Use of item-based non-linear mixed effects model to improve confidence in Phase III clinical trial decision-making

Carolina Llanos-Paez[1], Claire Ambery[2], Shuying Yang[2], Misba Beerahee[2], Elodie L. Plan[1], Mats O. Karlsson[1]

1. Department of Pharmacy, Uppsala University, Uppsala, Sweden
2. Clinical Pharmacology Modelling and Simulation, GlaxoSmithKline plc, London, UK

29th PAGE conference - 7th September 2021

Failures in clinical drug development

Phase II & III

Early termination of “poor” assets

	Reasons for failure	
	Phase II	**Phase III**
Efficacy	**48%**	**55%**
Safety	25%	14%
Strategy	21%	14%
Others	6%	17%

Harrison, R. Nat Rev Drug Discov. 2016; 15:817–818

Failures in clinical drug development

Phase II & III

Early termination of “poor” assets

	Reasons for failure	
	Phase II	**Phase III**
Efficacy	**48%**	**55%**
Safety	25%	14%
Strategy	21%	14%
Others	6%	17%

Efficacy failures:

- ✗ Insufficient knowledge about target population
- ✗ Insufficient sample size
- ✗ Large uncertainty in drug efficacy estimate
- ✗ Insufficiently powerful analysis methods

Harrison, R. Nat Rev Drug Discov. 2016; 15:817–818; Lalonde et al. Clin Pharmacol Ther. 2007;82(1):21–32 ; Erdmann et al. BMC Med Res; Methodol. 2020; 20(1):253

Chronic Obstructive Pulmonary Disease
(COPD)

- Outcomes of interest in efficacy trials:
  - ✔ Airway obstruction (Forced Expiratory Volume in 1 second – FEV_1)
  - ✔ Dyspnea (Transition Dyspnea Index - TDI)
  - ✔ Health status (St. George Respiratory Questionnaire - SGRQ)
  - ✔ Exacerbations
  - ✔ **Patient reported Outcomes (PROs):**
    - Reported directly by the patient
    - Increasingly used in drug development

How do we “measure” disease?

PRO in COPD

- **EXACT** – 14 questions related to COPD symptoms
- **E-RS:COPD** – 11 questions specifically related to respiratory symptoms

Item number	Item-level construct	Score	Symptom construct
7	Breathless today	0 - 4	
8	Breathless with activity	0 - 3	
9	Short of breath – personal care	0 - 4	**Breathlessness**
10	Short of breath – indoor activity	0 - 3	
11	Short of breath – outdoor activity	0 - 3	
2	Cough frequency	0 - 4	
3	Mucus quantity	0 - 3	**Cough and sputum**
4	Difficulty with mucus	0 - 4	
1	Congestion	0 - 4	
5	Discomfort	0 - 4	**Chest symptoms**
6	Tightness	0 - 4	
12	Tired or weak	0 - 4	
13	Sleep disturbance	0 - 4	Additional attributes
14	Scared or worried	0 - 3	

EXACT total scores uses all 14 items with logit scoring transformed to a 0 to 100 interval-level scale; E-RS:COPD scores are based on summation to yield ordinal-level scales with a total score ranging from 0 to 40

Evidera. EXACT® Program. 2020. https://www.exactproinitiative.com/content/

How are such data typically analysed?

Mixed Model Repeated Measures (MMRM)

- Total score analysis
- Non-ignorable missing data requires a sensitivity analysis
- Time is handled as discrete variable

Item number	Item-level construct	Score	Symptom construct
7	Breathless today	0 - 4	
8	Breathless with activity	0 - 3	
9	Short of breath – personal care	0 - 4	**Breathlessness**
10	Short of breath – indoor activity	0 - 3	
11	Short of breath – outdoor activity	0 - 3	
2	Cough frequency	0 - 4	
3	Mucus quantity	0 - 3	**Cough and sputum**
4	Difficulty with mucus	0 - 4	
1	Congestion	0 - 4	
5	Discomfort	0 - 4	**Chest symptoms**
6	Tightness	0 - 4	
12	Tired or weak	0 - 4	
13	Sleep disturbance	0 - 4	Additional attributes
14	Scared or worried	0 - 3	

Sum of scores

Alternative analysis

Item-based response model (IRM)

- Mathematical models expressing the **probability** of the particular response to a **scale item** as a function of an **underlying trait** or **latent variable**

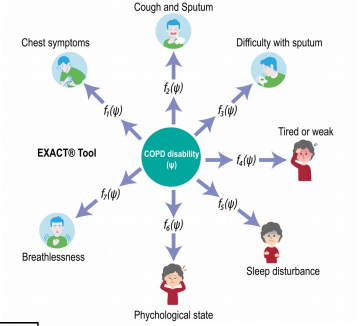
- All questionnaire information is used

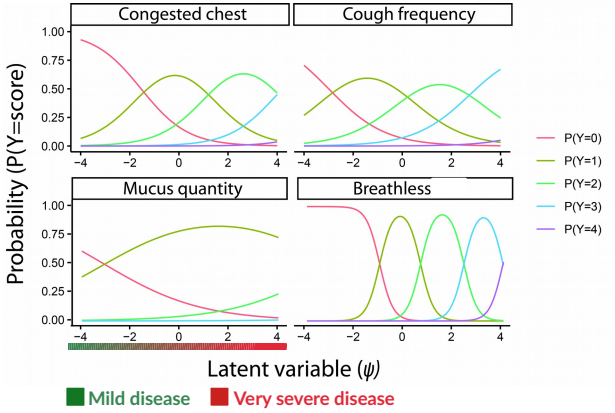
- Handles non-ignorable missingness

- Increased power to detect drug effect

$$P\left(y_{ij} \geq k\right) = \frac{e^{\left(a_j\left(\psi_i - b_{j,k}\right)\right)}}{1 + e^{\left(a_j\left(\psi_i - b_{j,k}\right)\right)}}$$

$$P\left(y_{ij} = k\right) = P\left(y_{ij} \geq k\right) - P\left(y_{ij} \geq k+1\right)$$





IRM in COPD

EXACT PRO data

Phase II

Observational Study > AAPS J. 2019 Apr 26;21(4):60. doi: 10.1208/s12248-019-0319-9.

A Novel Method for Analysing Frequent Observations from Questionnaires in Order to Model Patient-Reported Outcomes: Application to EXACT® Daily Diary Data from COPD Patients

Eva Germovsek [1], Claire Ambery [2], Shuying Yang [2], Misba Beerahee [2], Mats O Karlsson [1], Elodie L Plan [3]

> AAPS J. 2021 Jun 2;23(4):79. doi: 10.1208/s12248-021-00600-1.

Improved Decision-Making Confidence Using Item-Based Pharmacometric Model: Illustration with a Phase II Placebo-Controlled Trial

Carolina Llanos-Paez [1], Claire Ambery [2], Shuying Yang [2], Maggie Tabberer [3], Misba Beerahee [2], Elodie L Plan [1], Mats O Karlsson [4]

Phase III?

IRM in COPD

EXACT PRO data

Phase II

Observational Study > AAPS J. 2019 Apr 26;21(4):60. doi: 10.1208/s12248-019-0319-9.

A Novel Method for Analysing Frequent Observations from Questionnaires in Order to Model Patient-Reported Outcomes: Application to EXACT® Daily Diary Data from COPD Patients

Eva Germovsek [1], Claire Ambery [2], Shuying Yang [2], Misba Beerahee [2], Mats O Karlsson [1], Elodie L Plan [3]

> AAPS J. 2021 Jun 2;23(4):79. doi: 10.1208/s12248-021-00600-1.

Improved Decision-Making Confidence Using Item-Based Pharmacometric Model: Illustration with a Phase II Placebo-Controlled Trial

Carolina Llanos-Paez [1], Claire Ambery [2], Shuying Yang [2], Maggie Tabberer [3], Misba Beerahee [2], Elodie L Plan [1], Mats O Karlsson [4]

Phase III?

Clinical Trial > Adv Ther. 2018 Jan;35(1):56-71. doi: 10.1007/s12325-017-0650-4. Epub 2018 Jan 8.

Once-Daily Triple Therapy in Patients with COPD: Patient-Reported Symptoms and Quality of Life

Maggie Tabberer [1], David A Lomas [2], Ruby Birk [3], Noushin Brealey [3], Chang-Qing Zhu [4], Steve Pascoe [5], Nicholas Locantore [5], David A Lipson [5 6]

Phase III, randomized, double-blind, double dummy, parallel-group, multicenter study to evaluate once daily ***fluticasone furoate/umeclidinium/vilanterol*** (**FF/UMEC/VI**) (100 ug/62.5 ug/25 ug) inhalation powder versus twice daily ***budesonide/formoterol*** (**BUD/FOR**) (400 ug/12 ug) in patients with **COPD**

UPPSALA UNIVERSITET

Aim & Methods

To illustrate how a **new methodology** to analyse **PRO data** based on a longitudinal **IRM improves confidence** in a **Phase III clinical trial** endpoint compared with a **MMRM** analysis.

Aim & Methods

To illustrate how a **new methodology** to analyse **PRO data** based on a longitudinal **IRM improves confidence** in a **Phase III clinical trial** endpoint compared with a **MMRM** analysis.

Step 1[a] – IRM

- **Item characteristic functions**
- ✓ Graded response model (a logistic transformation to model each item)
- ✓ Independent occasion approach

Step 2[a] - Longitudinal model - modeling Ψ_i:

- Use the estimates of Ψ from step 1 as DV[b]
- Different parameters between treatment arms
- Pre-specified Weibull model
- Smoking status and geographical regions on $\psi_{i,t=0}$

a: Analysis performed using NONMEM v.7.4.4 with an Intel FORTRAN compiler and PsN v.5.1.0
b: Schindler et al. Pharm Res 2018;35:122
Ψ_i: Individual latent variable

Aim & Methods

To illustrate how a **new methodology** to analyse **PRO data** based on a longitudinal **IRM improves confidence** in a **Phase III clinical trial** endpoint compared with a **MMRM** analysis.

Step 1[a] – IRM

- **Item characteristic functions**
- ✓ Graded response model (a logistic transformation to model each item)
- ✓ Independent occasion approach

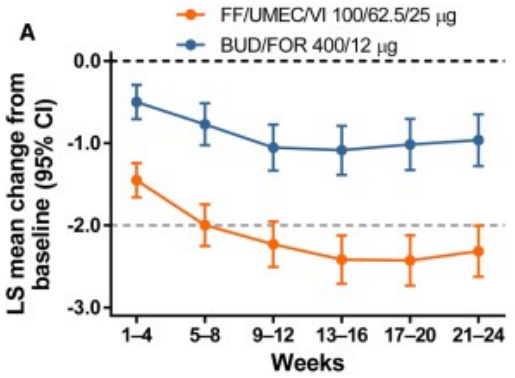
Step 2[a] - Longitudinal model - modeling Ψ_i:

- Use the estimates of Ψ from step 1 as DV[b]
- Different parameters between treatment arms
- Pre-specified Weibull model
- Smoking status and geographical regions on $\psi_{i,t=0}$

Step 3[a] - Total score (RS-total)[c] simulations and CI calculation for each treatment arm

- Perform simulations including parameter uncertainty ($PRIOR)
- Calculate 95% CI –Mean difference in total score between baseline and four week intervals:
  - ✓ Week 0 – 4, week 5 – 8, week 9 – 12 ...

Published MMRM results

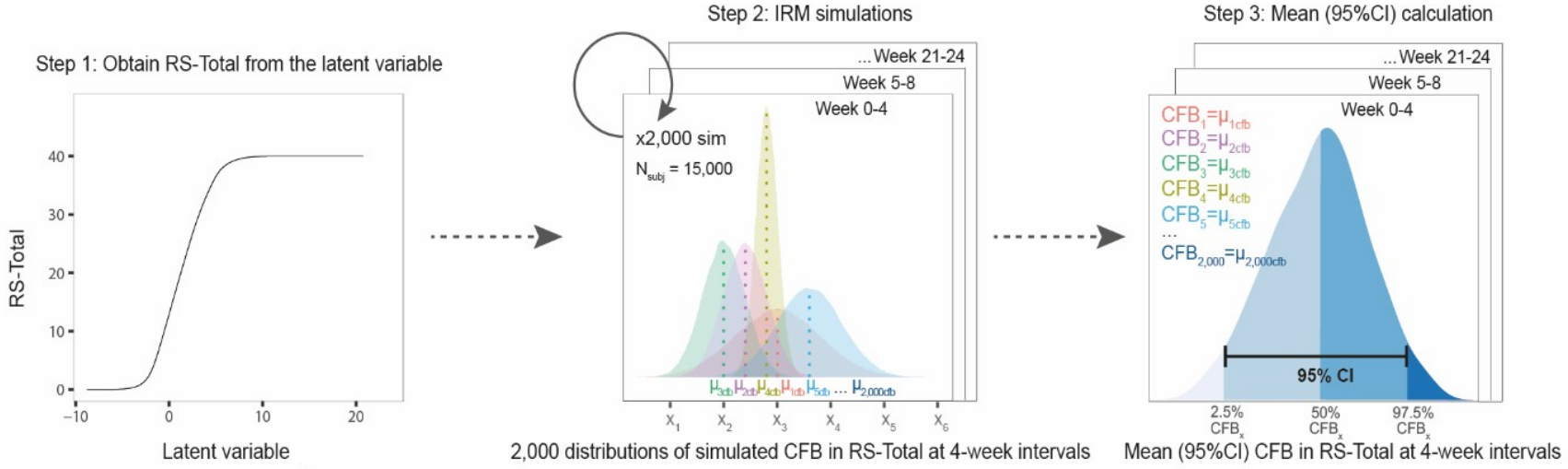


Tabberer et al. Adv Ther 2018; 35:56-71

a: Analysis performed using NONMEM v.7.4.4 with an Intel FORTRAN compiler and PsN v.5.1.0
b: Schindler et al. Pharm Res 2018;35:122
Ψ_i: Individual latent variable
c: Total score from the E-RS:COPD questionnaire

Simulations

Including parameter uncertainty



2,000 distributions of simulated CFB in RS-Total at 4-week intervals

Mean (95%CI) CFB in RS-Total at 4-week intervals

Simulations included uncertainty for each parameter value ($PRIOR)

NWPRI:

- Multivariate normal distribution for THETAs
- Inverse Wishart distribution for OMEGAs

Sample size (N)

$$N = \left(\frac{CI_{MMRM}}{CI_{IRM}} \right)^2$$

UPPSALA UNIVERSITET

Data

Baseline characteristics	FF/UMEC/VI (n=907)	BUD/FOR (n=894)
Age (years)	64.2 (8.56)	63.6 (8.73)
FVC (L)	2.84 (0.80)	2.87 (0.79)
FEV_1 (L)	1.25 (0.46)	1.24 (0.45)
Male (n)	675 (74%)	658 (74%)
Smoker (n)	396 (44%)	392 (44%)
COPD GOLD disease status	Moderate: 298 (33%) Severe: 501 (55%) Very severe: 107 (12%)	Mild: 1 (0.1%) Moderate: 290 (32%) Severe: 477 (53%) Very severe: 124 (14%)
RS- total score[a]	12.2 (5.85)	12.9 (5.96)

Values are mean (SD) or n (%); a: scores were calculated as the mean value during baseline period defined as from day -14 to day -1; FVC: forced vital capacity ; FEV_1: forced expiratory volume in one second; GOLD: global initiative for chronic obstructive lung disease; RS-total score ranged from 0 to 40; In this analysis, nine patients were excluded from the intention to treat population of the FULFIL clinical trial (NCT02345161) for the following reasons: absence of E-RS:COPD score data for the whole study period (4 patients), dispensing errors (4 patients), and missing recorded time (1 patient).

Parameter estimates

Longitudinal model

Parameter	FF/UMEC/VI Value (RSE)	BUD/FOR Value (RSE)
Baseline ($\Psi_{i,t=0}$)[a]	0.33 (0.23)	0.29 (0.28)
Time of response (T_R)	0.08 (0.03)	0.08 (0.03)
Maximum response (R_{MAX})	-0.31 (0.11)	-0.16 (0.25)
γ	9.27 (0.13)	16.9 (0.75)
Offset	-0.27 (0.08)	-0.08 (0.30)
ω^2 Baseline	1.09 (0.03)	1.47 (0.03)
ω^2 Time of response	0.45 (0.03)	0.46 (0.03)
ω^2 Maximum response	0.89 (0.04)	0.98 (0.06)
ω^2 Offset	0.37 (0.05)	0.42 (0.05)
RUV	0.32 (0.02)	

a: Effects of smoking status and geographical regions on $\psi_{i,t=0}$ were included in the model as these covariates were considered in the MMRM analysis

$$\Psi_i(t) = \Psi_{i,t=0} + R_{MAXi} \cdot \left(1 - e^{\left(-\left(\frac{\ln(2)}{T_{Ri}} \cdot t\right)^{\gamma}\right)}\right) + Offset_i$$

- Weibull describes a late onset effect
- Offset near-instant or potentially symptomatic drug effect

UPPSALA UNIVERSITET

Parameter estimates

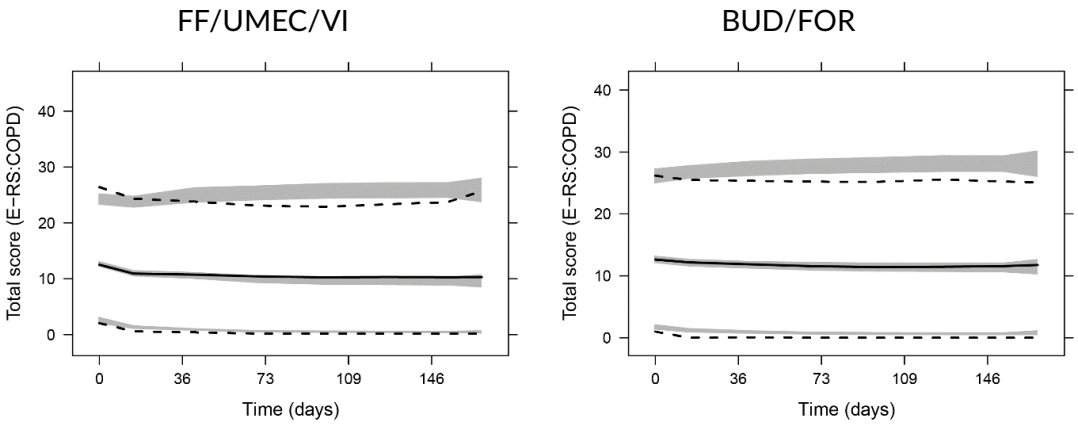
Longitudinal model

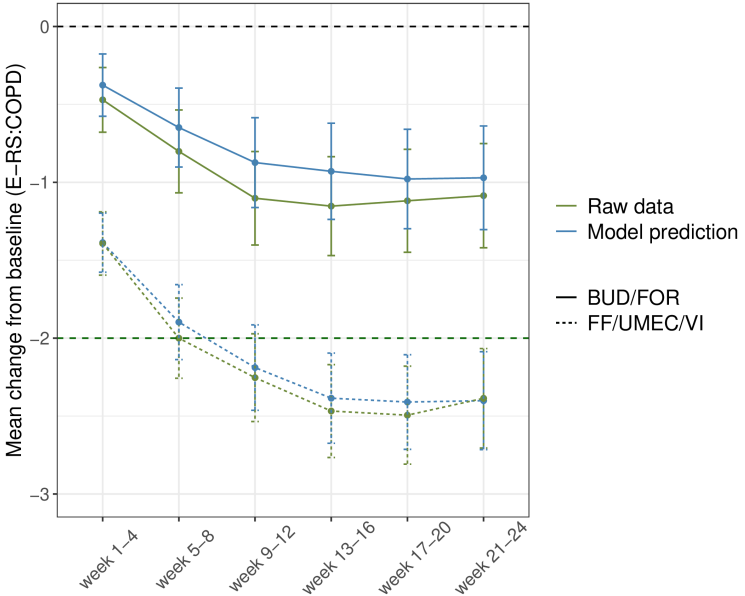
Covariates	Value (RSE)
Smoking effect	0.13 (0.62)
Region 1	-0.64 (0.13)
Region 3	-0.61 (0.13)
Region 4	-0.20 (0.40)
Region 5	-0.41 (0.27)
Region 6	-0.98 (0.12)

Region 1 (21%): Germany, Greece, Italy
Region 2 (24%): Russian Federation, Ukraine
Region 3 (21%): Bulgaria, Hungary, Romania, Slovakia
Region 4 (18%): Czech Republic, Estonia, Poland
Region 5 (6%): China, Republic of Korea
Region 6 (10%): Mexico

	Correlation	Value (RSE)
FF/UMEC/VI	ω^2 Maximum response~ω^2 Offset	11% (0.24)
	ω^2 Baseline~ω^2 Offset	-12% (0.20)
BUD/FOR	ω^2 Maximum response~ω^2 baseline	-13% (0.17)
	ω^2 Baseline~ω^2 Offset	-10% (0.27)

Model predictive performance

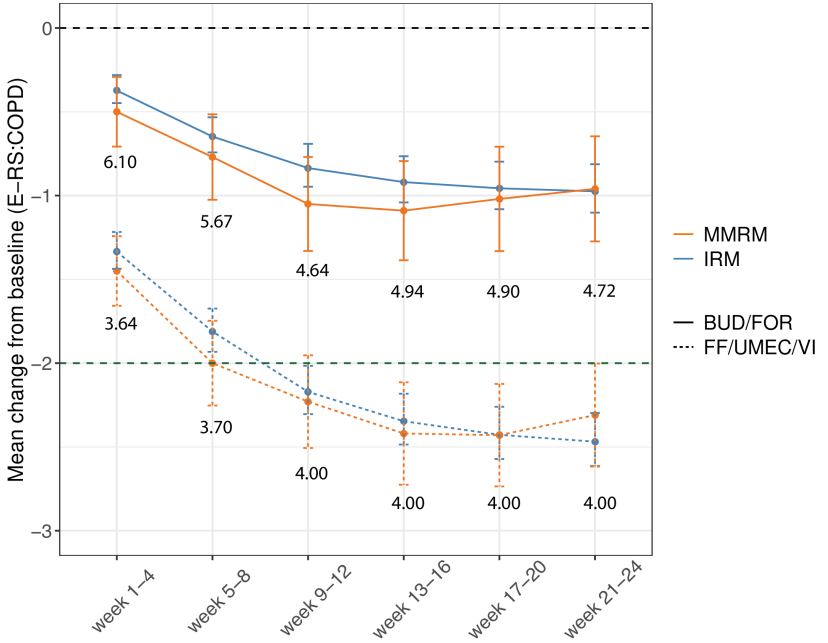




Lines are the 2.5th, 50th and 97.5th percentile of the observed data, and grey areas are the corresponding 95% confidence interval from model simulations (N=500)

Precision in outcome measures

- ✓ The IRM improved precision of the estimated drug effect compared to MMRM
- ✓ A **4.0 (FF/UMEC/VI)** and **4.72 (BUD/FOR)** times larger sample size for the MMRM analysis to achieve the precision obtained with IRM at week 21 – 24



Discussion & conclusion

- Advantage of using a non-linear mixed effect model-based analysis of Phase III clinical item-response level data, over MMRM for the same clinical endpoint in both methods (IRM and MMRM)

- Proposed methodology is applicable to any item-level data

- IRM may improve decision-making in drug development:
  - ✔ Prediction of outcomes from short term to long term study
  - ✔ Smaller studies sizes and more informative analysis

Acknowledgments

- **Maggie Tabberer** for her contribution to the data analysis and interpretation
- **GSK** funded this study in the form of a research payment to Uppsala University
- **Pharmacometrics group** Uppsala university
- **Scientific organising committee** 29th PAGE 2021

Conflict of interest: CL-P, EP and MOK declare that they have no conflict of interest. CA, SY, and MB are GSK employees and hold GSK shares