

## ***Modeling & Simulation using metadata***

***away from inventing the wheel again***

Klaas Prins, PhD  
Senior Scientist  
Strategic Consulting Services  
Pharsight Corporation



# Outline

- **Metadata**
  - Capture process
  - Entry process
  - Caveats of using metadata
- **Case study**
  - M&S in Rheumatoid Arthritis
- **Conclusion**

# The danger of trusting summary statistics.

A statistician once waded through a river that was  
on average 1 meter deep...  
... and he drowned.

*Godfried Bomans (Dutch writer)*

# Searching for summary statistics is unnecessarily expensive.

- Year-over-year increase in multimillion budget allocated to assess and prevent the risk of 'another 9/11'.
- Well, we all know the average risk of another 9/11 is the same every year:
- ...1 / 365.25

# Models provide rational framework for integration and organization of disparate compound/biologic knowledge

## Model-based drug development

- An integrated, data-driven, model-based decision-making methodology
- Integrate all relevant public and proprietary data, spanning:
  - discovery to clinical
  - in-house data to competitors' information
  - healthy volunteers to patients
  - Build probabilistic model(s) of the compound's attributes and product profile in the context of a competitive landscape
- Simulate development scenarios of interest

Understanding and embracing uncertainty is key to quantitative decision making and effective risk management

# Why should we consider quantifying metadata containing published material?

- Placebo has been studied to death in a very heterogeneous patient population.
  - You will never be able to study placebo to this extent as a function of this many patient and/or trial characteristics in this many patients.
- In your head, you are always making comparisons of your drug versus comparator responses reported in public trial reports.
  - Don't you want you know how (im)precise such comparisons actually are?
- The clinical program may not provide for study of a certain comparator, yet the comparison is urgently warranted by the team
  - The required information may be accessible by quantifying public domain data.
- Statistical methods are in place to pool model predictions based on public domain with model predictions based on in-house data.
- Broaden the horizon of drug development team
  - qualitative sense: more drugs & more factors of impact
  - quantitative sense: distributions rather than point estimates.

# Public Data Sources

- Public data sources on competitors and analogues from the clinical and scientific literature, collectively called “meta data”, are diverse and include:
  - Published journal articles
  - Regulatory documents (e.g., FDA Summary Basis for Approval)
  - Package inserts or promotional materials from the drug manufacturer
  - Published abstracts or poster presentations
  - Meeting proceedings
  - Web documents (e.g., press releases about new clinical trial results)
  - Online clinical trial registries
- Digitization software is available to accurately capture data from graphs.
- All data are formatted and entered in a meta database that undergoes multiple quality design and check stages

# Data and its sources need to be discussed extensively before being entered into any data base.

- Building a data base = librarians task? Think twice!!
- Capturing data requires repeated extensive discussion with
  - Librarians
  - Data base specialists
  - Modeling scientists / Biostatisticians
  - Clinical Experts (clinical pharmacologists, MDs)
- Clinical expertise is crucial
  - Clinicians typically have read the key papers and understand the nature and features of the endpoints, populations, etc...
  - What endpoints are hot and which are not?
  - Ensure endpoint models built are useful in development discussions
  - Defining assumptions
- Clinician's involvement from the start creates buy-in and commitment
  - Clinician's are the modeling & simulation effort's clients.

# Caveats of metadata capturing and processing.

- **Be aware of sources of bias**
  - Publications tend to be overenthusiastic on drug effect
  - Limited data availability may lower the criteria for database entry
- **Observed case / LOCF or both?**
- **What metrics are reported?**
  - Mean or median
  - Proportions
    - Prop. Pts with at least one point improvement vs baseline,
    - Prop. Pts. With at least 25% improvement vs. baseline
- **Important to go after baseline values!**
- **Placebo-controlled or also include open label studies?**
- **Patient level data / summary statistic / both**
  - Separate models
  - Joint summary statistic & patient data models: full bayesian analysis (WinBUGS)
- **Can we combine multiple endpoints that really mean the same thing?**
  - Increase data density by combining through conversion or challenged assumptions

# A meta data base : what does it look like?

- Spreadsheet program (e.g. Excel®) containing the data.
- Endnote / Reference Manager files containing references that are linked via a unique number to the spreadsheet.
- Screenshot of the data base for Diabetes:

	A	B	C	D	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP
	Reference	Protocol	Mod	Who	Time	T unit	Endpoint	Matrix	Fasted measurement	Variable	Value	Value unit	SD	SE
1														
2	91	core study		IDC	0	wks	HbA1C	blood	fasting	mean	7.81	percent		0.11
3	91	core study		IDC	4	wks	HbA1C	blood	fasting	mean	7.70	percent		0.12
4	91	core study		IDC	8	wks	HbA1C	blood	fasting	mean	7.79	percent		0.11
5	91	core study		IDC	12	wks	HbA1C	blood	fasting	mean	7.89	percent		0.14
6	91	core study		TAB	0	wks	HbA1C	blood	fasting	mean	7.62	percent		0.08
7	91	core study		TAB	4	wks	HbA1C	blood	fasting	mean	7.35	percent		0.08
8	91	core study		TAB	8	wks	HbA1C	blood	fasting	mean	7.23	percent		0.09
9	91	core study		TAB	12	wks	HbA1C	blood	fasting	mean	7.16	percent		0.11
10	91	extension study		IDC	0	wks	HbA1C	blood	fasting	mean	7.75	percent		
11	91	extension study		IDC	4	wks	HbA1C	blood	fasting	mean	7.58	percent		0.15
12	91	extension study		IDC	8	wks	HbA1C	blood	fasting	mean	7.61	percent		0.12
13	91	extension study		IDC	12	wks	HbA1C	blood	fasting	mean	7.77	percent		0.14
14	91	extension study		IDC	16	wks	HbA1C	blood	fasting	mean	7.83	percent		0.16
15	91	extension study		IDC	24	wks	HbA1C	blood	fasting	mean	7.90	percent		0.22
16	91	extension study		IDC	36	wks	HbA1C	blood	fasting	mean	7.96	percent		0.21
17	91	extension study		IDC	52	wks	HbA1C	blood	fasting	mean	8.36	percent		0.24

List of randomly selected metadata bases built at Pharsight. There are actually databases available on 18 Therapeutic Areas.

Project	Abstracts Reviewed	Studies Included	Drugs Included
Sleep	195	51	Diazepam, Flunitrazepam, Flurazepam, Lorazepam, Lormetazepam, Triazolam, Zaleplon, Zolpidem
Asthma	118	31	Beclomethasone dipropionate, Montelukast
Asthma/COPD	570	104	Beclomethasone dipropionate, Fluticasone, Montelukast, Albuterol, Theophylline, Cilomilast, Ipratropium, Tiotropium, Salmeterol
Depression	940	129	Citalopram, Duloxetine, Fluoxetine, Fluvoxamine, Nefazodone, Paroxetine, Reboxetine, Sertraline, Venlafaxine
Schizophrenia	620	90	Aripiprazole, Chlorpromazine, Clozaril, Haloperidol, Olanzapine, Risperidone, Ziprasidone, Quetiapine, Amisulpride
Diabetes	750	106	Glyburide, Glipizide, Metformin, Glimepiride, Tolazamide, Chlorpropamide, Exenatide, Liraglutide, Pioglitazone, Rosiglitazone, Troglitazone, Vildagliptin, Sitagliptin
Rheumatoid arthritis (RA)	170	38	Adalimumab, Anakinra, Etanercept, Methotrexate
Gastro-esophageal reflux disease (GERD)	250	70	Cisapride, Esomeprazole, Lansoprazole, Metoclopramide, Mosapride, Omeprazole, Pantoprazole, Rabeprazole, Tegaserod, Zaccopride

## Case study is focused on Rheumatoid Arthritis (RA)

- RA = therapeutic area with multiple endpoints & new, promising drug molecules
- Competition is fierce as standard of care anti-TNF- $\alpha$  antibodies show very good efficacy/AE profiles.
- Various pharma companies are investing in NCEs for RA treatment aiming to at least equate the standard of care.
- Abstracts Reviewed: 170
- Studies Included: 38
- Treatments Included
  - Placebo, Adalimumab, Anakinra, Etanercept, Methotrexate
- Scope of Therapies
  - Drugs intended to mitigate the signs and symptoms of rheumatoid arthritis

# RA database details (drugs, classes, endpoints)

<b>Drugs</b>	<b>Class</b>	<b>Endpoints</b>
Adalimumab	anti-TNF alpha	ESR, SJC, PGA, MDGA, SDAI, HAQ, ES, TSS, JSN, Serious Infection
Anakinra	interleukin antagonist	SJC, TJC, MDGA, PGA, HAQ, CRP, ESR, ACR20, ACR50, ACR70, SDAI, CRP, DAS28, TSS, SJC, Serious Infection
Etanercept	anti-TNF alpha	ESR, DAS28, PGA, MDGA, SJC, TJC, ACR20, ACR50, ACR70, DAS28, CRP, SDAI, TSS, ES, JSN, CRP, Infections, Serious Infection
Methotrexate	antimetabolite (dihydrofolic acid reductase inhibitor)	ACR20, ACR50, ACR70, TSS, JSN, ES, CRP, TJC, SJC, DAS28, MDGA, SDAI, PGA, ESR, Infections, Serious Infection

..and of course the entire package of endpoints is available for placebo treatment!

# The hypothetical client was at start of the project receptive to M&S and use of metadata analysis.

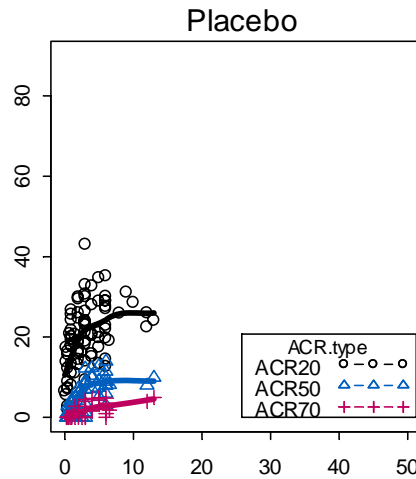
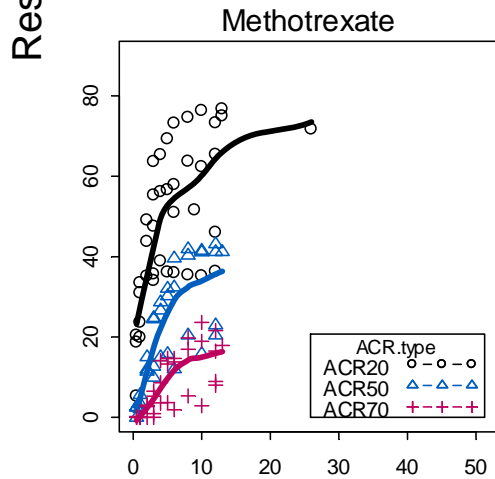
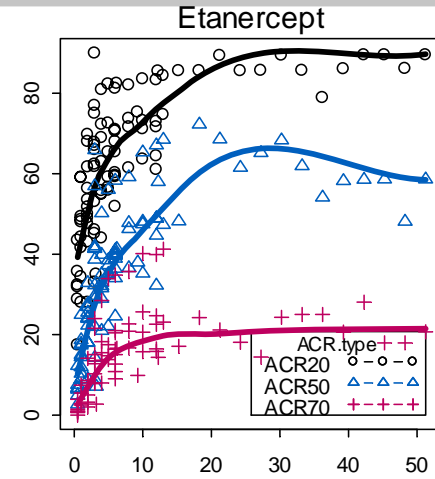
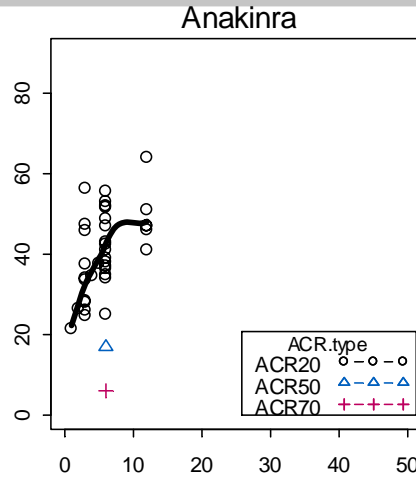
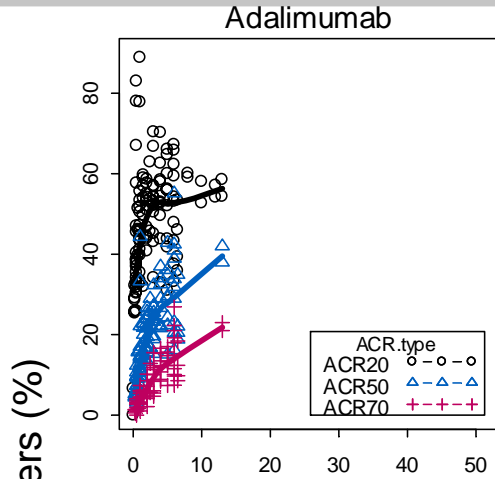
- No uncertainty range defined around Critical Success Factors
  - “ACR20 Enbrel = 80% at steady state effect” => this statement is far too precise
  - Need to be non-inferior to Enbrel (what does that exactly mean/imply?)
- Clinical team’s perception of competitor landscape was maybe qualitative to semi-quantitative, but not fully quantitative
  - Example : client was convinced Enbrel beats MTX in terms of efficacy, but was not aware of the associated probabilities across all uncertainties.
  - Knew roughly the effect of marketed doses, but did not know the full dose-response curve of competitor drugs
- They decided they wanted fully integrated models for ACR20, ACR50, ACR70 for the most prominent competitors in the field.

Aim was to have a quantitative reference cadre of the competitive landscape in place for when new data of their lead compound currently in phase IIa would come in.

# We built a comprehensive ACR efficacy model using the metadata set

- ACR – American College of Rheumatology
- For example: ACR20 = Proportion of patients showing:
  - 20% improvement in tender and swollen joint counts
  - 20% improvement in  $\geq 3$  out of 5 other measures of RA activity
- A joint ACR20, ACR50, ACR70 model
  - Relatively rich data set
  - Model as function of drug, time, dose and other covariates
  - Acknowledge drug-specific differences in time course
  - Dose response for nearly all competitors
- Simulations from this model were uploaded to DMX<sup>®</sup>
  - communication and visualization tool to interactively display multidimensional simulation results

# Time course of ACRn scores of a selection of anti-RA drugs & Placebo on as collected from the literature



Time (weeks)

Dots are observed published data  
 Lines are lowest smooths

## An ACRn Model was established.

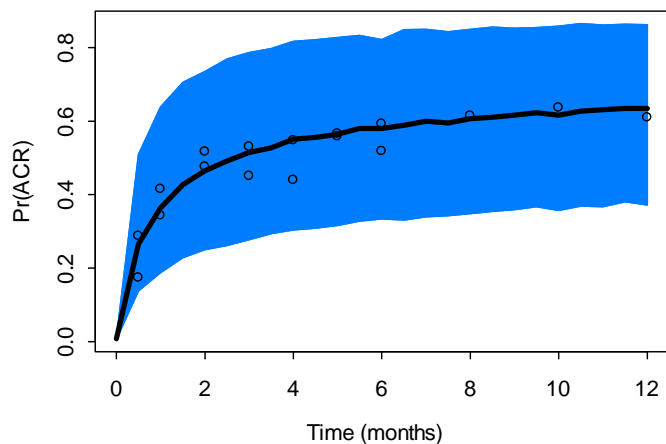
- Joint estimation of ACR20, 50 and 70 ( $f(ACR)$ )
- Logit-transformed fraction of responder data was modeled assuming normal distribution.
- Saturable time course model:  $t/(t_{50}+t)$

$$\text{logit}(E(t, drug, dose)) = \frac{\text{structural model}}{t_{50} + t} + \text{trial error model} + \text{residual}$$
$$\text{logit}(E(t, drug, dose)) = \frac{f(ACR) \cdot Emax_{drug} \cdot t}{t_{50} + t} + N(0, \omega^2) + N(0, \sigma^2 / N)$$

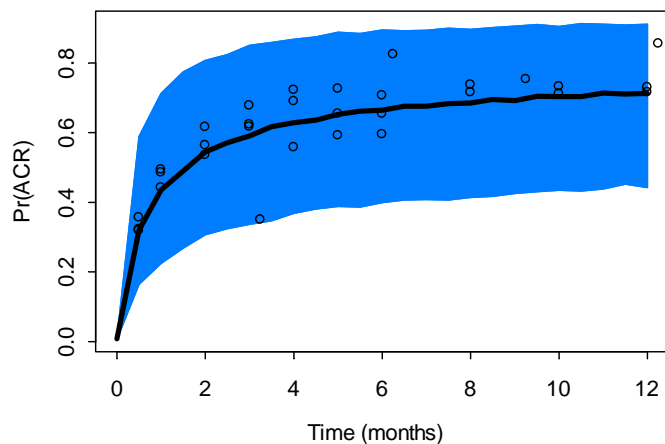
- Different effect sizes for each treatment ( $Emax_{drug}$ )
- Difference in results across trials was modeled by an additive random inter-trial error ( $\omega^2$ ).
- Simple additive residual error model ( $\sigma^2$ ), weighted by sample size.

# Visual Predictive Check for Etanercept confirmed the model was doing a good job in describing the data.

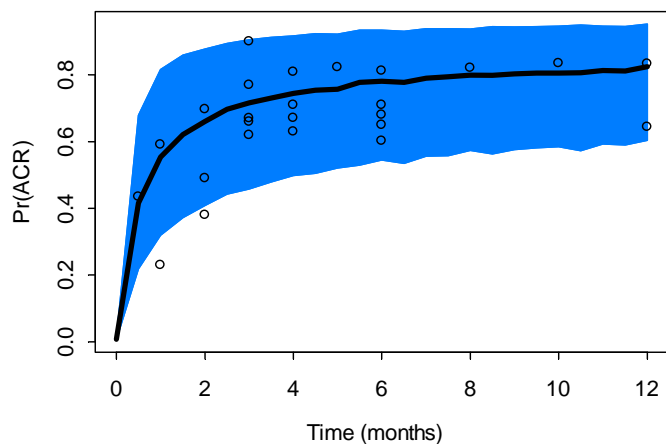
Etanercept TYPE=20 MTX=0 Dose=10



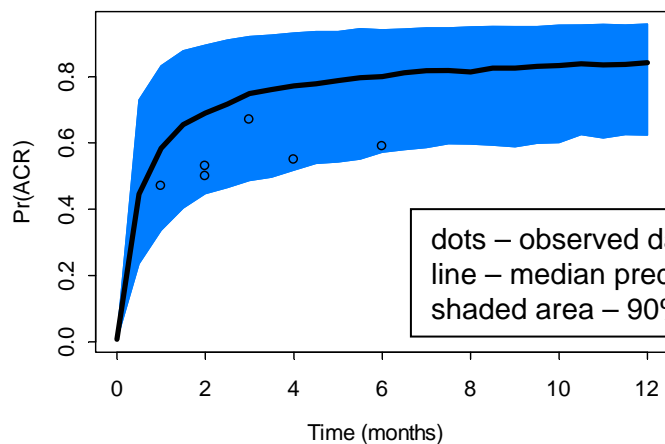
Etanercept TYPE=20 MTX=0 Dose=25



Etanercept TYPE=20 MTX=1 Dose=25



Etanercept TYPE=20 MTX=1 Dose=50



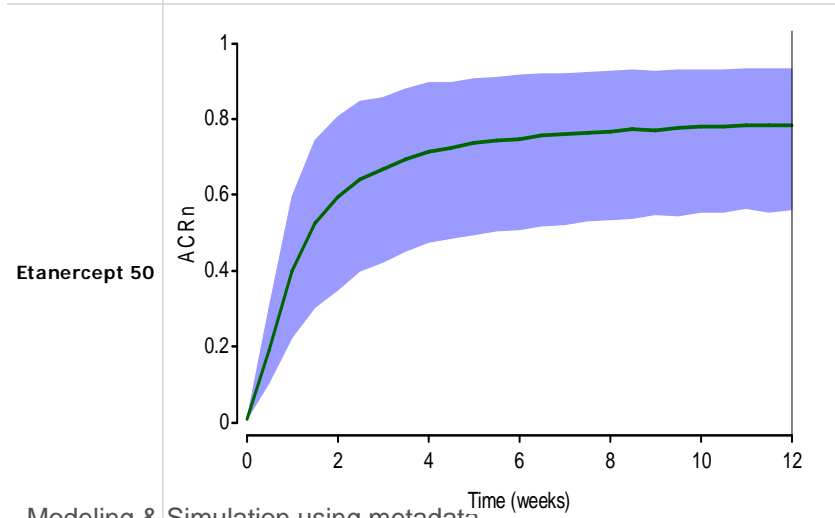
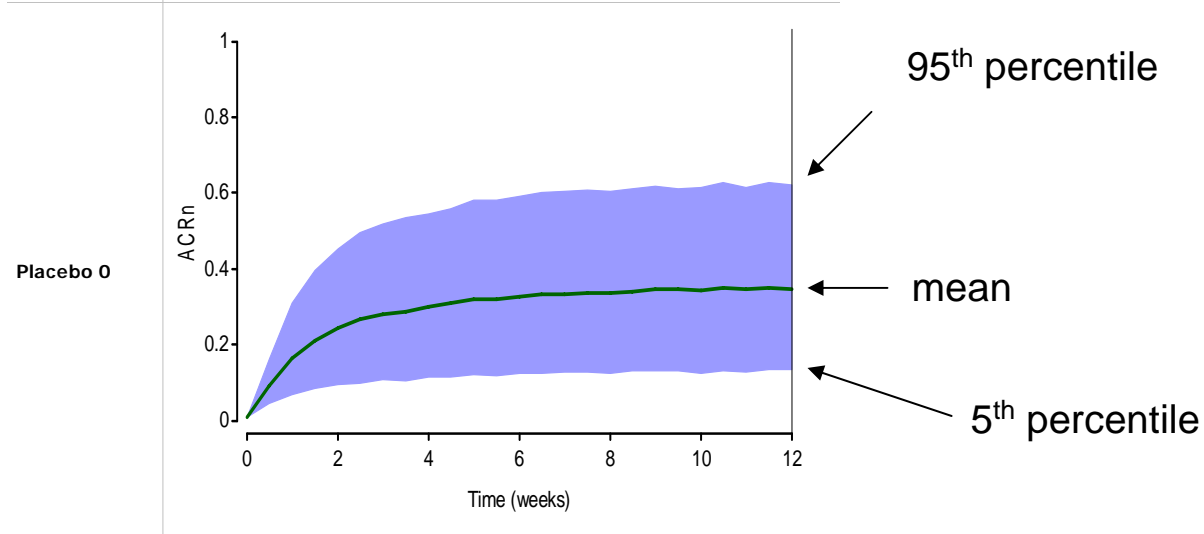
dots – observed data  
line – median prediction  
shaded area – 90% prediction interval

## Simulations of expected mean $\Pr(\text{ACR})$ given certain covariate combinations can now be explored to address development team questions.

- Predictions of the model using the uncertain parameter estimates were obtained for given sets of covariates
  - done by sampling from the multivariate parameter estimate distribution
- To each prediction a single random draw from the inter-trial error distribution was added.
  - This reflects our conservativeness towards these summary statistic models that typically originate from heterogeneous trial designs and populations.
- These predictions reflect the state of knowledge of the mean response in the infinite patient population.
- The resulting grid of response predictions was explored in DMX<sup>®</sup>.
- DMX<sup>®</sup> is a communication and visualization tool that facilitates point-and-click exploration of predictions of multivariate models.

# Screen shot 1 of the DMX<sup>®</sup> viewer containing these simulations. Plots and table highlighting mean drug response & uncertainty.

2 Plots  
 Treatment  
**ACRn vs Time (weeks)**  
 ACR type: 20



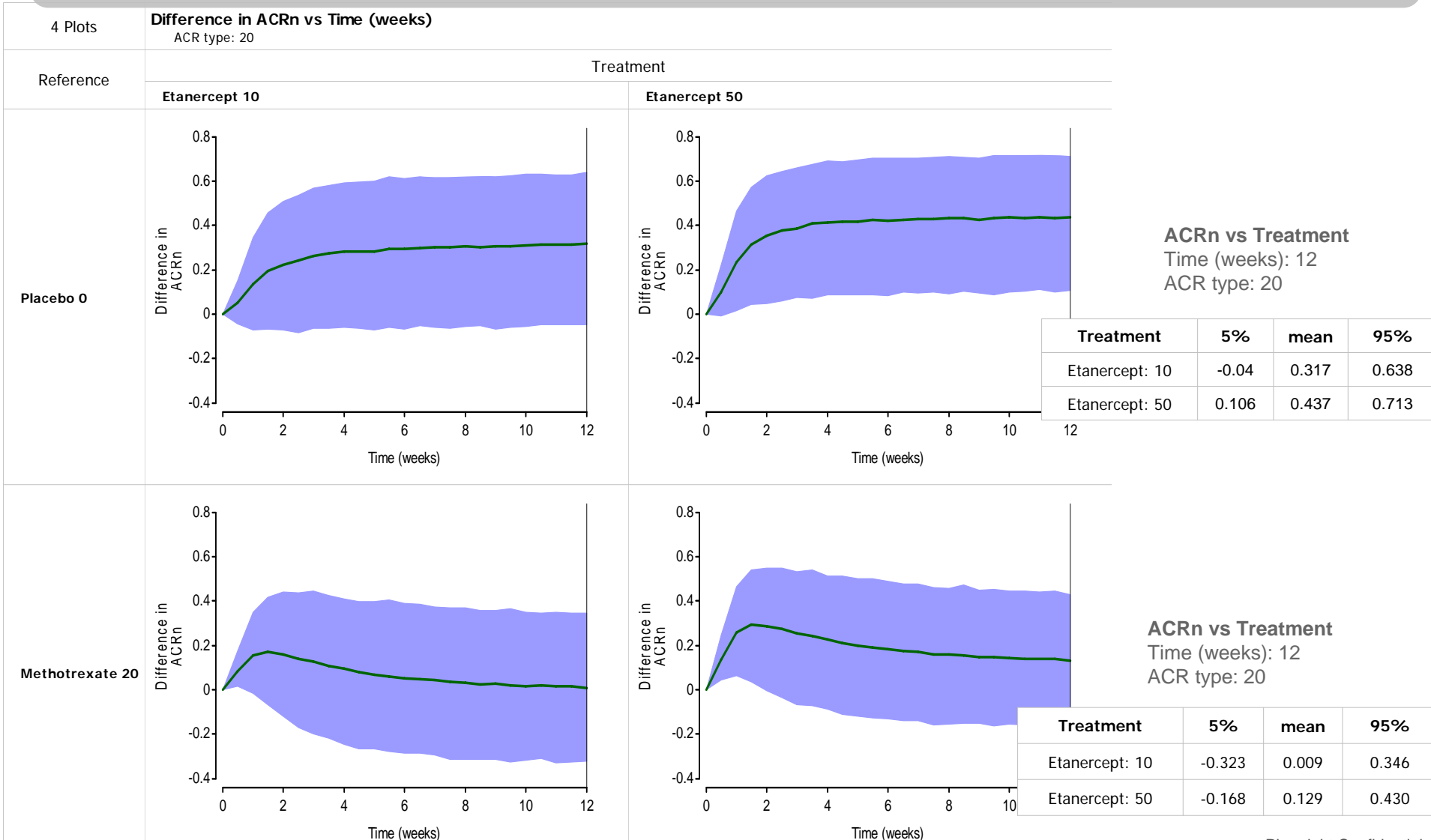
### ACRn vs Treatment

Time (weeks): 12

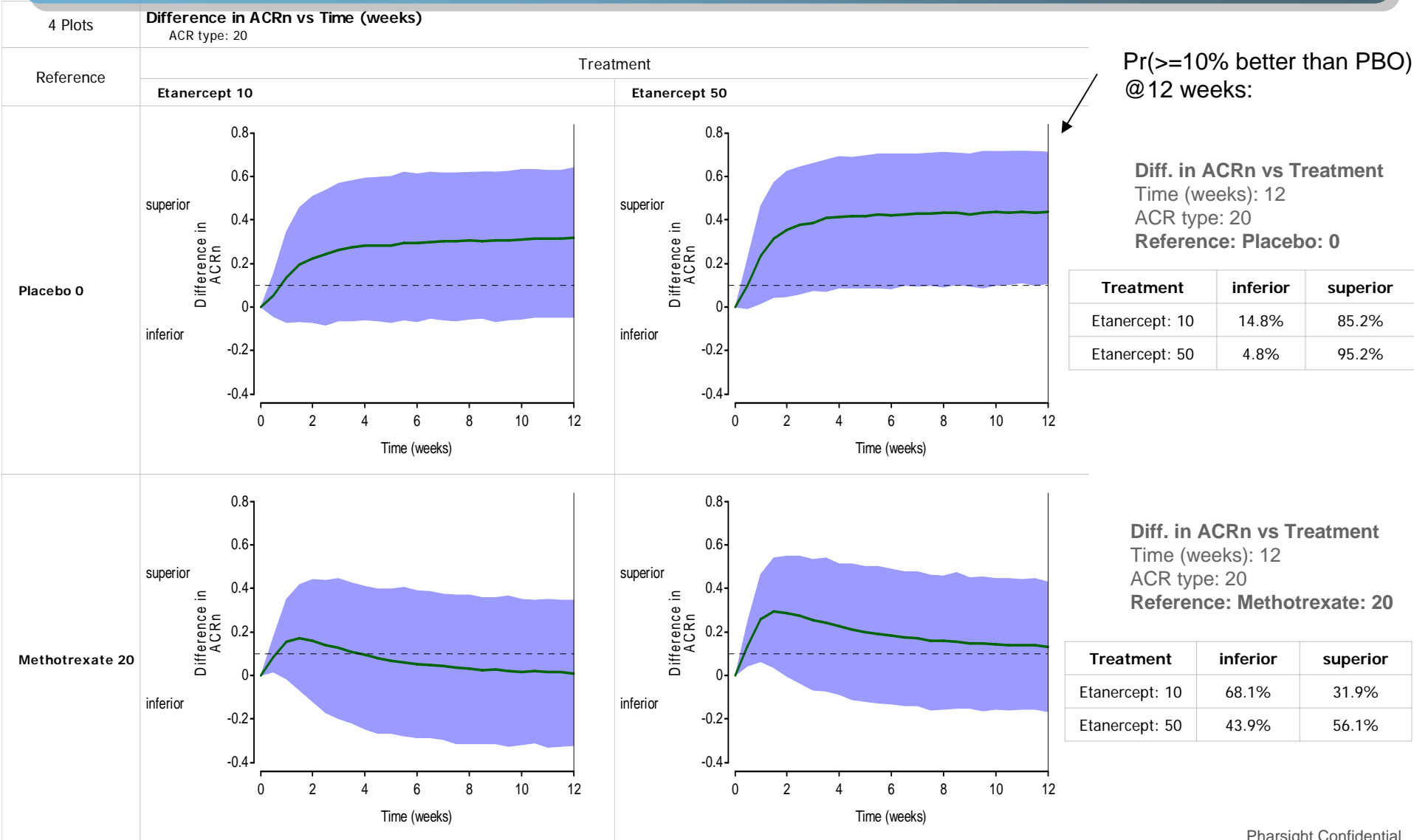
ACR type: 20

Treatment	5.0%	mean	95.0%
Placebo: 0	0.132478	<b>0.347253</b>	0.619311
Etanercept: 50	0.559614	<b>0.784595</b>	0.931161

# Screen shot 2 of the DMX<sup>®</sup> viewer containing these simulations. Plots highlighting mean drug response vs. reference treatment & uncertainty.



# Screen shot 3 of the DMX<sup>®</sup> viewer containing these simulations. Plots of drug response vs. reference treatment with uncertainty and tables with probabilities of hitting a defined target.



# The quantitative imprint of the competitive landscape has been established using published data!

- Simulations in DMX are currently used by clinicians to reference drug performance vs. competitors.
- They have a quantitative understanding of the RA therapeutic area based on real (published) data.
- They have predictions of efficacy across many covariates incorporating all uncertainties captured during modeling of the available literature data
- Note that this all was done using published data only!
- This framework is a living object:
  - Can continuously be updated with new trial data being published.
  - Can continuously be updated when new internal (patient) data comes in.
  - Creating an ever growing quantitative picture of the client's pipeline in relation to the competition.
  - DMX<sup>®</sup> software has been specifically designed for these type of analyses.

# Was the gain in knowledge from this effort unique to performing this type of metadata analysis?

- Yes, no pharma company will execute trials investigating efficacy/safety of all compounds in all patient settings.
- Creates a state of knowledge far beyond what any pharma company could learn from their development program.
- More precise estimate of placebo/drug effect
  - across multiple covariates
  - in a wider/more heterogeneous patient population
- Pharma does not have to *invent the wheel again*; the information is out there and ready to be picked up from the street.
- All it needs is data management, appropriate modeling technology & clinical expertise to quantify the information with its uncertainties.

# We or the client itself could have done more...and still can!

- Metadata sheet contains many different endpoints.
- Models developed were for Signs and Symptoms of RA (ACRn)
- We could have established models for structural joint damage reversion claims
  - E.g. Joint Space Narrowing
- Adverse events models
  - E.g. rate of serious infections at injection site
- The client still has the metadata sheet.
- Should they need additional models to be set up, any modeler could be working on this task the moment the decision has been taken.
- A huge time efficiency advantage.

Predictions from models based on publication data can be pooled with predictions from models based on patient data.

- Depends on the required inferences from these models.
- Typically the Drug Development Team wants to know about the mean drug response in the entire patient population
- Requires simulations to be set up accordingly
  - Publication data model: the distribution of model predictions across uncertainty are inferences of mean response.
    - Could require our conservative approach of adding inter-trial error draws
  - Patient data: the distribution of expected mean predicted trial results in trials with 'infinite' sample size.
  - Then pool these and explore using for example DMX<sup>®</sup>.

# Conclusion

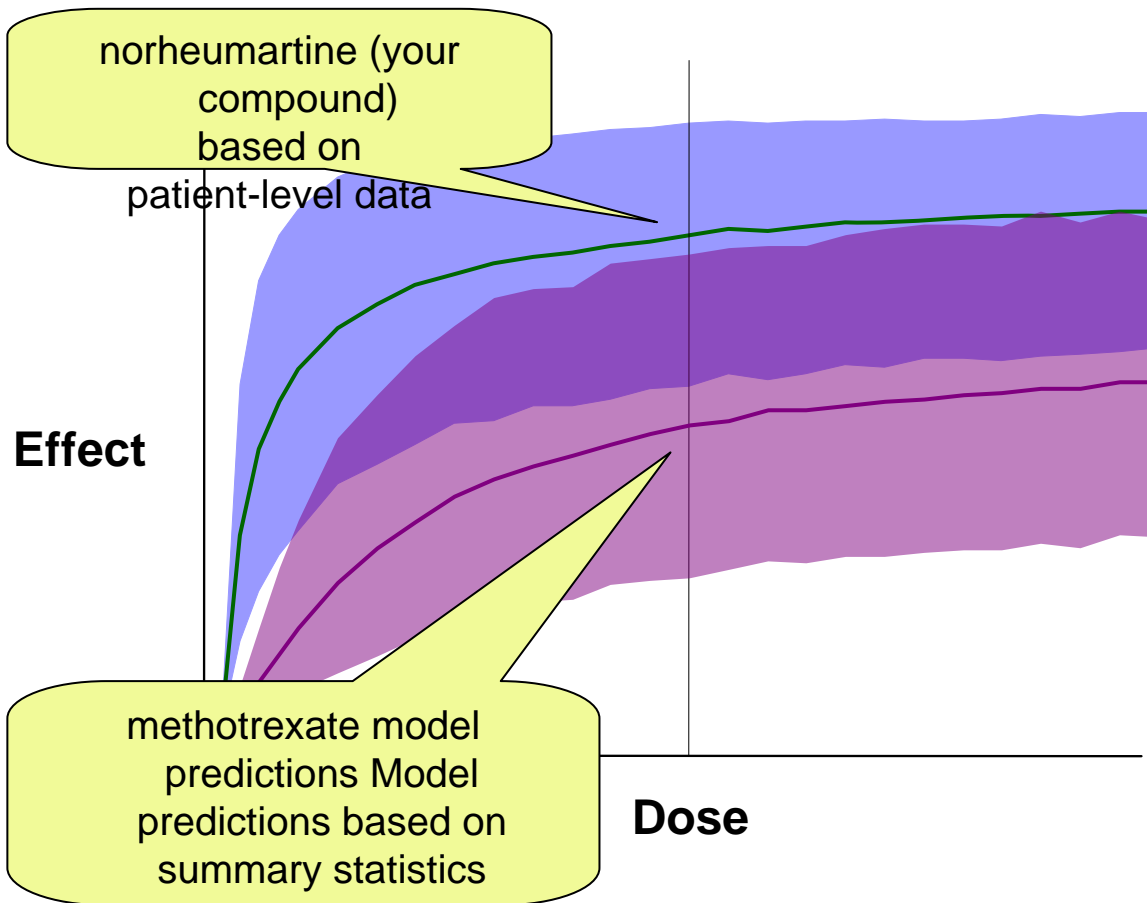
- We need to move from discussion of publication point estimates to probabilistic estimates of (competitor) drug response.
- This requires quantifying public domain data in metadata models.
- Metadata capture is an extensive process requiring involvement of multiple disciplines of pharma industry.
- Statistically sound metadata models can be established from which predictions can be created that are useful inputs for development discussions.
- Metadata model predictions can be pooled with model predictions from patient data models.
- DMX<sup>®</sup>-like communication of metadata model predictions facilitates drug development discussions and, among other benefits, takes the regression analysis out of the black box for clinicians.

# Backup Slides



# Beginning with the end— what we ultimately want is a robust prediction of the future to help answer strategic questions

## Example: a new drug's dose-response vs. that of marketed comparators



### Possible Strategic Questions

- What is the probability our drug can be superior to comparator at dose X?
- Which dose gives us the highest chance of being superior? Non-inferior?
- Before we invest in a clinical trial, is there any chance we can meet our success hurdle?
- What is the effect of co-therapy? How does response react in populations with a higher biomarker level at baseline? Should we tailor our entry criteria?

# Summary metric (e.g., population mean, or median) data vs. patient-level data: Both are useful and “poolable”.

- We use the data that are fit-for-purpose to the question
  - **Summary metric:** Often teams want to know about the effect of a drug as expressed in a population-level metric (e.g., the mean, median, etc. response)
  - **Patient-level data:** Other times, teams need to know about predictions for *individual* patient responses and the variability among patients
- For developing a drug, the population outcome matters: is the drug going to be successful in the target patient population?
  - still decades away from a fully integrated ‘right drug for the right patient’
  - Model-based drug development thus requires drug models for the population
  - (This does not contradict the need for clinical trial simulation work to optimize trial outcome)
- Typically, pooling summary statistic and patient-level data models entails making predictions of the response in the entire patient population.
- This requires different simulation techniques for both types of models.

# Technical details for integrating summary statistic and patient-level data models.

- **Statistically sound approaches exist for both**
  - Mixed effects techniques allow to distinguish between the parameter uncertainty and that caused by the variability between studies (summary statistics) and individuals (patient-level data)
  - Exploration and appropriate incorporation of covariates can explain systematic variation between studies from different literature sources
- **How do we actually simulate the expected mean/median response in the target population?**
  - summary statistics models
    - Add a single draw from inter-trial variability estimate to each prediction across model parameter uncertainty.
  - Patient-level data models
    - Simulate a super trial (10,000 patients or so) for every parameter draw from the uncertain multivariate model parameter distribution. Take the trial mean or median.
- **The resulting simulations can simply be pooled as they are making inferences about the same thing!**
  - Using appropriate communication tools, such as DMX<sup>®</sup>, one can make relative inferences from patient-level data models vs. summary statistic models

# RA database details (data dictionary and abbreviation definitions)

<b><u>abbreviation</u></b>	<b><u>definition</u></b>
ACR 20	20 percent improvement on the American College of Rheumatology criteria
ACR 50	50 percent improvement on the American College of Rheumatology criteria
ACR 70	70 percent improvement on the American College of Rheumatology criteria
ACR 90	90 percent improvement on the American College of Rheumatology criteria
CRP	c-reactive protein
DAS	disease activity score
DAS 28	28-point disease activity score
ES	erosion score
ESR	erythrocyte sedimentation rate
HAQ	Health Assessment Questionnaire
JSN	joint space narrowing
MDGA	physician's global assessment
PGA	patient's global assessment
SDAI	specific disease activity index
SJC	swollen joint count
TJC	tender joint count
TSS	total Sharpe score