UNDERSTANDING AND ADDRESSING POTENTIAL BIAS IN PATIENT-REPORTED OUTCOMES FROM CLINICAL TRIALS

ISPOR Barcelona Workshop

Tuesday 13 November

14:00-15:00

Prof. Olivier Chassany
EA 7334, Patient-Centered Outcomes Research, Université Paris-Diderot
Health Economics Clinical Trial Unit, hôpital Hôtel-Dieu, AP-HP
olivier.chassany@aphp.fr

Health authorities are asking for PRO assessment in dossiers

From rejection to recognition of PRO

FDA

Guidance for Industry[1]
Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)

December 2009
Clinical/Medical

emea
European Medicines Agency
Pre-authorisation Evaluation of Medicines for Human Use

London, 27 July 2005
Doc. Ref. EMEA/CHMP/EWP/139391/2004

REFLECTION PAPER ON THE REGULATORY GUIDANCE FOR THE USE OF HEALTH-RELATED QUALITY OF LIFE (HRQL) MEASURES IN THE EVALUATION OF MEDICINAL PRODUCTS

efsa European Food Safety Authority

European Food Safety Agency

EFSA Journal 2015;XXX:YYYY

DRAFT SCIENTIFIC OPINION

Guidance on the scientific requirements for health claims related to the gastro-intestinal tract, the immune system, and defence against pathogenic microorganisms[1]

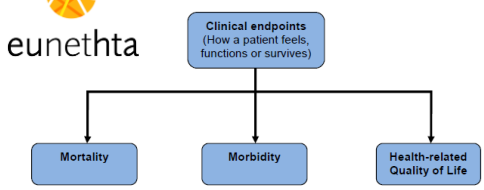
EFSA Panel on Dietetic Products, Nutrition and Allergies[2,3]

European Food Safety Authority (EFSA), Parma, Italy

EunetHTA Guideline Clinical Endpoints

European Network HTA Joint Action
WP5 - Update on Methodology guidelines
February 2013

eunethta



Requests from Health Authorities and changes in study design

HTA requests

- For more “real life” studies (especially post marketing authorization)
- With **patient perception**
- Meaning :
  - **Open cohort** study
  - Randomized **open** trial

Shift in study design

- Placebo less and less ethical
- Superiority : often unreachable goal (HIV, anticoagulants …)
- Shift for **Non-inferiority** design (“less robust”)
- Often **open** as blind is not feasible or desirable

Health authorities (especially HTA) :
Too many (potential) biases
→ non eligible study for review

Cohort
Open
Non-inferiority
PRO

Discrepancies among Agencies

On the use of PRO measures in oncology studies (EMA 2016)

Whilst the concern in relation to bias in open label studies remains, it might well be that data of **clinical interest a priori can be produced only under open label randomised** controlled trial conditions.

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man
The use of patient-reported outcome (PRO) measures in oncology studies

Draft agreed by Oncology Working Party	December 2013
Adopted by CHMP for release for consultation	22 May 2014
Start of public consultation	17 June 2014
End of consultation (deadline for comments)	30 November2014
Agreed by Oncology Working Party	November 2015
Adopted by CHMP	1 April 2016
Date for coming into effect	1 November 2016

Oncologic disease: considerations on blinding (FDA, 2018)

FDA recommends that a sponsor use a placebo-controlled design only in selected circumstances (e.g., …), or with certain trial design features (e.g. …, when the endpoint intended to support a labeling claim has a **high degree of subjectivity, such as patient-reported outcomes**).

FDA

Hematologic Malignancy and Oncologic Disease: Considerations for Use of Placebos and Blinding in Randomized Controlled Clinical Trials for Drug Product Development
Guidance for Industry

DRAFT GUIDANCE

August 2018
Clinical/Medical

Schizophrenia Research 168 (2015) 498–504

Contents lists available at ScienceDirect

Schizophrenia Research

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

Qualify: a randomized head-to-head study of aripiprazole once-monthly and paliperidone palmitate in the treatment of schizophrenia

Dieter Naber [a], Karina Hansen [b], Carlos Forray [c], Ross A. Baker [d], Christophe Sapin [b], Maud Beillat [b], Timothy Peters-Strickland [d], Anna-Greta Nylander [e], Peter Hertel [e], Henrik Steen Andersen [e], Anna Eramo [f], Jean-Yves Loze [g], Steven G. Potkin [h,*]

[a] Department for Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany
[b] Lundbeck SAS, Quai du Président Roosevelt 37-45, 92445 Issy-les-Moulineaux, France
[c] Lundbeck Research US, 215 College Rd, Paramus, NJ 07652, United States
[d] Otsuka Pharmaceutical Development & Commercialization, Inc., 508 Carnegie Center, Princeton, NJ 08540, United States
[e] H. Lundbeck A/S, Ottiliavej 9, 2500 Valby, Denmark
[f] Lundbeck LLC, 4 Parkway North, Deerfield, IL 60015, United States
[g] Otsuka Pharmaceutical Europe Ltd, Gallions Wexham Springs, Framewood Road, Wexham SL3 6PJ, United Kingdom
[h] Department of Psychiatry and Human Behavior, University of California, Irvine, 5251 California Ave, Irvine, CA 92617, United States

Primary objective : to compare the effectiveness of 28-week treatment with AOM 400 to PP (both long-acting injectable antipsychotics) in adult patients with schizophrenia on the Heinrichs-Carpenter Quality-of-Life Scale (QLS) as primary endpoint using a non-inferiority hypothesis.

Primary endpoint : Heinrichs–Carpenter Quality-of-Life Scale (QLS)

- **Clinician-rated scale** derived from a semi-structured patient interview widely used in psychopharmacological evaluation of treatments for schizophrenia
- QLS measures effects beyond functioning: richness of personal experience, quality of interpersonal relations, productivity in occupational roles
- 21 items in 4 domains:
  - Interpersonal Relations (8 items)
  - Instrumental Role (4 items)
  - Intrapsychic Foundations (7 items)
  - Common Objects and Activities (2 items)
- **Primary analysis: QLS total score change from baseline to week 28**

(Lewis et al., 2006) (Heinrichs et al., 1984)

Justification of the NI margin ?

Potential bias	Risk to set a large NI that could lead to demonstrate falsely NI
Margin ?	**5-point difference on the QLS total score**
Justification	5-6 points represent the **MCID** and is a clinically relevant difference in the evaluation of antipsychotic drug efficacy based on previous trials, i.e. between 1st & 2nd-generation antipsychotics and between aripiprazole and SOC. **Usual to set the NI margin as the half of the difference observed in previous studies** between the comparator and placebo. • But no study of paliperidone palmitate versus placebo based on the QLS questionnaire. • Meta-analysis of 6 comparative trials of olanzapine vs placebo: mean difference of the total QLS score was 10 points, which reinforces the 5-point as NI margin

Jones et al, 2006; Taylor et al, 2008; Dunayevich et al, 2006

Sample size

Potential bias/risk	Risk to i too muc enough
Sample size	**286 = 220** attrition ra of previou
Calculation	- 5-point N - SD 15 - α 0.05 - Power 80% - (Hypothesized treatment difference of 1)

sealed envelope™ ★ THE ORIGINAL INTERNET AND TELEPHONE

POWER (SAMPLE SIZE) CALCULATORS

Significance level (alpha)	5%
Power (1-beta)	80%
Standard deviation of outcome	15
Non-inferiority limit, d	5

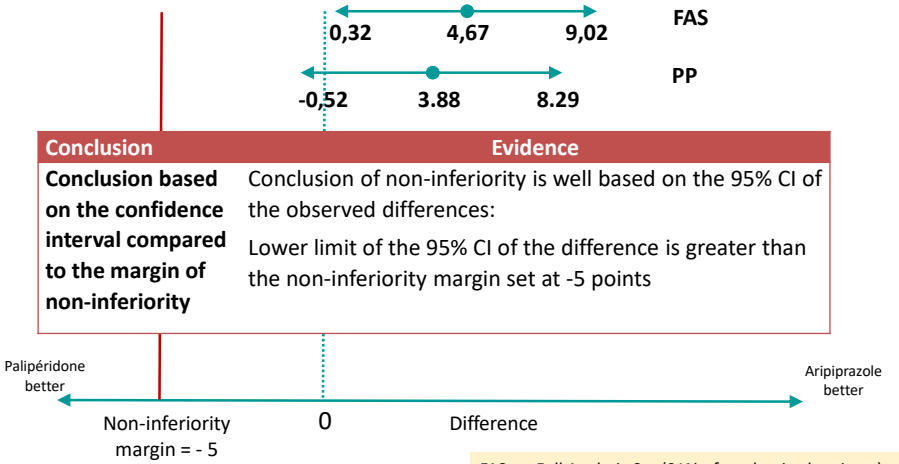
Calculate sample size

Sample size required per group	112
Total sample size required	224

Addressing the bias of lack of internal validity

Internal validity	Evidence
1. Prior demonstration of the efficacy of the treatment under study	Aripiprazole (AOM) has demonstrated clinical efficacy and has Marketing Authorization (MA)
2. Prior demonstration of the efficacy of comparator	Paliperidone palmitate (PP) has been shown to be effective. It is the most commonly used atypical antipsychotic drug in most European countries
3. Experimental conditions similar to previous trials of comparator efficacy demonstration	Eligibility criteria are similar to previous trials
4. Appropriate dosage and conditions of administration of treatments (especially comparator)	**At the 24th week of treatment (at last injection, dosage was 387±34 mg for AOM 400 and 110±3.6 mg for PP) in line with MA**
5. Confidence in the quality of the monitoring of the trial (difficult to check by reading the publication)	~ 30% of the patients did not complete the 28 weeks: 29.7% (AOM), 36.7% (PP) consistent with previous trials 7 patients were lost to follow-up (2 AOM, 5 PP)

Demonstration of non-inferiority consistency of both ITT & PP analyses based on IC ?



Conclusion	Evidence
Conclusion based on the confidence interval compared to the margin of non-inferiority	Conclusion of non-inferiority is well based on the 95% CI of the observed differences: Lower limit of the 95% CI of the difference is greater than the non-inferiority margin set at -5 points

FAS Full Analysis Set (91% of randomized patients)
PP Per Protocol (79% of randomized patients)

Checking the quality of the study and its eligibility for review by a health Authority (HTA)

Non-Inferiority (NI) checklist	Yes	No
Justification of NI margin (predefined)	√	
Sample size based on NI margin	√	
Prior demonstration of the efficacy of comparator	√	
Experimental conditions similar to previous trials of efficacy demonstration of the comparator	√	
Appropriate dosage and conditions of administration of treatments (especially comparator)	√	
Confidence in the quality of the monitoring of the trial (difficult to check by reading the publication)	√	
Results presented in per protocol AND in Intent to treat analysis : **consistency of both analyses**	√	
Conclusion based on the **Confidence Interval (95% CI) of the difference** between treatments compared to the predefined margin of NI	√	

Internal validity

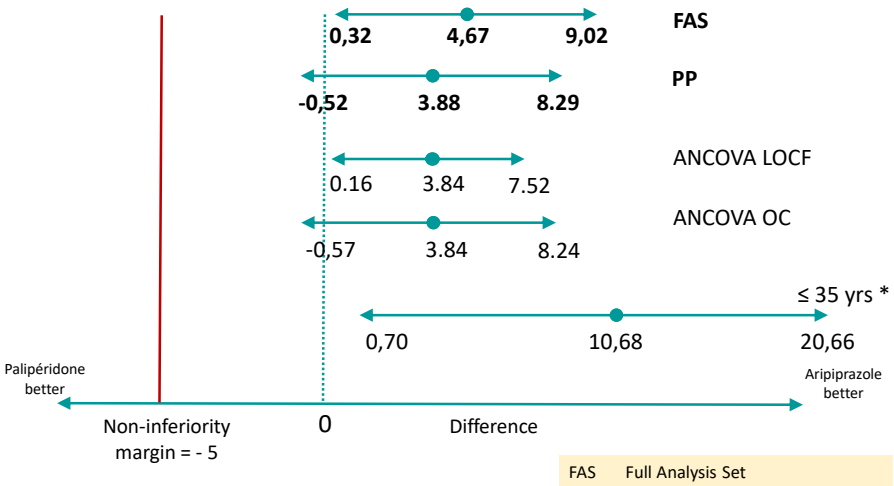
Piaggio G, Elbourne DR, Pocock SJ, Evans SJW, Altman DG, for the CONSORT Group. Reporting of noninferiority and equivalence randomized trials. Extension of the CONSORT 2010 statement. JAMA. 2012; 308(24): 2594-2604.

Addressing the bias of lack of blind

Potential bias	Lack of double-blind: subjectivity of assessment
Justification	The study setting was close to "real life" while keeping high methodological quality Blind was not desirable in this context, as the objective was precisely to the capture the patient perception on these 2 treatments
Solution	As the primary endpoint is a Clinician-Reported Outcomes (CRO), **PROBE** has been applied (i.e. independent assessor blinded to treatment) to QLS and efficiency scale (IAQ) (secondary endpoint)

PROBE: Prospective Randomized Open, Blinded Evaluation

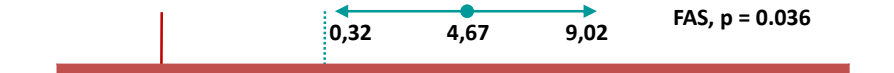
Demonstration of non-inferiority
consistency of sensitivity analyses



* Subgroup analysis prespecified by stratification

FAS	Full Analysis Set
PP	Per Protocol
LOCF	Last Observation Carried Forward
OC	Observed cases

Switching NI to superiority



Potential bias	Risk to falsely conclude to superiority
Justification	Since the lower limit of the 95% CI of the difference between aripiprazole and paliperidone is greater than 0, the difference is statistically significant in favor of aripiprazole: p = 0.036 in ITT (FAS)
Solution	**Prespecified in the protocol : if met, a predefined test of superiority would be conducted** **There is no supplementary analysis, just to present the p value of the difference between groups**
Interpretation	The difference of 4.7 points is close to the minimal clinical difference (MCID). A recent study confirms that the MCID of the QLS questionnaire is 5.3 (Falissard et al 2015).

Palipéric bette

azole er

ward

OC	Observed cases

Switching NI to superiority

Quite difficult to understand (and accept) to define a **non clinically relevant NI margin** at beginning and to conclude finally that the difference observed between groups (a little bit lower than the NI margin) is **clinically relevant** for a superiority claim…

What else could support the demonstration of efficacy ?

Demonstration of superiority
Supported by consistency across II endpoints

Population FAS (at wk 28 or change BL-wk 28)		Mean of the difference (IC95%), OR or %	P
Investigator's Assessment Questionnaire (IAQ) - relative effectiveness (efficacy, safety and tolerability) of antipsychotic medications	ClinRO PROBE	-1.49 (-2.94 ; -0.05)	P = 0.043
CGI-S Clinical Global Impression Severity scale	ClinRO	-0.28 (-0.48 ; -0.09)	P = 0.004
• Responders (%)		OR = 2.26	P = 0.01
CGI-I Impression of Improvement responders (%)	ClinRO	OR = 2.51	P = 0.0032
Work Readiness Questionnaire (WoRQ)	ClinRO	-1.16 ± 0.40 (-1.96 ; -0.37)	P = 0.004
• Patients ready to work according to clinician		20%	
• Patients not ready to work at baseline and ready at wk 28		14.2%	
Arizona Sexual Experience Scale (ASEX)	PRO	-1.162 ± 0.399	
• % with sexual dysfunction at wk 28		OR = 0.80 (0.48 ; 1.32)	
• % with sexual dysfunction at baseline and without at wk 28		9.5%	
Subjective Well-being under neuroleptic treatment (SWN-S)	PRO	1.00 (-2.40 ; 4.42)	P = 0.56
Tolerability and Quality of Life (TooL)	PRO	-0.70 (-1.51 ; 0.12)	P = 0.095

Demonstration of superiority
Supported by prespecified relevant subgroup analysis

Population FAS		Mean difference (IC95%) or OR	p
QLS total score	**(~2 MCID)**	**10.68** (0.70 ; 20.66)	P = 0.037
IAQ		-2.65 (-5.28 ; -0.02)	P = 0.048
CGI-S		-0.44 (-0.83 ; -0.06)	P = 0.026
WoRQ		-2.70 ± 0.85 (-4.41 ; -0.99)	P = 0.0026
• Patients not ready to work at baseline and ready at wk 28		OR = 2.67 (1.39 ; 5.14)	P = 0.003
ASEX : % of patients with sexual dysfunction at wk 28		OR = 0.60 (0.24 ; 1.46)	

Potential bias	Non comparability of subgroups
Justification	Important to demonstrate the effectiveness of treatments in the young population of schizophrenics, for an optimal care early in the disease and to act on the risk of desocialization
Solution	Stratification allows in case of positive result on the overall population, to perform this subgroup analysis (≤ 35 vs.> 35 years)

Conclusion

- Subjectivity of the patient is what we want to capture
- NI trial is not free of potential biases (especially lack of blind), but these can be anticipated, minimized or balanced:
  - Adequate methodology (e.g. when possible **PROBE**)
  - High quality of the follow-up
  - Clear report of analysis
  - **Interpretation of the observed difference:**
    - **Compared to MCID**
    - **Presentation as responders**
    - **Consistency across endpoints, across studies**

Open NI trial even with COA (ClinRO, PRO) is eligible for review by agencies: Regulators may use a checklist to easily check the quality of the trial

Voting question

Given the (unbiased and objective) presentation of this NI trial, what is your perception ?

a) Non-inferiority has been demonstrated

b) Superiority has been demonstrated

c) The difference between groups is clinically relevant

d) A claim in the Summary of Products Characteristics could be granted

e) Biases remain and preclude any formal conclusion

Live Content Slide

When playing as a slideshow, this slide will display live content

Poll: Given the (unbiased and objective) presentation of this NI trial, what is your perception?