MJ, Sturniolo GC, Kim K, Mescoli C, Rugge M, Medici V. Wilson disease: histopathological correlations with treatment on follow-up liver biopsies. World J Gastroenterol 2010;16:1487-1494.

60.Arnon R, Calderon JF, Schilsky M, Emre S, Shneider BL. Wilson disease in children: serum aminotransferases and urinary copper on triethylene tetramine dihydrochloride (trientine) treatment. J Pediatr Gastroenterol Nutr 2007;44:596-602.

61.Bauer P, Brannath W. The advantages and disadvantages of adaptive designs for clinical trials. Drug Discov Today 2004;9:351-357.

62.Eichler HG, Bloechl-Daum B, Bauer P, Bretz F, Brown J, Hampson LV, Honig P, et al. "Threshold-crossing": A Useful Way to Establish the Counterfactual in Clinical Trials? Clin Pharmacol Ther 2016;100:699-712.

63.Brakenhoff TB, Roes K, Nikolakopoulos S. Bayesian sample size re-estimation using power priors. Stat Methods Med Res 2019;28:1664-1675.

Table 1. The medical needs in WD

• Complete reversal of symptoms is not always achieved.
• Some patients experience slow progression of disease during treatment.
• Unwanted effects may prevent use of the most effective drug.
• Long-term adherence to therapy is a major problem and may be related to unwanted drug effects, dosing, cold storage, cost, etc.
• Early drug-induced neurological deterioration has been reported with all available treatments.

Table 2. Endpoints in trials for patients with Wilson Disease

Hepatological Endpoints

- The *clinical important hepatological endpoints* include fibrosis progression and development of cirrhosis and its complications (ascites, esophageal varices, jaundice, hepatic encephalopathy).
- No measure has been validated as *a hepatological surrogate endpoint* but the likely candidate is fibrosis progression/regression assessed by transient elastography or MRE
- *Surrogate markers* should include clinical scores in cirrhosis (MELD, Child Pugh)
- *Exploratory endpoints* may include peripheral fibrosis markers [FIB-4 index, AST/platelet ratio~~n~~ (APRI), Enhanced liver index (ELF)], markers of inflammation and quantitative liver function tests (galactose elimination capacity, LiMax® test or lidocaine clearance test)
- *Exploratory endpoints* also include ALAT, AST and other liver function tests to monitor treatment safety

***Identified areas of research*:**

High priority:

- Prospective validation in large cohorts of WD patients of transient elastography (Fibroscan®, ARFE or MRE) as possible surrogate markers for fibrosis regression/progression and development of cirrhosis in the individual patient.

Others:

- Prospective validation of markers of inflammation and quantitative tests of liver function as endpoints.

Neurological endpoints

- The use of a common neurological rating scale will facilitate comparison between studies and is recommended
- At the present time the panel recommends the use of UWDRS as an important *neurological endpoint*
- No measure has been validated as a *neurological surrogate endpoint* or *surrogate marker*.
- *Exploratory endpoints* may include MRI, Evoked potentials, psychiatric disease and the use of drugs to treat psychiatric disease.

Identified areas of research:

High priority:

- Development of a neurological score that is less complex and with good correlation to the physical wellbeing of the patient
- Prospective validation in large cohorts of WD patients whether changes on MRI described in a reproducible way parallel clinical neurological development in the individual patient.
- Development of specific measures to evaluate psychiatric disease as well as Quality of Life in WD patients.

Others:

- Prospective validation of evoked potentials and cerebrospinal copper as endpoints.

Endpoints related to assessment of Copper metabolism

- No measure of copper metabolism has been validated as a *surrogate endpoint.* The most likely candidates are NCC, Exchangeable copper (CuEXC), and 24-h Urine Cu after at 48 hours “drug holiday”.
- The 24 h urine excretion on current treatment or after 48h “drug holiday” may be included as *surrogate marker*.
- *Exploratory endpoints* may include optical coherence tomographic assessment of KF ring intensity.

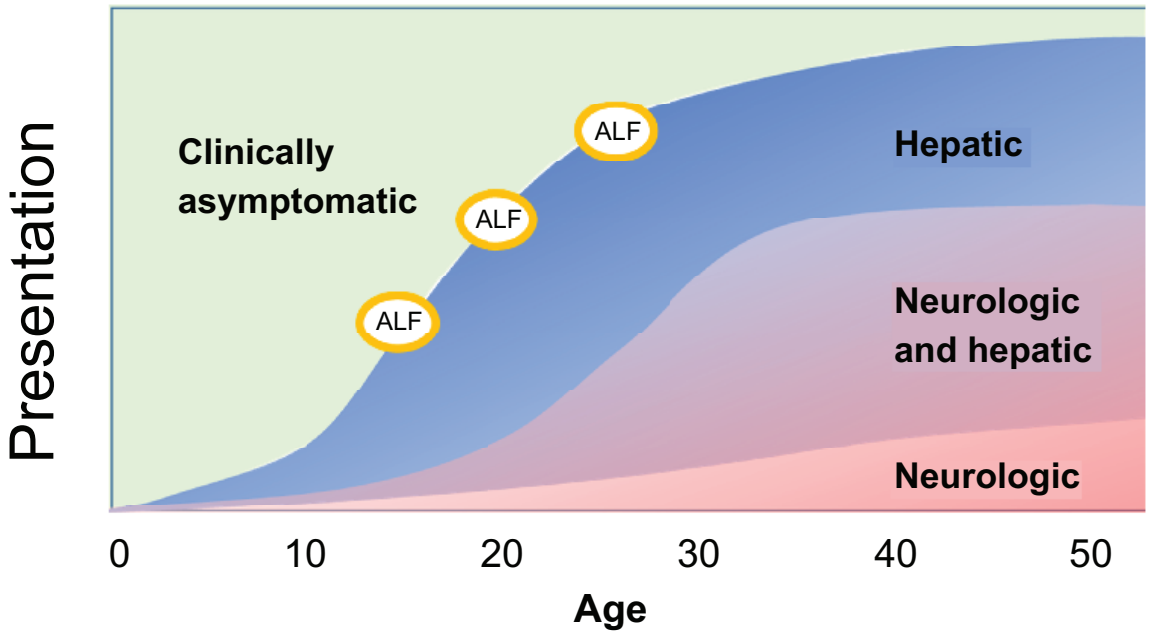
***Identified areas of research*:**

High priority:

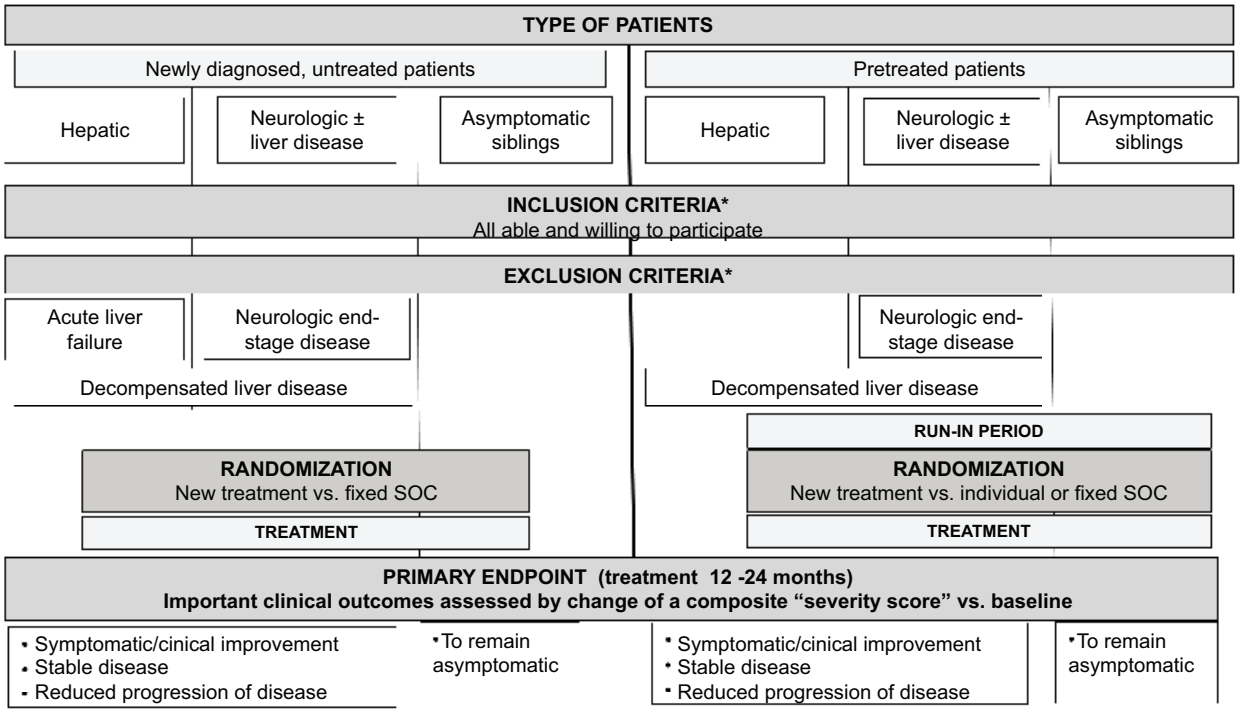
- Prospective validation in large cohorts of treated WD patients as to whether NCC, CuEXC or 24-h urinary Cu after 48-H drug holiday are predictive of important clinical endpoints.
- Development and validation of methods to quantify plasma Cu that is bioavailable.

Others:

- Prospective validation of assessment of KF ring intensity by use of optical coherence tomography as an endpoint
- Development of methods that quantifies intracellular effects of copper



hep_32074_f1.eps



hep_32074_f2.eps