THE FORUM FOR COLLABORATIVE RESEARCH

# LIVER FORUM 9

## Summary of Proceedings

**Wednesday, July 10, 2019**
**Paris, France**

1608 Rhode Island Avenue NW
Washington, DC 20036
www.forumresearch.org

THE FORUM
For Collaborative Research℠

Berkeley Public Health

# CONTENTS

# SESSION I: NASH CLINICAL TRIALS PART I - DEEP DIVE INTO RECENT TRIALS

## Lessons Learned from Trials to Date

**Presenter:** Mary Rinella, Northwestern University

**Slides:** https://bit.ly/3dpXWgf

Lessons Learned from Completed Phase 2 Trials

- PIVENS: Central pathology is critical
- FLINT: Fibrosis can improve with a metabolic MoA; ALT can predict histologic response
- GOLDEN: Milder disease and looser endpoint → high placebo response; Fibrosis improves when NASH improves
- NGM: Steatosis and ALT can improve quickly; Histologic response seen with 12-week biopsy endpoint
- Centaur: Surprising results due to compelling MoA; May have been offset by redundant pathways or limited target engagement
- Multiple studies: High screen failure rate → 50%; Resolution of NASH strongly tracks with fibrosis improvement

Endpoints in Early Phase 2

- ALT and PDFF
  - Currently used as initial signal of efficacy for new drugs
  - Ability to assess efficacy is dependent on the MoA
  - ALT: elevation associated with long-term mortality; and reduction results in histologic improvement, and correlates with resolution of NASH
  - MRI-PDFF: ≥5% absolute reduction and ≥30% relative reduction are benchmarks that have been established, but are based on limited data; and associated with histological improvement
- Steatosis
  - Easy to measure, with good agreement on histology and accurate non-invasive measures; degree of steatosis associated with increased metabolic risk
  - Improvement is not linked to disease outcomes; and steatosis decreases as the disease progresses
- Histologic read-outs of trials
  - ALT improves and tracks with every trial that has shown histologic improvement
  - Seladelpar trial: PDFF relative change from baseline in treatment arms was modest, but the placebo response was 20%, which is very high, compared with an average 5% seen across other trials. On the other hand, ALT dropped significantly, which also signals an important change.
  - REGENERATE: reductions seen in ALT, AST, and GGT, and a non-significant decrease in steatosis. Lobular inflammation and ballooning, components used to characterize the presence of steatohepatitis did significantly decrease.

Impact of Placebo Response

- NASH resolution has the lowest placebo response (12%), and NAS improvement >2 with no worsening of fibrosis has the highest amount of placebo response (28%). Fibrosis improvement of >1 stage is in the middle (16%)
- Factors influencing placebo response: disease activity at baseline, endpoint, weight-loss are being accounted for in trials currently. Other factors such as vitamin E use, dietary

macronutrients, changes in activity level / intensity of exercise, and alcohol intake are not being tracked very well.

Limitations of Current Histologic Endpoints

- Resolution of NASH with no worsening of fibrosis
  - Data presented from the REGENERATE trial demonstrated that the percent of patients achieving 'NASH resolution' differed significantly based on the endpoint definition:
    - NASH resolution with no worsening of fibrosis: (i) pathologist overall histopathologic assessment of "no fatty liver disease" or "fatty liver disease (simple or isolated steatosis) without steatohepatitis"; (ii) NAFLD Activity Score (NAS): hepatocellular ballooning = 0 and lobular inflammation = 0 or 1; and (iii) no increase in fibrosis stage from baseline]
      - No significant difference between either treatment arm and placebo
    - Gestalt: resolution of definite NASH with no worsening of fibrosis: (i) resolution of definite NASH (i.e., absence of steatohepatitis) based on pathologist overall diagnostic assessment and (ii) no worsening of fibrosis stage from baseline.
      - Significant difference between 25mg treatment arm and placebo
  - Despite significant improvement in the individual components of hepatocellular ballooning and lobular inflammation, the improvement in NASH resolution was not statistically significant.
    - Issues with inter- and intra-observer variability on individual components
    - NAS discriminates among steatohepatitis, but is not intended to replace a diagnosis.
- Reduction in fibrosis with no worsening of NASH
  - Data from the REGENERATE trial demonstrated a statistically significant reduction in fibrosis, using a definition of improvement in fibrosis by ≥1 stage and no worsening of lobular inflammation, hepatocellular ballooning, or steatosis
  - Data from MGL-3196 trial showed no difference between treatment arm and placebo in ≥1 stage fibrosis reduction when using traditional staining; however, there was a statistically significant improvement observed when using Second Harmonic Generation microscopy (automated quantification of fibrosis on liver biopsy)
    - Fibrosis should be measured on a continuous scale, as stages may not accurately reflect disease burden- the liver collagen burden is not linear across fibrosis stages and there is significant variability, especially at greater stages of fibrosis.

Clinical Trials in Populations with Cirrhosis

- Challenges in trials of cirrhosis include a prolonged compensated phase, and that patients with more advanced disease may reach outcomes quickly, but may be too advanced to see therapeutic efficacy.
- Data from the Simtuzumab trial[1] demonstrated no response in either treatment arm compared to placebo, for either ≥20% reduction in HVPG, or a reduction in HVPG to <10 mmHg.
  - Mean HVPG at trial entry was 12 mmHg, and 68% of patients had clinically significant portal hypertension (HVPG ≥ 10 mmHg)

[1] Harrison, S. et al. Simtuzumab Is Ineffective for Patients With Bridging Fibrosis or Compensated Cirrhosis Caused by Nonalcoholic Steatohepatitis. Gastroenterology. 2018;155(4):1140-1153.

- Additional data from the Simtuzumab trial[2] in patients with bridging fibrosis demonstrated that change in hepatic collagen content and ELF score (cut-off of 9.76) were both predictive of progression to cirrhosis.
    - Over 24-month follow-up period, 21.5% progressed to cirrhosis
- The Simtuzumab trial[2] in patients with compensated cirrhosis further validated that HVPG > 10 mmHg is predictive of clinical events, as well as baseline ELF score (cut-off of 11.27)
- Emricasan and Galectin[3] trials, both demonstrated no reduction in HVPG using intention-to-treat analysis.
    - In post-hoc analysis of the Emricasan data, patients with a higher HVPG (≥16) seemed to favor the treatment vs. placebo
    - In post-hoc analysis of the Galectin data, improvement was seen in patients without presence of varices and lower HVPG at baseline

Lessons Learned from Recent Trials

- Early phase trials: both steatosis and ALT can predict histological response- ALT has been more consistently predictive of histologic improvement.
- REGENERATE: success in phase 3 NASH trial is possible, but the histological endpoint of NASH resolution needs further refinement or reconsideration
- MADRIGAL: better mechanisms to measure fibrosis improvement on a linear scale are needed
- SIMTUZUMAB: huge amount of information learned about the natural history of NASH
- STELLAR 3/4, GALECTIN, EMRICISAN, CENICRIVIROC: additional appraisal of evidence prior to initiating phase 3 is needed; need to adapt more stringent stopping rolls so there are fewer failures; cirrhosis as a population needs to be sub-stratified

## Transitioning from Pre- to Post-Accelerated Approval

**Presenter:** Vlad Ratziu, Hôpital Pitié Salpêtrière

**Slides:** https://bit.ly/3730cc4

Regulatory pathways using surrogate endpoints

- Using surrogate endpoints allows for trials to be completed quicker than they would otherwise be able to, and with fewer patients.
- “Generally accepted surrogates” are endpoints that have been clearly linked in epidemiological and natural history studies to outcomes. The link to hard clinical outcomes is well-established, that if you improve the surrogate, you will improve the outcome.
    - Approval is based on whether the drug improves the ‘generally accepted surrogate’
    - There are not yet ‘generally accepted surrogates’ for NASH
    - Also called “validated surrogate endpoint”[4]
- “Reasonably likely surrogates” are endpoints that do not yet have definitive proof, but appear to be correlated with hard clinical outcomes.
    - Resolution of NASH and ≥ 1 stage reduction in fibrosis are the current endpoints ‘reasonably likely to predict clinical benefit’.
        - The link of these ‘reasonably likely’ surrogates to hard clinical outcomes, including progression to cirrhosis, is not fully established.

[2] Sanyal, A. et al. The Natural History of Advanced Fibrosis Due to Nonalcoholic Steatohepatitis: Data From the Simtuzumab Trials. Hepatology. 2019. [epub ahead of print]
[3] Chalasani, N. et al. A multicenter, randomized, double-blind, PLB-controlled trial of Galectin-3 inhibitor (GR-MD-02) in patients with NASH cirrhosis and portal hypertension. Journal of Hepatology. 2018;68(S1):S100-S101.
[4] BEST (Biomarkers, EndpointS, and other Tools): https://www.ncbi.nlm.nih.gov/books/NBK453484/

- Conditional approval: A drug must demonstrate that it improves one or both of the 'reasonably likely' surrogates. Full approval is pending the results of an outcome study which demonstrates that the surrogate improves hard clinical outcomes.
- The two 'reasonably likely to surrogates are achievable:
  - Resolution of NASH with no worsening of fibrosis has been achieved in 4 trials: Aramchol, Resmetirom, OCA, Elafibranor
  - ≥1 stage fibrosis reversal with no worsening of NASH has been achieved in 2 trials: Cenicriviroc, OCA

- Questions:
  - Does improvement in disease activity predict less progression to cirrhosis to the same extent as NASH resolution? What extent of improvement in disease activity is sufficient to slow down disease progression?
    - The level of disease activity is correlated with fibrosis progression
      - Data from NASH CRN (PIVENS and FLINT)[5] demonstrated that patients achieving a qualitative change in the diagnosis of NAFLD/NASH (i.e., going from definite NASH to probable NASH) had the highest improvement in fibrosis stage.
        - NASH resolution was the strongest predictor of fibrosis improvement
      - Data from GOLDEN trial, demonstrated nearly linear relationship between improvement in the components of disease activity (ballooning and inflammation) and progression of fibrosis.
        - Greater improvement in activity results in less progression in fibrosis.
    - By requiring that drugs induce resolution of NASH, there is the risk of missing drugs that may also be effective that may show a clinical benefit despite not fully resolving NASH.
      - Alternatively, trials require a stringent criteria of efficacy to prove that a drug works, and anything less than full resolution may not be convincing enough evidence.
  - What would be the impact of a trial that demonstrates resolution of NASH but has no impact on fibrosis?
    - Typically fibrosis improvement should follow resolution of NASH – but with trail duration reducing (2 years→1½ years→1 year), there may be insufficient time to see a significant improvement in fibrosis.
  - Does one stage improvement in fibrosis truly predict less progression to fibrosis?
    - What is ultimately important at the group level is how many patients worsen, not how many patients improve.
    - If a drug demonstrates a significant improvement in fibrosis for patients in a treatment arm compared with placebo, but the amount of patients that worsen remains the same, this is not a positive outcome as the same amount of patients will progress to cirrhosis, resulting in missing the endpoint and trial failure.
      - A marker that can take into account both the improvement and worsening in fibrosis is needed to be able to fully understand the drug's anti-fibrotic effect.
  - If the interim analysis is successful, should the remaining patients still be biopsied at the interim time point?

[5] Brunt, E. et al. Improvements in Histologic Features and Diagnosis Associated With Improvement in Fibrosis in Nonalcoholic Steatohepatitis: Results From the Nonalcoholic Steatohepatitis Clinical Research Network Treatment Trials. Hepatology. 2019; 70(3):522-531.

        - The interim analyses are typically done in about half of the population that will be included in the outcome trial, about 700-1000 patients.
        - Is there a benefit to biopsying patients after a successful interim analysis? Or rather would it be better to save the biopsy for later, so that patients may stay in the trial longer?
    - How many times do the "reasonably likely surrogates" need to be validated to move to "generally accepted surrogates"?
        - The question remains as to how many trials would need to show the relationship between the 'reasonably likely' surrogates and clinical outcomes for the surrogate to become 'generally accepted'.

Outcome/ post marketing confirmatory trials

- Trial retention in outcome trials becomes difficult because of the availability of an approved drug, trial fatigue, non-responder drop-out, and desire to not be on placebo
    - Approved drug: It is important to make the point that conditional approval is not definitive proof of efficacy of relevant endpoints, and is a 'temporary' designation; the outcome trial is necessary to make the link to clinical outcomes and for the drug to get final approval.
        - This applies to drugs in development, as well as drugs already approved for other indications (i.e., pioglitazone and vitamin E)
        - Particularly patients that have improvement at the interim analysis should be encouraged to continue in the trial instead of switching to an approved drug that they may not be a responder to.
    - Trial fatigue: trials need to work to simplify the follow-up in order to reduce burden on patients; at the same time, issues may arise in-between visits if there is too much time → 3-month visits are a reasonable length and reinforces adherence to the trial.
    - Non-responders: what if the interim analysis with the second biopsy shows no response?
        - Studies with biopsies done at 3, 6, 9, or 12 months might consider these patients to be 'slow responders', with not enough time to show improvement for some patients. Studies with biopsy done at 18 months is less likely to be the case of a slow response.
    - Placebo: once a drug is approved, patients may be concerned about being on placebo when there is a treatment option available. This can be addressed by designing trials so that there are fewer patients on placebo than in the treatment arm(s), such as 1:2 randomization. Additionally, reinforcing that the development of cirrhosis will be identified in the trial if there is indication that the disease has progressed.
- Completing an outcomes trial is crucial for the validation of the 'reasonably likely' surrogate, for definitive drug approval, and to justify long-term therapy.
    - Without a confirmatory trial, conditional approval may be withdrawn.
    - It remains to be determined whether efficacy results demonstrated from 1-2 year trials can be extrapolated to long-term therapy
        - Rosiglitazone study[6] demonstrated an anti-steatogenic response after 1-year, but continued treatment (measured by biopsy after an additional 2 years) did not show further reduction in steatosis, and prolonged therapy did not result in enhanced benefit.

[6] Ratziu, V. et al. Long-term efficacy of rosiglitazone in nonalcoholic steatohepatitis: results of the fatty liver improvement by rosiglitazone therapy (FLIRT 2) extension trial. Hepatology. 2010; 51(2):445-53.

        - In the Cenicriviroc study[7], the benefit seen after one year of the trial was not confirmed at two years (however there were many patients lost to follow-up in the active arm at the year two analysis).
        - Outcomes trials will offer standardized way to see whether histologic assessment years later will confirm the initial responses, or, show an enhancement in efficacy.
            - Determine the duration of treatment: many years vs. treatment until the timepoint that trial has shown efficacy.
    - Improved understanding of the natural history in the placebo arm, which might then eventually allow for the elimination of the placebo arm altogether.

Impact of conditional approval on the clinical landscape

- In the event of a drug receiving conditional approval, how would things change?
    - Early access programs
        - Some countries will likely set up programs to allow access to drug
        - Limited number of patients and likely to include the sickest patients and those that are excluded from clinical trials, as a result, this is not likely to impact the recruitment/retention of ongoing trials
    - Other clinical trials
        - Other trials may demonstrate an enhanced benefit compared to the approved drug, or may have a higher response rate, or work in a combination (either with approved drug or other drug).
            - This is particularly important to explain to patients when considering clinical trials.
    - Outcome/ confirmatory trials
        - Only by completing the outcome trial and demonstrating success will the drug be able to be considered standard of care.
            - Until this point is reached, there is no basis to replace the control arm with an arm of the conditionally approved drug.
    - Market availability
        - How will patients be selected to be prescribed the drug?
            - Histology: likely to be recommended, since liver biopsy is being used in the phase 3 trials, and this would allow providers to assess if the patients fit the same profile as those included in the trial.
            - Non-invasive: very likely that providers will use non-invasive methods for diagnosis despite that trials are built on histology.
                - Biomarker panels that are being developed to specifically fit definition of active NASH plus fibrosis (i.e., NIS 4)
                - Combination of elastometry, CAP, and AST
                - Clinical algorithms[8,9]
                    - The problem with non-invasive methods is the will likely never be better than 0.8 or 0.85 AUROC. At some point, will need to make a determination on what is ‘good enough’ for clinical practice.
                    - With variability of liver biopsy between 10-20%, achieving an AUROC of 0.99 or 1.0 is actually not the

[7] Friedman, SL. et al. A randomized, placebo-controlled trial of cenicriviroc for treatment of nonalcoholic steatohepatitis with fibrosis. Hepatology. 2018; 67(5):1754-1767.
[8] Younossi, ZM. et al. Algorithms Using Noninvasive Tests Can Accurately Identify Patients with Advanced Fibrosis Due to Nash: Data from the Stellar Clinical Trials. Hepatology. 2018; 68(6):1451A-2A.
[9] Anstee, QM. et al. Routinely Available Noninvasive Tests Discriminate Advanced Fibrosis Due to Nash in the Phase 3 Stellar Trials of the ASK1 Inhibitor Selonsertib. Hepatology. 2018; 68(S1): 953A.

best approach because the comparison is being made to an imperfect reference standard.

- Mathematical modeling[10] demonstrating that we may already have the best fibrosis markers (not NASH): "best scenario where liver biopsy accuracy is highest (sensitivity and specificity of biopsy are 90%) and the prevalence of significant disease 40%, the calculated AUROC would be 0.90 for a perfect marker (99% actual accuracy) which is within the range of what has already been observed".

- How will patients be assessed for treatment response?
  - Ideally, analyses of large phase 3 data will ultimately result in development of an algorithm that will be able to predict response
  - If ALT is increased at baseline, then reduction in ALT is a good marker of response.
  - Depending on mechanism of action, liver fat content could be justified to assess response of the drug.
  - Fibrosis markers could be used for long-term assessment.

## Discussion

C: It will be 7-8 years before trials readout on outcomes data, and it will be extremely difficult to keep patients on placebo, especially patients with more advanced disease. Need to re-think this concept, especially for development of combination approaches.

- Two phase 3 trials: REGENERATE and RESOLVE-IT, can chart the drop-out rate over time to see the impact; and once a drug is on the market, things will become more difficult.
- Drugs receiving conditional approval are not going to be standard of care until the final approval; however, there is off-label use of some drugs already and trials will have to be very creative with retention strategies.
  - Trials might also consider different statistical methodologies and new definitions of population estimates, trying to take into consideration patients that will take drugs that are already in the market.

Q: How much data/ how many patients/how long of follow-up is needed to be able to confidently understand the natural course of the disease and outcomes enough to be able to eliminate the placebo arm?

- Once the two currently ongoing phase 3 trials are complete, there will be a lot of data and a lot of information to analyze. This, in addition to the data from the NASH CRN, will help to address this question.
  - Data from the NASH CRN will be presented at the 2019 Liver Meeting which describes prospective clinical outcomes data from the registry
- There is a huge need to be able to do the phase IV clinical trials with a natural history control arm because retention is going to be a problem. Even if an investigator is able to 'convince' a patient to stay in the trial, when patients go home and talk to their family and/or their primary care doctor, they might change their mind.
  - It does not seem likely that the placebo arm in phase III trials could be replaced anytime soon because a natural history arm would not have safety data for the drug. Trials need to be conducted long enough to be able to adequately define the safety issues with a drug against a placebo arm.

[10] Mehta, SH. et al. Exceeding the limits of liver histology markers. Journal of Hepatology. 2009;50(1):36-41.

    - The phase IV trials will need longer natural history data then what is currently available from phase II and III. The NASH CRN and other longer-term datasets will be useful to address this issue.

Q: Currently, in LITMUS and NIMBLE, we are looking at comparison of non-invasive biomarkers to biopsies. Should we change the comparator, the biopsy, from qualitative to a quantitative? Do we have to redo the biopsies before we do the non-invasive biomarkers?

- One opinion is that NASH is a gestalt diagnosis and should be more qualitative and less quantitative.
    - Would be cautious about the gestalt approach- pathologists are struggling with the issue of reproducibility between pathologists, and the reproducibility of gestalt assessment of NASH is not known. 'Gestalt' is more a concept then a definition, and it needs to be further evaluated.
- With the classification of fibrosis, there is more weight placed on architectural changes compared with pure amount of fibrosis. The best predictor of fibrosis progression is fibrosis stage at baseline; however, the current NASH CRN classification stages are too broad. Stage 3 includes both small bridging septa, as well as very abundant septa (10-15 years of natural history represented). This scale should be further refined, vs. introducing something more quantitative.
    - Agree that there should be more granularity in the evaluation of fibrosis. A semi-quantitative score with automated analysis for example.
- As a field, there needs to be more transparency about how biopsies are collected and how they are interpreted. That will really help the field move forward a lot more quickly and a lot more consistently.
    - The Liver Forum could be instrumental in suggesting approaches towards analysis of liver biopsies in a consistent and transparent manner.
    - Agree that developing best practices for histologic assessment, particularly in advanced phase trials, would be very valuable.
        - Pros and cons of single assessment vs. dual assessment, vs. panel assessment, vs. taking the initial read and the end of study read, vs. mixing them all up and doing a reread.
        - Head-to-head, the trials are not comparable because they do not all follow the same format.
- Concern about the idea that clinical trials could be conducted without a liver biopsy. The screen failure rate in clinical trials due to patients not meeting histological criteria is greater than 50%, and these are patients that have been selected/ referred to the trial for screening from non-invasive biomarkers. With such a high screen failure, the non-invasive techniques currently being used are inadequate to select patients for trials. It will likely be a long time before biopsy can realistically be replaced.
    - Not all patients referred to clinical trials have gone through a rigorous non-invasive assessment- this may vary by trial site and by investigator.
    - If it becomes a choice for a patient to enroll in a trial and need to have a biopsy, vs. getting a prescription for a drug based on non-invasive measures, this will be another factor making trial recruitment and retention difficult.
        - If patients are selected for treatment based on non-invasive measures, we need to be thinking about how to monitor response, because while there are theories on measures that could work, there is a lack of data on this.
        - Data from REGENERATE trial should allow for comparison of changes in serum fibrosis markers in responders and non-responders.
            - Important to be able to see a change early-on, it would not be that helpful if a change is seen only after a year or more.
    - Only about 7% of those in clinical practice feel comfortable biopsying patients, so not realistic for physicians to biopsy patients before prescribing a treatment.

    - Alternatively, a non-invasive biomarker could be correlated with clinical outcomes and avoid the need to correlate to biopsy.

Q: For metabolic anti-steatotic mode of actions, we know that change in fat fraction is a good early signal to facilitate decision making. What are good signals to look for in early trails aiming to evaluate tolerability, safety, and efficacy in drugs with anti-inflammatory, anti-fibrotic modes of action?

- This will depend on the specific mechanism of action of the drug- it is currently unclear how to assess early efficacy in purely anti-fibrotic drugs, and also purely anti-inflammatory drugs.
    - Possibly PRO-C3 could be used to look at anti-fibrotic response
    - ALT would be difficult without an anti-inflammatory mechanism
    - PDFF would not work without a steatotic mechanism
- Very little is known about inflammation in the liver and the measures with which to assess it are very crude.

Q: Is there a linear correlation between ALT and NASH resolution? ALT and fibrosis resolution? And for the data on ALT improvement, is this driven mostly by people with high ALT?

- ALT and NASH resolution: a 10 IU decrease and a 17 IU decrease were shown in PIVENS and FLINT, respectively, to correlate with NASH resolution.
- ALT and fibrosis resolution: this remains unclear, no biological basis that a mitochondrial enzyme in the hepatocyte would have anything to do with fibrosis.
    - Although if you consider NASH resolution as the driver of fibrosis, there is an indirect biological relationship. Studies which have shown fibrosis benefit have also shown an improvement in ALT, but that is related to the improvement of NASH.
    - Early data from OCA showing trends in ALT predicting fibrosis response and early ALT changes predicting later stage.
        - Biopsy is likely underestimating some response – looking at ALT or other non-invasive markers of fibrosis, there is a correlation with histologic responders, but also an impact in non-responders as well. This suggests that histologic 'non-responders' might not actually be non-responders.
            - Perhaps they are slower to respond or their overall liver health is improved, but the biopsy didn't show a stage improvement in fibrosis
                - Represents issues with categorical nature of biopsy, timing of assessment, and whether histologic assessment really gives a full picture of underlying disease activity.
    - In placebo arm of CENTAUR study, changes in ALT and AST from baseline had a correlation with worsening of biopsies between baseline and year one. Though it is important to note that the study was not powered for this and the confidence intervals are wide.

Q: Is there any additional information on site effects from the GOLDEN-505 study, in terms of potential association with lifestyle recommendations provided, weight loss, macronutrient composition, or anything that could explain the site effect and high placebo response?

- Across different sites and study centers, and especially across different countries, the standard of care is very different. Weight loss, dieting, exercising, can impact on certain endpoints, particularly resolution of NASH.
    - To avoid a site effect from standard of care, it is important to have all arms of the trial in each trial site, that way if there is a site with very high level of standard of care, it is not only impacting the placebo arm.
- Patients recruited to the GOLDEN-505 trial had NAS ≥3 – the high placebo effect is primarily coming from the NAS 3's, which taught the field a valuable lesson about recruiting early-stage patients, as mild patients will resolve quicker.
- With the current definition of NASH resolution, the placebo response is between 6-13%, which is fairly low (vs. 30-40% in ulcer trials).

C: The basis for using NASH resolution as a surrogate is that biologically, steatohepatitis is underlying the disease, and therefore resolving the steatohepatitis will improve the disease state; and improvement in steatohepatitis is associated with improvement in fibrosis.

- Additionally, natural history studies show people with pure steatosis have fewer liver outcomes than people with steatohepatitis.
    - It's possible that patients transition back and forth between steatosis and steatohepatitis, and you are only catching different levels of activity at different points in time.
    - There is still a role for activity scores- which have greater reproducibility and also correlate with improvement in fibrosis.
    - Maybe activity score should be weighted – the MELD score is not just a combination of creatinine, bilirubin, and INR – the components are weighted based on relevance. For example, a modified NAS definition with similar criteria, with more weight given to ballooning might better characterize the extent of disease activity.

C: Looking at the Simtuzumab data, the progression of fibrosis from F3 to F4 was 20% within two years, and this is not aligned with what has been seen in thousands of patients in the NASH CRN. One theory is that since there were previously no trials for the advanced fibrosis population, there was a cohort of patients that had been ineligible for other trials and waiting, and once a trial for advanced fibrosis opened they were all enrolled. It is possible that this group of patients was further along in their disease.

- We do not know the natural history of progression because there is not a second biopsy in clinical practice at year two, so it is difficult to know if 20% progression is out of line or not.
- It's possible the patients were more advanced at baseline, or maybe only patients with clear bridging were selected, etc, but also consider the Simtuzumab trial for cirrhosis, where there was a 15% spontaneous reversal from cirrhosis to F3.
- The REGENERATE trial also had between 15-20% worsening of one stage of fibrosis after 18-months of follow-up, so not sure the Simtuzumab data is that unusual.
    - It depends on stage of fibrosis at baseline- F3 patients progress to F4 at a rate of about 5-10% per year. Within a 3-4 year timeframe, there are very few patients that progress from F2 to F4, which makes estimating sample size very challenging.
- By now, centers do not have a 'backlog' of patients waiting a long time to be referred to trials and are being recruited actively, which might have an impact on placebo response rates. For example: a patient goes to a diabetes clinic, gets a liver scan, is told liver stiffness is high and referred to trial screening for a biopsy. This may be the first time the patient is hearing that something is wrong with their liver, and now they are getting a liver biopsy and entering a trial – these events may spark patients into action where they change their lifestyle behavior.
    - Especially in early phase trials, which are 12-16 weeks, this could have a big impact on placebo variability. How patients are being recruited could make early phase trials more complicated.
    - MGL study recruited patients from diabetes sites that had not recruited NASH patients before and found that it was only the earlier stage patients that really had an impact, and it was really only the weight loss that had an impact. Sites can have a big effect on weight loss so that is important to be aware of and adjust for at both early and later stages.
    - REGENERATE trial did not see significant site interactions, and the global study results from over 350 sites were consistent with the US-based study with 9 sites.

C: All the drugs tested so far have very small effect sizes when you look at placebo adjusted response rates. This is a big struggle for the field - if we had drugs with very high effect sizes, this could overcome the issues of the placebo response rate.

- Using the example of ulcer trials mentioned previously, there was ~40% placebo response, but that was easily overcome because the drugs had such high response rates.

C: Are we holding the NASH field to an unnaturally high standard when it comes to biomarkers? Both LDL cholesterol and HbA1c are noisy biomarkers but are good enough to be included in clinical practice guidelines and inform how patients are treated.

- Another part of the problem is that AUROC is an artificial construct that implies sensitivity and specificity are equally weighted for each diagnostic, which is not correct.
- Just like more than one drug might be needed to treat NASH, perhaps a combination of markers will be needed.

C: most market research indicates that payers would accept costs around $8,000-$12,000 – and if the cost is higher, it seems likely they will require biopsy for access to treatment, which would be supported by the label.
C: there are changes in fibrosis biomarkers early on, however, it is not known how much these changes reflect a change in the histological stage of fibrosis, and there needs to be a distinction between markers of fibrogenic drive, and markers of collagen content in the organ.
C: there is a lot of focus on non-invasive markers of fibrosis, and not as much on disease activity- this is something that needs to be kept in mind and addressed.
C: there's currently a very interesting initiative on trying to develop a platform trial for NASH that some companies are engaged in. That will have the potential to introduce a real paradigm shift in the way we conduct clinical trials.
C: Vilar-Gomez data showed that patients needed to lose 5-10% of body weight in order to impact histology- is this magnitude of weight loss being seen in trials? Particularly when comparing 12-16 week studies looking at ALT and liver fat, vs. phase 2B or 3 studies based on biopsy. Is weight loss possibly not as relevant for the longer and later phase trials?

- The longer the trial, the less impact of weight loss because it cannot be sustained long term – even in trials for obesity. While still a consideration for short-term trials, it is less of an issue in phase 2B and 3 trials.
- It's important to note that in the Vilar-Gomez data, only 45% of patients that reached 10% weight loss had improvement in fibrosis – so there are also a majority of patients who were non-responders to weight loss.

# SESSION II: NASH CLINICAL TRIALS PART II - FUTURE DIRECTIONS

## Future of Trial Design and Endpoints for NASH

**Presenter:** Arun Sanyal, Virginia Commonwealth University

**Slides:** https://bit.ly/2GWFKyt

Objectives for NASH therapeutics

- Pathway of NASH development
    - Excess metabolic substrate delivery to liver, cytokines, microbial products
        - Cellular stress, if sever enough, will kill the cell
            - Regenerative response and inflammatory response
                - Persistent inflammation triggers fibrogenic response
                    - Development of cirrhosis
- Two treatment philosophies:
    - Target upstream drivers of disease – metabolic, cell stress, inflammation
    - Target fibrogenic engine and reduce the fibrosis progression
- Full understanding of a drug's mechanism of action is critical to understanding what it can biologically achieve.
    - For example, it would not make sense to have a steatosis endpoint for a pure anti-fibrotic or a drug working on the interplay of inflammation and fibrosis.

Outcomes and competing risks

- Prior to the development of cirrhosis, NASH outcomes are predominantly cardiovascular
- After the development of cirrhosis, liver outcomes become more prominent – though cardiovascular outcomes are still a major outcome.
    - This has led to the current 'generally accepted surrogate' of reduced progression to cirrhosis for pre-cirrhotic patients; and reduced outcomes for patients with cirrhosis.
- Historically, NASH drug development has been focused on the liver; however, it is part of a multi-system disease with multiple competing risks to the patient.
    - Patients with NASH frequently have T2DM, diastolic dysfunction, impaired renal function, cardiovascular disease.
    - Results in multiple competing risks for outcomes
        - Despite being very good now at treating variceal bleeding and ascites, mortality does not improve because the patient is still dealing with multiple other factors that also have increased risk of mortality.
- When designing trials, need to take into account the patient population and the natural course of the disease to understand what outcomes can be realistically expected.
    - Clinical outcomes occur in patients even at the fatty liver stage- though typically not liver-related- there is an increased risk of development of hypertension, chronic kidney disease, cardiovascular disease, and HCC.
    - Liver outcomes occur primarily later on in the disease.
- Considerations for clinical trials: mechanism of action
    - The approval path should be considered in the context of use for intervention and mechanism of action
        - Scenario A: the mechanism of action impacts the root cause of disease (obesity, insulin resistance, metabolic inflexibility, systemic metabolic inflammation, etc)
            - May have beneficial effects on diabetes, heart disease, kidney disease, and liver disease
            - Could extend what is currently being done for T2DM to include cardiac, liver, renal endpoints as appropriate for the mechanism of action
                - Conduct an outcomes trial instead of using surrogate endpoints
        - Scenario B: the mechanism of action affects liver disease activity with or without major systemic effects.
            - Pre-cirrhotic: improve liver disease activity in the short term, and reduce progression to cirrhosis in long term.
            - Cirrhotic: reduced progression to consideration for transplant or clinical decompensation.
        - Scenario C: the mechanism of action affects fibrosis without major systemic effects
            - Pre-cirrhotic: prevent progression to cirrhosis
            - Cirrhotic: prevent liver decompensation or reverse fibrosis
    - The choice of target should take into consideration the redundancies in pathogenic pathways
        - There are an increasing number of cellular processes that are affected on the pathogenic pathway as you move downstream from metabolic overload to the development of fibrosis.
            - By only focusing on a downstream component, it is unlikely to have much impact on the drivers of disease; though combination strategies could play a role.

Patient Populations: NASH vs NAFL

- It may be time to challenge the concept of NASH vs NAFL
  - Example of Chronic Persistent Hepatitis vs. Chronic Active Hepatitis
    - Were long considered to be two distinct clinical entities
    - One was a progressive disease and treated, the other was not progressive and not treated
    - Ultimately, it was determined to be the same disease with waxing and waning of disease activity.
  - NAFL vs NASH
    - NAFL is considered mild, NASH is considered more sever, more progressive
    - Longitudinally, the disease waxes in wanes
    - The same molecular mechanisms are involved in both, and the difference is in the level of severity
- Ultimately need better measures of disease activity- different opinions of what 'steatohepatitis' is depending on who is asked
- Fibrosis progresses across a continuum, and multiple stages coexist in the same liver – this will make separating F3 and F4 in a biomarker study impossible.
- Considerations for clinical trials: populations
  - Scenario A: need to consider a scenario where the systemic mechanism of action has an impact on multiple end-organs of interest.
    - Systemic risk factor profile: need to look at the severity status of individual end organs and holistically model outcomes based on the entire patient.
  - Scenario B: pre-cirrhotic population and the mechanism of action is anti-disease activity
    - Must have NAFLD with high activity scores – need enough disease activity, whether measured by NAS or SAF or a new scoring system that weights ballooning and inflammation more heavily than steatosis.
    - Consider removing NASH as a requirement- a gestalt diagnosis is not quantifiable and the Kappas for calling something 'NASH' are low
  - Scenario C: pre-cirrhotic population and the mechanism of action is anti-fibrotic
    - Stage 3 population is key target where prevention of cirrhosis is feasible and will have the biggest impact
    - Stage 2 population allows more patients to be included but will dilute the effect size and increase the sample size needed to show reduced progression to cirrhosis and outcomes

Endpoints

- If drug has a systemic mechanism of action where the hypothesis is that multiple end organs will improve, consider composite endpoint extending what has been done for type 2 diabetes drug development
  - Consider the use of estimands
  - Consider alternate models, such as based on desirability function
- It is time to push back and challenge traditional paradigm of anchoring disease assessment to histology
  - There is an argument to be made for using alternate surrogates to assess disease activity – accumulation of fat, cellular injury/death, and inflammation
  - Of liver histology, AST/ALT, MRI-PDFF, 2D MRE, and cT1, all have some amount of misclassification rate
    - Liver biopsy has a risk of death, lowest ease of deployment, and lowest patient preference.
    - Except for fibrosis, Kappa scores for histology components are modest/low, particularly for ballooning and lobular inflammation

- For drugs with a "front end (metabolic-cell stress)"-targeted MOA, reducing disease activity is the key to reducing disease progression
    - Both resolution of NASH, as well as improvement in NAS, robustly predict fibrosis improvement- should consider using something which can be reproduced more reliably if going to use a histological surrogate.
- Considerations for clinical trials: endpoints
    - Scenario A: pre-cirrhotic population and mechanism of action of anti-disease activity
        - Conditional approval could also potentially be considered on the basis of extensive safety data and some composite of non-invasive markers
        - Consider eliminating NASH resolution as an endpoint
            - Improvement in NAS is more reproducible and should be re-considered
        - Keep progression to cirrhosis as the generally accepted surrogate
    - Scenario B: pre-cirrhotic and cirrhotic population and mechanism of action is anti-fibrotic
        - Stage 3: consider reduced progression to cirrhosis, instead of fibrosis regression

## Estimands in NASH Clinical Trials

**Presenter:** Peter Mesenbrink, Novartis Pharmaceuticals

**Slides:** https://bit.ly/3nP7whq

Review of estimand concept

- The original ICH E9: Statistical Principles for Clinical Trials was developed in 1997
- To address gaps in the guidance, an addendum has been being developed to focus on estimands, and there have been position papers written by the FDA, EMA and PDMA
- An estimand precisely defines the treatment effect of interest in a clinical trial
    - In the clinical trial setting, the population, endpoint, and treatment comparison of interest is defined as a standard practice
    - However, trials often do not adequately account for the diversity in patient journeys, and how this may impact the assessment of clinical benefit of the investigational treatment
        - This is particularly relevant for NASH due to the heterogeneous patient population

Events impacting a patient with NASH in a clinical trial

- Intercurrent events are defined by the ICH E9/R1 guidance as "events that occur after treatment initiation and either preclude observation of the variable or affect its interpretation"
- Intercurrent events may be related to the disease, or may be related to the effects of the treatment – how can this be captured?
    - Patients may discontinue their treatment due to an adverse reaction
    - Patients may start alternative treatments before the endpoint of interest is measured or observed
    - Patients may experience a clinical event before the endpoint of interest is measured or observed (such as a cardiovascular event or renal failure)
        - Events such as liver transplant are difficult to interpret due to geographical differences
    - Patients may die due to causes unrelated to the underlying disease before the endpoint of interest is measured or observed
- These events may have a positive or negative effect on treatment, and this may not be directly captured through the endpoint or clinical outcome of interest.

    - Is it possible to collect the data? And is the data able to be accurately interpreted to understand the effect?
- Some of the events noted may impact the interpretation of the endpoint or clinical outcome of interest
    - Ex. Impact of changing type 2 diabetes medication
- Some events may prevent the actual measurement of the endpoint or clinical outcome
    - Ex. If a patient receives a liver transplantation, changes in fibrosis can no longer be measured
- The existence of intercurrent events raises the important question: how to define the treatment effect of interest ('estimand') in the study population, for the relevant primary endpoint in the presence of varied intercurrent events?
    - For a complex disease such as NASH, with many pharmacological targets, there will likely not be a 'one size fits all' solution, and will depend on the mechanism of action.

Example: T2DM – Medication Changes

- Patient has controlled T2DM, is on a stable dose of liraglutide, and continues on that dose for 6-months. After 6-months, there is a worsening of T2DM which requires an increase in dose of liraglutide.
    - Is the intercurrent event the change in T2DM medication, or the worsening of T2DM?
    - Could the increase in dose of liraglutide have an effect on liver histology efficacy endpoints?
    - If the assessment of the effect of Drug X is confounded, what should be the estimand of interest?
- How to handle:
    - Treatment Policy:
        - The treatment effect estimated remains the same, regardless of change in T2DM and/or worsening of disease
        - For histology endpoints, the treatment comparison focuses on the effect of the drug plus potential changes in medications over the time period, versus control and the potential change in the medication over the time period.
        - The assumption is that these changes will happen equally across treatment groups.
    - Hypothetical:
        - The treatment effect estimated in the hypothetical scenario where we assume that no changes occur in T2DM medication over the time period. Values of medication changes would be assumed missing and imputed based on observed data.
    - Composite:
        - The treatment effect estimated based on a composite endpoint where patients who change T2DM medication are considered non-responders to treatment.
        - Change the histology endpoint at 12/18 months which is already considering multiple scenarios to a more complex composite endpoint.
    - The strategy may be different depending on the treatment or combination of treatments that are being studied

Thinking process for developing estimands recommended in ICH E9 addendum (draft)

- Therapeutic setting and intent of treatment determine trial objective
- Identify the intercurrent events that could impact that trial objective
- Discuss strategies to address the intercurrent events
- Construct estimand(s)
- Align choices on trial design, data collection and method of estimation

- Identify assumptions for the main analysis and suitable sensitivity analyses to investigate these assumptions
- Document the chosen estimands

Estimands in NASH Clinical Trials Working Group

- Both the Standard of Care Comorbidity Management working group and the Standard of Care Lifestyle Management working groups are working on NASH estimands, though possibly didn't know it.
- New working group to focus specifically on estimands in NASH
    - Objectives of the group:
        - Identify what are the different types of estimands to be considered in NASH clinical trials (different estimands will be needed for short-term mid-term and long-term endpoints)
        - What are the points to consider in determining the best way to handle missing data caused by intercurrent events for the different types of estimands
        - What are the intercurrent events that should be collected in NASH clinical trials so that their impact on efficacy can be accounted for?
        - Develop a commentary manuscript for a peer-reviewed journal on NASH estimands and what should be considered when developing a new NASH therapy
        - Update Liver Forum on progress and importance of this on the future in the design of NASH clinical trials
        - Synergize with SoC Liver Forum working groups
    - Possible intercurrent events for histological endpoints (draft):
        - Changes in diet
        - Changes in exercise
        - Changes to dose in T2DM medication
        - Changes to dose in lipid lower medications
        - Changes to dose of Vitamin E if taken concomitantly
        - Stopping of treatment prematurely due to adverse event/reaction
        - Decreasing of dose of study treatment due to adverse event/reaction
        - Death
        - Liver Transplant
    - Possible intercurrent events for long-term clinical outcomes (draft):
        - Changes in baseline disease co-morbidities (new and/or worsening)
        - Changes to doses for concomitant medications to treat disease co-morbidities
        - Changes to dose in T2DM medication
        - Stopping of treatment prematurely due to adverse event/reaction
        - Decreasing of dose of study treatment permanently due to adverse event/reaction
        - Liver Transplant
        - Changes to standard of care for NASH and adding of a new approved NASH treatment to manage the condition

## Discussion

C: Regarding estimands, most of the items cited as being potential intercurrent events are known and can be dealt with/ adjusted for using current statistical methods. There seem to already be methods in place- though perhaps they are not sufficient. It would be interesting to show a real life example of a clinical trial and re-analyze the results taking into account the estimands and compare with the classical assessment - from any field, not necessarily NASH.

- This approach will likely be used in the development of the manuscript from the working group
- An important aspect of this approach is to have good data to demonstrate that the intercurrent events truly impact the natural course of the disease, and to what extent they have an impact.
- A goal would be to predict outcomes at the patient level and being able to have a better sense of the probability of an outcome based on baseline disease state.

C: A challenge for clinical trials will be as more patients enter trials from a primary care setting where frequently their comorbidities are not being adequately controlled.
C: Concern about changes in dietary habits in a trial and patients 'indulging' more because they are in a trial and on a drug

- Example of PCSK9 trials (cholesterol lowering), where patients in the trial stopped taking statins – which impacted the trial because the drugs worked better with a statin. In this case, the effect of the drug was powerful enough to show a difference, even with the change in statin use.
- What is an intercurrent event for one mechanism of action may not be for another- for example a GLP-1 agonist will have weight loss and diabetes improvement, which would not be appropriate to treat as an intercurrent event.

C: Must consider what is known to have an impact and what can be documented- all without being too restrictive on the trial implementation.

- This is particularly important for large clinical trials – smaller early phase trials with very defined settings and frequent patient visits, are very different from the larger phase 3 trials.
- Need to think carefully about practicality- taking the example of cardiovascular outcomes, based off diabetes trials, would need ~5,000-10,000 patients in a trial in order to control for CV outcomes and be able to assess primary outcomes of interest for NASH.
    - If cardiovascular disease is a common morbidity of patients, should it be part of the composite outcome itself? And consider preventing cardiovascular events from occurring as a clinical benefit.
- The mechanism of action is very important to consider; you would have different populations and not expect to see the same biological processes affected with liver-specific compounds vs. compounds with systemic effects.
    - NASH occurs in the context of multiple comorbidities that are linked pathophysiologically. Even if a drug is completely liver-centered, the patient still typically has multiple comorbidities that can improve or worsen and thus also alter the course of the liver disease.

Q: What happens if these intercurrent events are not taken into account? Is the concern that the difference versus placebo might not be there? Or, is it only the size effect that might be impacted?

- A little bit of both, particularly since the effect sizes are already small – difficult to know if the effect size is really due to the drug.
- This is important to consider for trials in the cirrhosis population, where multiple factors can result in mortality. Despite ~70 years of doing trials in the cirrhosis population, we cannot make gains in reducing mortality because of this exact problem. Because there are competing risks for mortality that are simultaneous and contemporaneous, even if one factor is addressed, another one kills the patient. If these approaches are not developed, drugs developed for cirrhosis are going to continue to fail.
- What is the point of a clinical trial, to maximize the effect size and efficacy, or to show efficacy in real life conditions?
    - Both – in addition to showing efficacy, need to an understanding of whether the drug is going to work in the real world.
    - Regulatory perspective is typically in favor of a treatment policy estimand which assumes that randomization will take its course. However, when dealing with small

effect sizes, this has the potential to for making a type 2 error and determining that a drug is not as effective as it may actually be, due to confounding factors.

C: In the case of HCV before DAA's when patients were treated with interferon and ribavirin, clinical trials were more successful at curing the disease than what was seen in the real-world data. Once DAA's were used for treatment, the clinical trial and real-world data became closely aligned. Though HCV and NASH are very different, will this issue of estimands in NASH go away once very efficacious drugs become available?

- The estimands will evolve based on what the treatments are, what the objective of the trial is, and how trials are being evaluated.

C: The FDA composite endpoint is all cause mortality- this would be an endpoint and not an intercurrent event.

C: Estimands are not an analytical method, but instead defining what is clinically relevant. Consider instead what you would want to tell the patients being prescribed the treatment- what can be expected, on average.

C: Not everything can be adjusted for in a trial- need to consider what is relevant for the context of use, and if it is possible to estimate the event. Not everything that is interesting can be estimated and adjusted for.

Q: Implementing estimands introduces yet another variable into the mix, and may or may not be exactly the same for each trial. How could someone then try to compare the efficacy of different drugs, run in different programs?

- Issues with this type of implementation have been seen in the diabetes world. Some of the drugs that have been more recently approved had to apply estimands as their primary endpoints, whereas drugs in the same class previously did not. Those numbers are therefore clearly different and not clearly comparable. How would we deal with that and do meta-analysis of this data?
- The goal of ICH E9 addendum was to move to a more generalized framework, because per-protocol populations have many biases that are associated with them. The goal was to have a lot of the principles built into the estimand. One challenge will be that it there is not a one size fits all approach and companies may do things differently. However, if we have an understanding of what are the things that we should be collecting across trials, we at least can begin collecting that data and may have the ability in the future, through data sharing, to be able to do cross comparative analyses.
    - Should this be done only if the drug fails be a small margin and you are trying to understand what happened?
    - Estimands definitely have value in that regard. Alternatively, it is still important from a comparative effectiveness perspective, to understand under different settings, different assumptions, different estimands, how the effect is interpreted based on the questions asked and your underlying assumptions.
    - You may have one result for your primary estimand, then do a sensitivity analysis and your results are not overly robust, that would bring into question the overall robustness of your effect size.

C: The concept of intercurrent events was not something invented by the estimands- they've always been there. If you do the trial planning without considering estimands, you will still have an estimand, it will just be an unconscious estimand based on some assumptions that you were not completely aware of. The e9 addendum tries to bring the message across that you need to think about estimands when planning a trial. Regulators are very keen on evaluating what comes out of the trial and may ask for different estimands to check the robustness of the results.

- Using an estimand framework is really not doing anything differently; it's just doing it prospectively, instead of retrospectively. The great opportunity the Liver Forum provides is to come to some agreement on the major types of estimands that we should be applying across trials. This would allow us to better look at these different numbers, even though there's still no direct comparisons, and have some common ground on the major ones like missing

biopsies, for example. That will be common across all studies, and can move the field forward.

C: The approach for evaluating NASH activity as a continuum vs a dichotomous approach (NAFL vs NASH) and the comparison to historical example of chronic active vs. chronic persistent is relevant to NASH. When the NASH CRN criteria was developed 20 years ago, the experience from clinical trials did not yet exist. As with viral hepatitis, stages were defined and scoring system developed, and was adjusted as time progressed and experience was gained.

- When reviewing the literature on inflammation, hardly anyone 'resolves' to zero- is a patient staying at inflammation of 1 considered resolution of NASH? Should we be looking at potentially qualitative changes in inflammation, where a patient goes from injury to healing? And should we be enrolling patients with higher inflammation?
    - One of the challenges is that the time course of resolution of the individual features of steatohepatitis are not locked in step. Fat is the fastest thing to move, and we know it can change within a couple of weeks. Inflammation is there to mop up dying cells, so if based on the mechanism of action, you've got certain cells that have now been killed, you will have some residual inflammation, which is actually a good thing.
        - There is good inflammation and bad inflammation, many biopsies show a regression in ballooning but have inflammation still, or even increase in inflammation, which can be interpreted as clearing up the dead cells.
        - Inflammation based on what can be seen on histology is probably not a good marker of regression.
    - We have to think in terms of the biology of what is going on- if the fat and the ballooning get better, that is much more important. You also don't want a massive flare of inflammation, but we don't really see that.
    - Inflammation has the poorest Kappa and the score can fluctuate between a 2 and a 3. It is hard to understand how much of that is just the artifact of how we measure it, versus how much of that is just real biology.
    - Having straightforward measures of fat, inflammation, ballooning, but with maybe some weightage to the ballooning, probably makes sense.
- The idea of having a spectrum defined by the severity of activity makes some sense because there are many cases that cannot be classified and patients progress and regress in their disease. The problem is when discussing with other stakeholders, like payers, not narrowing down the population to a much smaller number may worry them. We have to be careful also to take this into account.
    - If you have a drug with a systemic effect, such as one that increases weigh loss and improves metabolic milieu, there is a reasonable possibility that all end organs will benefit, and then it could make sense to do a study in a broader population with a broader set of endpoints. The objectives are completely different than with a drug that has a liver-specific MOA.

C: Inter-observer variability of histology reading is a problem- what is most important for clinical trials, considering the use of central pathologists, is the intra-observer variability. The intra-observer variability for each feature is much better than the inter-observer variability, and these results should also be noted.

- Large trials that are/have been enrolling have had screen failures in the range of 3:1. There is a lot to learn from these biopsies about progression, and the number of biopsies we see fatty liver disease and fibrosis but no ballooning. As well as the discrepancy between the central reader and the local pathologist with determination of NASH vs no-NASH. Collectively sharing this information will help us to learn more about pathology and how it should really be defined.
    - One of the things that is really missing is gathering sufficient data on the patients who do not meet the trial criteria. Due to cost and consent issues, many companies are not collecting data on these patients. This is a missed opportunity and the Liver

Forum can perhaps help address this to pool data on these patients across all the studies.
    - At a future Liver Forum meeting, it would be a good idea to talk about and work on how to create a registry for patients that fail trial screening, so that prospective data can be generated and begin to create the natural history cohort that everyone will one day need for studies.

C: There is no difference in nature between the resolution of steatohepatitis, and improvement of activity, given that the components of activity are part of the definition of steatohepatitis. The only difference is in the extent of the improvement. When you want to have a surrogate marker, one that is intended to become a good surrogate of clinical outcomes, you need to define it in a very stringent way.

C: With these large studies with placebo arms that go on for years, it seems that there could be some important aspects here that could help with the estimands as far as what drives treatment effect and placebo. This could be one way to collaborate within the Liver Forum, to actually pull all the data from the placebo arms and actually use that to define better estimands.

- The shared placebo concept makes a lot of sense- this concept exists for other disease areas within Transcelerate, who runs a placebo standard of care data pool. Companies have the ability to request placebo standard of care data from the individual trials, to test hypotheses such as these. This could be an opportunity for us to be able to look at some of these things, and collaborate with the Liver Forum to test some of these hypotheses.

C: Patients often decide to participate in clinical trials because of their relationship with their physician who they trust and has recommended a trial or research. Often patients still have questions about their disease, and why they are supposed to do certain things, such as changing diet and losing weight. More guidance is needed from the researchers, including explanations of 'why'.

- Physicians and study coordinators are the ones that really get patients enrolled in trials, as they have been developing a relationship with the patient and trust. Input from patients is a very important contribution to the development of study protocols.
- Patient understanding of the disease is very important – we know that the disease is a continuum, and those patients that are F0, F1 are not considered unmet medical need, but some will progress in their disease without any intervention. The field should partner with patient organizations to bolster understanding of the disease in this population which could prevent progression down the road.
- We need a holistic approach to treat not only the liver, but everything else that can be worsened by the liver condition or that coexists with the liver condition because of an underlying etiology. That does not mean that hepatologists have to tackle all of this alone, but to work with cardiologists, diabetologists, endocrinologists, etc. For clinical trials we need to design measures that are simple enough and focused enough so that we can easily understand an effect on one particular organ.
    - Though it is good to have support from other specialty areas, patients do not want to see multiple physicians, and take multiple medications prescribed by each specialist to address their medical needs.

# SESSION III: EUROPEAN REGULATORY UPDATES AND PERSPECTIVE

## Review of EMA Reflection Paper and Stakeholder Meeting

**Presenter:** Mark Ainsworth, Danish Medicines Agency

**Slides:** https://bit.ly/33SCnS9

EMA Reflection Paper Status and Contents

- The reflection paper is currently a draft. It has been released for consultation, after which it will be revised by the drafting group and sent for formal adoption by the CHMP.
- General considerations
    - NASH is a slow progressing disease, which makes it difficult to design studies that evaluate hard outcomes that are directly clinically relevant to the patient
    - Development must rely on interim evaluations with intermediate endpoints and confirmation of clinical outcomes later on.
        - Conditional marketing authorization requires the demonstration of an unmet medical need, and a positive benefit-risk evaluation of the available data.
        - There must be a balance between allowing drugs on the market as soon as possible for the benefit of patients, but also balancing that against allowing drugs on the market that do not have the desired effect
    - Histology is still considered the gold standard, though the limitations are recognized, and is required for inclusion and evaluation of efficacy in later stages of drug development. Regulators are aware that this is not an ideal tool, and the development of non-invasive methods and surrogate endpoints is encouraged.
- Overall Patient Population
    - In general, patients enrolled in trials should be representative of the target population and well balanced for demographics and concomitant disease.
    - Patients should be screened for features of metabolic syndrome, as well as other liver diseases that may be exclusion criteria.
- Non-Cirrhotic Population
    - Patients with fibrosis stage 1 should be excluded, as these patients have minimal risk for progression to end-stage disease.
    - Patients with fibrosis stages 2-3 should be included, and should have either NAS>5, or NAS≥4 with the individual components of lobular inflammation and ballooning ≥1.
    - The long-term endpoint for non-cirrhotic populations is a composite of histological diagnosis of cirrhosis, MELD >14, decompensation events, liver transplantation, and death.
    - The intermediate endpoint is a co-primary evaluation of resolution of NASH without worsening of fibrosis, AND improvement of fibrosis without worsening of NASH
        - The co-primary evaluation is proposed due to the uncertainties associated with this interim strategy. A certain level of certainty is needed before conditional marketing authorization can be granted.
- Cirrhotic Population
    - Inclusion can be based on historical biopsies with NASH, or a high likelihood of NASH based on biomarkers and comorbidity (such as T2DM or obestity)

  - Non-invasive inclusion criteria can be used in early stage trials, with certain conditions
  - Patients included should have undergone relevant lifestyle interventions and comorbidities should be treated with stable doses at inclusion. Comorbidities may also be considered for use as stratification factors
  - Clinical endpoints of all cause death and liver decompensation events can be used
    - Can be used as an "all comer" strategy without an intermediate endpoint
    - Could also be a late stage endpoint in a strategy with an intermediate endpoint, followed by hard outcomes
  - If the intermediate endpoint strategy is used, histological reversal of fibrosis from stage 4 to stage 3 or less is possible
    - Data supporting that reversal of cirrhosis is associated with reduction in risk is currently lacking
  - In patients with late stage cirrhosis, lowering of MELD score and the lowering of HVPG may be possible.
- Mode of Action
  - Usual MOA includes reduction of fat toxicity, reduction of inflammation, and resulting in a subsequent reduction of fibrosis
  - Other treatment targets may be possible
  - If the molecular target is fibrosis, "resolution of NASH" may not be possible. It could be possible to use an intermediate endpoint strategy with increased level of evidence of success- for example regression of 2 stages of fibrosis without worsening of NASH (instead of 1 stage).
- Trial Duration
  - A two-year study period is recommended for intermediate evaluation.
  - The trial duration can be shorter or longer with justification such as the trial size, magnitude of effect, patient characteristics, MOA.
  - Studies should be long enough to demonstrate an effect.
- Target of Estimation (Estimand)
  - Adopting an estimand approach aligned with the recent addendum to the ICH E9 is encouraged to be done in a prospective fashion when considering the planning, design, conduct, analysis, and interpretation of data.
  - Sponsors must carefully consider what is to be considered an intercurrent event and how they will be handled.
    - Treatment discontinuation is a particularly important with long-term studies
    - Use of additional medication is also very relevant with the patient population
  - Intermediate endpoint evaluation: treatment policy strategy
    - Clear justification for choice of estimands needs to be provided
    - All outcomes of interest should be collected independently from the occurrence of an intercurrent event
    - Align statistical analysis with regard to imputation of missing values
  - Hard endpoint evaluation: treatment policy strategy
    - Important due to expectation that censoring of patients due to incomplete follow up may have a different prognosis.
    - The refusal of biopsy will need to be considered
- Combination Treatment
  - Caution is necessary since there are no approved treatments available

  - EMA guideline on clinical development of fixed combination medicinal products[11] is referenced, which states that combination treatment should be based on:
    - Valid therapeutic principles
    - Demonstration of the contribution of each of the combination partners
  - Before a combination is explored and investigated, the individual substances should have been fully investigated.
  - Patients that could be candidates for combination treatment include patients that have a high risk of progression, or have an insufficient response to mono-therapy
- Safety
  - Safety evaluation is challenging as the parameters that are used for evaluation of efficacy could also be the same parameters that are relevant for safety. For example an increase of ALT could be due to drug toxicity or could be due to insufficient effect of the drug.
    - It is important to identify true cases of drug induced liver injury (DILI) and use the instruments that are available for differentiating between lack of effect and DILI.
    - Rules for safety evaluation should be established, including stopping rules, thresholds for clinically relevant events, and use of experimental safety biomarkers.
  - NASH patients often have co-morbid conditions, and have a high risk of cardiovascular events. The EMA guideline on the assessment of cardiovascular safety profile for medicinal products[12] should be considered.
  - Further research into the natural history of NASH patients is encouraged.
  - Parameters related to cardiovascular risks such as lipid profile, glucose homeostasis, inflammatory parameters, occurrence of MACE should be evaluated.
- Children
  - NASH is a relevant health problem in the pediatric population, though there are challenges with drug development, and drug regulation in this population.
    - Ethical issues related to the use of a long-term placebo, as well as repeated biopsies.
    - Lack of information about how NASH behaves in children and how to interpret different histological patterns.
  - The adequate age range has yet to be determined – it is unclear if populations under 10 years old would be appropriate to include in a study.
  - The amount of data needed from adult trials to be able to start a trial in a pediatric population not yet known. Currently, the recommendation is to wait until long term outcome data becomes available.
  - Pharmacokinetic data must be collected in order to select the appropriate dose and to extrapolate safety data from adult patients.
  - Conducting clinical studies with a histological endpoint may be necessary; however, there are ethical concerns to consider, and possible studies may be restricted to older age groups and/or patients with more advanced disease.

EMA Stakeholder Meeting

[11] https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-development-fixed-combination-medicinal-products-revision-2_en.pdf
[12] https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-assessment-cardiovascular-safety-profile-medicinal-products_en.pdf

- The draft of the reflection paper was discussed at the EMA stakeholder meeting in London on December 3, 2018[13],[14]
- One of the primary issues of discussion was the co-primary endpoint of resolution of NASH without worsening of NASH, and improvement of fibrosis without worsening of NASH
    - The "and" is a key difference from the FDA guidance, which states the endpoint can be either of the two options. However, the FDA guidance does also allow an option for both endpoints.
        - Participants noted that one of the components alone may be important enough, and particularly from the patient perspective.
        - Others noted that the 'and' is not achievable for certain modes of action, i.e. anti-fibrotic drugs.
            - The proposal for considering a 2-stage improvement for fibrosis was considered to be too strict of a requirement
            - Conversely, others argued in favor of a 2-stage improvement, noting that a 1-stage improvement of fibrosis may occur with variability of histological evaluation, and the clinical relevance of 1-stage improvement has not been defined.
                - Improvement of 1-stage has also been seen in the placebo group.
            - Improvement of NASH without improvement of fibrosis may not be relevant to improve the prognosis of the patient.
    - The importance of consistency or alignment between agencies was noted as a serious issue
    - Concern raised that potentially valuable treatments may be missed, or that achieving both endpoints may take too long, and further delay time to effective treatment

## Discussion

Q: At what point in clinical development is a hepatic impairment PK study typically needed?

- If the target population is suffering from hepatic impairment, this data should be available relatively early in the drug development process- at phase 2 or at least before phase 3.
- Will depend on the metabolism absorption and excretion of the compound.

Q: What is the rationale behind the 2-year duration for intermediate endpoints?

- It is a balance between allowing sufficient time to observe the effect of the drug, and not to require too long of a duration. The reflection paper is written to provide flexibility with this point, and if a drug can demonstrate effect at an earlier time point, the duration can be discussed further.
- The 2-year duration is flexible and has been put in based on the proposals of clinical studies and the results of the phase 2 trials. The 1-year results from phase 2 trials were too narrow to evaluate, and to see the full effect, a longer treatment duration may be advisable.
    - Additional clarifying text may be helpful for readers of the reflection paper to understand that the duration has flexibility, and should consider factors including mechanism of action, onset speed of drug, population, etc.

[13] EMA Stakeholder Meeting Recording: https://www.youtube.com/watch?v=TAYO2gOEwBU&list=PL7K5dNgKnawa9_BzHhuWZIxIbejBBnqMT&index=3

[14] https://www.ema.europa.eu/en/events/european-medicines-agency-stakeholder-interaction-development-medicinal-products-chronic-non

- The draft FDA guidance recommends 12-18 months, there is not a firm or absolute requirement of how long a trial should be- it depends on the population, the mechanism, and the speed. The trial duration needs to be based on non-clinical and early clinical data.
- 2 years is a long time before having any readout on the efficacy of the drug; however, if the two endpoints are combined, 2-years may be needed to see the effect on both endpoints and avoid missing drugs that have a positive effect.
- Conditional marketing authorization will be based on the interim results and the ability to demonstrate a positive benefit-risk assessment during the specified duration- the trade-offs associated with this will need to be considered by the sponsors when setting the duration.

Q: The reflection paper notes that combination therapy should be considered for patients at high risk of progression - what would be the criteria for defining patients at high risk for progression?

- Combination treatment is accepted on the prerequisite that the individual components have demonstrated efficacy. Combination treatment is not usually recommended as a starting point for treatment. For certain diseases, combination has proven to be efficacious, with a reduced risk of side effects, but those disease areas have more experience than this area.
- The reflection paper provided two cases where combination treatment may be adequate from the start: patient population at high-risk of progression, and patient population that is insufficiently treated with an established or partly established medication.
    - This part of the guidance referring to combination treatment was developed in similarity to neighboring diseases, namely hypertension and type 2 diabetes.
        - For example with T2DM, the historical paradigm has been to prescribe Metformin, and if not sufficient, add on another therapy. It is uncertain whether this same paradigm will be the case for NASH.
        - Another more recent treatment development from T2DM is to hit hard at the disease from the start – this is a developing approach and is a moving target.
    - Need to show that each component of a combination treatment does some part of the work to address the underlying disease.
    - If combination treatment is the first line of treatment, it will be difficult to evaluate the safety of each of the drugs. Data on the safety and efficacy of each drug being used in a combination must be obtained first.
- The reflection paper notes that each component of a combination must be fully investigated- does that mean a phase 3 trial must be completed for each component before being allowed to put them together? What about a situation of a combination using a drug that is already on the market for different an indication?
    - The reflection paper states that the components can be evaluated one after the other, or at the same time. If a combination trial in phase 3 is done, the individual components would need to be included in any case.
    - It is questionable whether phase 2 data would be sufficient to demonstrate the added effect of putting two components together in comparison to the single components.
    - The expectation would likely be for a phase 3 trial to include the combination in addition to each individual component and to demonstrate that the combination performs better than each of the individual components alone.

Q: One of the factors required for accelerated approval is the existence of an unmet medical need- at what point would the authorities consider the medical need of NASH to be met, and the accelerated approval or conditional approval pathway no longer available? For example once the first drug is approved, is this pathway still available?

- Even when a drug is approved, there will still be subpopulations and special circumstances where there would still be an unmet medical need

- It is unlikely that one drug will address all the medical needs of this population- need to take into account efficacy, safety, and patient acceptability of the drug.
- A drug that is approved under accelerated approval is not considered to have 'met' an unmet medical need until it has full approval. Generally, it takes multiple drugs, and drugs targeting the different aspects of disease, before the unmet medical need would be considered to be met.

C: Considering the endpoint differences in terms of how they are defined in the U.S. and Europe and looking at the data that has come from phase 3 trials, it's worth considering if we should define responders at the patient level, instead of at the population level.

C: The current draft of the EMA reflection paper requests a co-primary evaluation at the population level of the two components, not a composite at the individual level. A composite at the individual level is of clinical interest and may lend support, but not what is requested in the reflection paper.

- With the co-primary evaluation, the response to both endpoints must individually be statistically significant.
- The reflection paper states that if due to reasons relating to mechanism of action, there are problems in achieving both endpoints, there is flexibility allowed. For example, if there is an anti-fibrotic that only fulfills the criteria for improvement of fibrosis, additional supportive evidence should be presented to strengthen the argument.
    - It is unclear what additional supportive evidence could be provided. The proposal suggests a 2-stage improvement in fibrosis may be relevant.
    - Similarly, this applies for anti-inflammatory or metabolic compounds that would only meet the resolution of NASH. It is unclear what the additional evidence should be, but developers should consider additional evidence that supports efficacy if only one component can be addressed.
    - There is no expectation that a truly potent anti-fibrotic would improve steatohepatitis, and there is concern about the impact the co-primary endpoint might have on drug development of anti-fibrotic compounds. Is there a different endpoint that could be constructed for these types of drugs?
        - The co-primary endpoint also pushes development into combination therapy because that would be the only way for many compounds to achieve both endpoints.
        - A better paradigm for NASH than diabetes may be oncology, where there is almost no monotherapy at all but they are starting with combination therapy.
        - The reflection paper would benefit from addressing combination therapy and how we could potentially do platform trials in NASH to be able to assess all these aspects.

Q: Could a factorial study in which the monotherapies are tested individually, and in combination, be done in phase 2b and then advance the combination to phase 3? Or is the expectation that the factorial design would progress into phase 3?

- The FDA has a draft guidance that looks at multiple ways to tackle problems relating to combination trials, including the infeasibility of doing a combination factorial design in these very large, long studies.
- If the proposed combination includes a drug that is already on the market, than a lower bar may be reasonable, and showing liver endpoints in a phase 2b trial may be enough.
- For a new drug with little information known about its efficacy or safety, a factorial design in phase 3 may be required.

Q: The reflection paper states that patients should have failed a trial of weight reduction before enrolling into a drug study. How would this be operationalized? Does the same recommendation apply to diabetes studies or other diseases where weight reduction improves the condition?

- This point has been misunderstood and the wording should be adjusted to be more clear.
- The intent of the statement in the reflection paper was that patients should have had an attempt to lose weight at some point in their history. This is usually checked at inclusion in a trial by asking the patient a simple question.
- This recommendation is also included in the European Obesity Guideline for Clinical Trials in Drug Development and was suggested by that group to include a similar statement in the reflection paper.
- The statement is not requesting a weight-loss trial before someone is included into a NASH trial; rather, it is recommending that at enrollment it should be checked whether a patient has attempted to improve their weight.
    - If a patient has never attempted to lose weight, then the recommendation should probably be that the patient attempt to lose weight and try to make lifestyle modifications. If after 3 months this has not worked, the patient could return and be included in the trial.
    - The point is that the patients entering NASH trials have tried lifestyle changes and still have NASH.
- Alternatively, it is generally not recommended to do any intervention 3-months before enrolling someone in a NASH trial because you want patients to enter the trial at a stable weight.

Q: What is the rationale behind waiting until clinical outcome trials are completed in adults to do pediatric studies?

- The rationale is safety and efficacy because there are many other variables to consider regarding pediatric patients. Body weight is more variable as a natural effect of growing- for example children may maintain their weight but still become leaner because they have grown taller. It is unknown how the histology will behave in relation to these changes in body composition, or how drug clearance may be impacted.
- It would not be appropriate to do a phase 3 trial with histology for children until there are clinical benefit outcomes are known in adults.
    - However, there are older children who are developing significant advanced fibrosis and there are young adults that are getting liver transplants for NASH, so the field must begin to move forward in a conscious way to address the needs of this population.
        - FDA has allowed some PK studies in children to begin that do not require liver biopsies, and use non-invasive endpoints.
        - EMA reflection paper is referring to the phase 3 data that needs to be generated in children, and PK studies or proof-of-concept studies could be possible in this population at earlier stages.

Q: In NASH cirrhosis, would histologic improvement of cirrhosis be acceptable for accelerated approval?

- The EMA reflection paper notes improvement of cirrhosis as a possible endpoint; however, also notes the uncertainty and lack of data on the impact of this improvement on prognosis.
- The FDA draft guidance for NASH cirrhosis populations acknowledges this is a complicated issue. Improving fibrosis to a pre-cirrhotic stage would need to be supported by clinical evidence that the function of the liver was improving.

# SESSION IV: WORKING GROUP UPDATES

## NASH Cirrhosis Working Group

**Presenter:** Arun Sanyal, Virginia Commonwealth University

**Slides:** https://bit.ly/33TSrTB

- The NASH Cirrhosis working group has three arms: case definitions; risk stratification; endpoints. The focus so far and the most progress has been made on case definitions, attribution of NASH as the etiology of cirrhosis, and risk stratification.
- The definition of cirrhosis for NASH trials was published previously from the Liver Forum, and now the focus is turned to how to attribute the cirrhosis to NASH.
- The working group developed criteria into 3 categories: definite NASH, probable NASH and possible NASH.
    - Definite NASH
        - Definite Criteria A
            - Histological evidence of cirrhosis: extensive architectural disruption, development of fibrosis with circumscription of regenerative nodules
            - Features of steatohepatitis: steatosis, hepatocellular ballooning, inflammation
            - Non-alcoholic nature of disease established by clinical assessment and AUDIT
        - Definite Criteria B:
            - Patients with a previous biopsy showing steatohepatitis, but who now have cirrhosis, either by clinical picture or imaging, or with biopsy
                - If biopsy, the biopsy is not showing clear cut features of fatty liver disease
            - One or more risk factors for fatty liver disease at the time of entry
        - Definite Criteria C:
            - Patients with current biopsy showing cirrhosis and steatosis only.
            - Two or more features of metabolic syndrome – one of which should be type 2 diabetes
    - Probable NASH
        - Probable Criteria A:
            - Patients with a previous biopsy showing steatosis, but now with cirrhosis, either by a clinical picture, imaging, or biopsy
            - Two or more features of metabolic syndrome
        - Probable Criteria B:
            - Patients with current or previous imaging showing steatosis
            - Two or more features of metabolic syndrome
        - Probable Criteria C:
            - Patients with “cryptogenic cirrhosis” without current or previous evidence of steatosis by imaging or steatosis/steatohepatitis by histology.
            - Two or more features of metabolic syndrome
    - Possible NASH
        - Possible Criteria A:
            - Cryptogenic cirrhosis patients without current or previous evidence of steatosis by imaging or steatosis/steatohepatitis by histology

        - Two or more features of metabolic syndrome
    - Possible Criteria B:
        - Patients with previously eradicated hepatitis C virus, or a remote history of heavy alcohol consumption, but who currently have evidence of cirrhosis and histological evidence of steatohepatitis.
            - Patients with a remote history of heavy alcohol consumption should have evidence they did not have cirrhosis at the time of stopping alcohol.

## Standard of Care: Comorbidity MGMT Working Group

**Presenter:** Raluca Pais, Hôpital Pitié Salpêtrière

**Slides:** https://bit.ly/2SOi4PA

- The aim of the working group is to develop a paper which provides simple recommendations on how to optimize the comorbidities in patients that are entering NASH clinical trials
    - The paper does not aim to substitute guidelines or recommendations that already exist for the treatment of the comorbid medical conditions, but rather to find the best way to fit these recommendations into the context NASH clinical trials.
- Introduction and selection of comorbidities
    - Decide to focus on the comorbidities which are the most prevalent and where treatment might have an impact on NAFLD/NASH histology or clinical outcomes: weight, type 2 diabetes, dyslipidemia, and hypertension.
    - Other comorbidities will be mentioned in the paper but not focused on, such as sleep apnea, renal function impairment, etc.
    - The paper will cover how to deal with these comorbidities before starting the trial, and what should be done to optimize the control of the comorbidities before and during the clinical trial.
- Term definitions:
    - Unstable medical condition: related to a condition that impacts the life expectancy of the patient (acute or chronic).
    - Uncontrolled comorbidities: may impact the course of liver disease, require hospitalization, discontinuation of study drug, or use of forbidden medication.
    - Optimal control of comorbidities: will not have perfect control, but should attain control within accepted boundaries for each condition.
        - Consider the best strategies to optimize patients without delaying the trial, and without a significant interaction with the drug tested.
    - Metabolically stable: no change or minimal changes in comorbidities and their treatment.
        - Some comorbidities such as type 2 diabetes and changes in weight can have an impact on the primary study outcomes.
        - Other comorbidities such as hypertension, or dyslipidemia, are more likely to impact secondary outcomes and safety evaluation.
        - The time requirement for metabolic stability is still under discussion, and whether this would vary according to stage of trial.
- Type 2 diabetes
    - Newly diagnosed type 2 diabetes (identified at screening):
        - Start treatment before inclusion in the trial – lifestyle recommendation, medical treatment according to local/national guideline.
            - When should liver biopsy be done—before or after start treatment?

        - Start trial without treatment for diabetes
    - Existing/uncontrolled type 2 diabetes:
        - Adjust treatment to try to achieve glycemic target—will not substitute for diabetes guideline.
- Arterial hypertension
    - Uncontrolled blood pressure will be defined ≥160/100mmHg.
        - More a safety issue than a confounder, particularly in phase 3 trials.
    - Treatment adjusted according to local guidelines, dose escalation of existing drug, or introduction of a new drug.
        - Anti-hypertensive drugs have no established effect on liver histology
- Dyslipidemia
    - Cardiovascular risk and the indication for statin therapy should be evaluated for each patient with NASH.
    - For patients not taking statin, a statin or non-statin dysliidemia therapy should be initiated if indicated.
    - If taking a statin but not controlled- up-titrate dose as necessary.
    - Questions include the accepted time interval for liver biopsy, if statins are initiated before starting treatment, should the start of trial be delayed? If so, for how long?
    - The paper will also discuss non-statin lipid lowering therapies, and differences between the lipophilic and hydrophilic statins
- Weight control
    - Weight changes can bi-directionally impact the severity of histological lesions.
    - No specific recommendations to achieve a target weight but rather to maintain a stable weight during the trial.
    - <5% weight changes, gain or lose, during the six-months prior to liver biopsy or prior to screening based on imaging or other criteria.
        - Questions include if the same recommendation should be used for phase 2 and 3 clinical trials? And if there should be an upper limit BMI since this can impact a drug's pharmacokinetics and pharmacodynamics.
- Run-in period
    - Not currently recommended in NASH clinical trials.
    - The purpose of the run-in period would be to stabilize the condition, optimize the management, and to assess the adherence to the treatment.
    - Currently recommended in type 2 diabetes and hypertension trials.
    - If applied to NASH, would it be used in both phase 2 or 3 clinical trials?
- Alcohol consumption
    - Will provide recommendations of how to document and/or monitor alcohol consumption in clinical trials, such as Timeline Follow Back, AUDIT, etc.

## Standard of Care: Lifestyle MGMT Working Group

**Presenter:** Sven Francque, Antwerp University Hospital

- The draft of the manuscript has been finalized with the proposed title: Standardization of Diet and Exercise in Clinical Trials of NAFLD-NASH: Recommendations from the Liver Forum
- The rationale behind the paper is to find the balance between the need to have some standardization on lifestyle management within the context of a clinical trial but not introducing an intervention within an intervention
    - Realistic balance between what is ideal or what you might like to do, and what is feasible in the context of a clinical trial—practical and logistical considerations, burden to the research team and patients, and cost for the sponsors

- The paper summarizes the data on how diet and exercise influences NAFLD-NASH and the influence it has on liver histology. For some patients, there will be a clear response and in others it is less pronounced but there is some effect that must be considered along with the possible contribution to placebo effect.
- The Working Group looked at historical and ongoing trials to assess what has been done and what is reported on lifestyle management within those trials, and came up with 3 categories of trials:
    - Details provided on nutritional counseling and exercise recommendations that were provided to participations, which are included in the protocol and reported on
    - Mention of some form of counseling but no further details provided
    - No mention of lifestyle management or counseling at all
        - A table of the different trials and categories is included in the paper to highlight the heterogeneity and lack of details in trials to date.
- Discussions of the paper brought up questions such as: how can lifestyle parameters be captured? What tools are available for the different components of lifestyle?
- Many patients have features of the metabolic syndrome, therefore should use examples of how trials in the fields of diabetes and obesity are conducted, particularly how they deal with the lifestyle component.
- The working group has robust discussions on whether there should be a lead-in period or period of stabilization, and reviewed other fields to find out that the recommendation of a lead in period for obesity was dropped in recent guidelines.
- Considerations for the recommendations:
    - From an ethical point of view, some lifestyle counseling must be provided to patients
    - Within a trial, lifestyle must be standardized as much as possible because changes within the context of a trial can have an impact on the results being measured
    - The group emphasized the importance of describing what is being done in the trial, standardizing it, suggesting a minimum of data which captures lifestyle factors of patients.
    - There are cultural and regional differences in dietary habits that must be considered and cannot be one-size fits all, especially for global trials.
    - Lifestyle recommendations provided need to be recorded in the protocol and in the reports of the trial.
        - Should also capture basic measurements on the patient so that changes can be analyzed and the impact on the trial can be understood.
- The paper has gone through several rounds within the working group and is currently being formatted for submission.

## Pediatric Issues Working Group

**Presenter:** Joel Lavine, Columbia University

**Slides:** https://bit.ly/2H2m1gY

- The Working Group's first manuscript is titled: Factors to Consider in the Development of Drugs for Pediatric Non-alcoholic Fatty Liver Disease
- Key issues to consider in pediatric NASH trials include:
    - NAFLD is the most common cause of chronic liver disease in developed countries
    - The majority of children are unaware that they have NAFLD because they usually do not go in for blood work. Are usually diagnosed after going to the doctor for something else and getting flagged for elevated liver enzymes.
    - Unlike adults, there are no confounders related to alcohol

    - The severity of histologic findings is only slightly reduced from that found in adults
    - There is a substantial subset of NAFLD that relates to histologic features unique to children that are not found in adults
    - There may be a difference in etiopathogenesis, long term outcomes, and variable responses to treatment.
    - The hepatic, cardiovascular, and endocrine adverse outcomes remain indeterminate.
- Factors that may help recognize children at risk for fatty liver include:
    - Presence of metabolic syndrome comorbidities, particularly obesity
    - Family history of fatty liver disease
    - Particular races and ethnicities
    - Older children/adolescents
    - Elevated serum aminotransferases, GGT
    - Hepatomegaly or acanthosis on physical exam
    - Evolving non-invasive imaging or serum biomarkers, less validated in pediatrics
- Regulatory considerations for pediatric trials in NASH include:
    - There are no phase 3 trials for children with NAFLD
    - FDA requires an initial pediatric study plan, and the EMA requires a pediatric investigation plan.
    - Both the EU and the U.S. have laws mandating the assessment of safe and efficacious drugs in children. However, children cannot participate in research for which there is more than minimal risk unless there is a direct benefit to them (not only societal benefit).
        - Approval must be granted by the FDA, NIH, and the IRBs at each institution
    - The Liver Forum is due to release the recommendations for pediatric NAFLD in 2019
- Treatment considerations for pediatric NAFLD/NASH trials include:
    - There have been two randomized controlled trials with histology endpoints conducted
        - Both trials successfully enrolled intended sample sizes and had complete histology after 1-2 years of treatment for over 86% of subjects.
        - All patients received lifestyle advice
        - Vitamin E improved NASH resolution significantly
    - Consideration of the unique pediatric-type subset histology in a significant proportion of subjects should be included in a priori stratification, as treatment response and clinical outcomes may differ.
    - Race/ethnicity, pubertal development, gender, and age are all important in design and analyses
    - Drug dosing, formulation, route pose hurdles need to be addressed
    - Mechanism of action ideally will address other co-morbidities
- The paper was submitted to Gastroenterology in April 2019, received ‘reject with hope’ decision in May along with request for additional components including lay summary and graphical abstract.
    - The manuscript addresses topics including epidemiology, regulatory mandates and concerns about conducting trials in children, and the current state of therapeutic development.

## Discussion

Q: Do patients with compensated cirrhosis have decreased liver function? If it is not decreased, how can liver function improve, as requested by the FDA? How would improvement in liver function be measured?

- If a patient has synthetic dysfunction, while they do not have clinical decompensation, they are well on their way to clinical decompensation and that may become a risk stratification issue
- Even if the patient is otherwise asymptomatic, it tells you that they are further along in their disease and the risk of a clinical event is higher
- Tools such as HepQuant, methacetin, breadth test, gadoxetate can measure improvement in a patient who is clinically compensated
- For patients who have histologic cirrhosis but have normal total bilirubin, no platelets, and no varices, it will be more challenging to determine evidence of improvement in function
    - For patients that have abnormal bilirubin, platelets, evidence of disease can be measured clinically by other biomarkers and non-invasive tests that may not be accepted as a single validated surrogate endpoint at this point, but the totality of the data can be considered if there appears to be clinical improvement along with histologic improvement.
    - Stratifying these populations is very important.

Q: In the first group of "definite cirrhosis", if there is a patient that had Hepatitis C 15 years ago, and a biopsy showing steatohepatitis and imaging showing cirrhosis, why should those patients be excluded?

- Clarification is needed because in the pre-cirrhotic population, patients with eradicated hepatitis are eligible.
- A critical question is after hepatitis C is cured, how long does it take before a patient will reach a steady state? There are ongoing anti-fibrotic trials in this population.
- Secondly, if a patient had hepatitis C, it must be certain that they did not have cirrhosis when the hepatitis C was cured. If that is confirmed and the patient now has cirrhosis and features of steatohepatitis, they can enroll in a trial.
    - Similarly, once an alcoholic patient stops drinking or consuming health amounts of alcohol, if they already had cirrhosis at that point, it is uncertain if their disease is purely due to NASH. So confirmation of no cirrhosis at that time is needed to increase the confidence that the cirrhosis is due to NASH and not the alcohol use.
- The three categories of NASH (definite, probable, and possible) are needed because in the early phases of clinical trials, very tight populations are needed to do a proof-of-concept and dose-ranging, and look at biomarkers in a smaller group of patients.
    - As drug development proceeds, a broader category of patients is needed because the population should resemble the patients who are going to get the drug when it goes on the market.
    - In phase 4 patients that have alcoholic hepatitis or hepatitis B should be included in order to obtain data in those subgroups of patients
    - Any population selected must be justified.

Q: How is the working group recommending that changes in lifestyle are measured? Tools such as Fit Bits? Taking pictures of food? Measuring calories?

- The group has developed a pyramid which includes increasing levels of quality of data that be collected.
    - At a minimum, there should at least be some questionnaires that capture data at baseline and throughout the trial.
    - Additional measures can be done including the use of Fit Bits or other tools/devices but that is up to the top of the pyramid as it is more intensive approach.
    - The paper was left open to have a general basic recommendation.
    - More data can be collected but it is always a balance between feasibility and increasing the intervention type of what is being done.

Q: How do regulators view the collection of real-life evidence on the impact of lifestyle modification?

- The FDA is very supportive of gathering real life data that supports clinical trial evidence

Q: Is there an impact of change in physical activity that is independent of changes in weight?

- There is data that indicates that change in physical activity has some impact- the consensus was that basic information on physical activity should be captured to be able to compare trials to some extent. It may also help with understanding what the final results of the treatment are. It is understood that it is not easy to capture.
- Physical exercise, independent of weight changes, may change the amount of liver fat content, but it is uncertain whether it has an impact on inflammation and fibrosis.
- It is also important to be sensitive of the study burden to the patient (i.e. asking them to fill out a lot of forms)

C: It is important to do a run-in period prior to performing a biopsy for a clinical trial- if you do a biopsy, then tell patients to follow these recommendations, then start a drug, there may be a lot of variation on who complies and who doesn't.

C: Practically speaking, there are logistical challenges of conducting large phase 3 trials with hundreds of sites all around the globe. Ethically, it is important to give diet and exercise recommendations to patients, but what this looks like in practice will vary from country to country.

- The duration of the trial is an important issue to consider- patients may drastically change their lifestyle in a short 12-week trial, vs. a longer trial it would be unlikely to maintain for a longer period of time.