The Decision-Equivalence Metric (DEM) for Evaluating Surrogate Endpoints for Accelerated Clinical Development

ISoP/ASA joint workshop series
Sep 12th, 2023

Presenters: Chandni Valiathan, PhD – MBMA group lead, Clinical Pharmacology and Pharmacometrics
John Maringwa, PhD – Director, MBMA group, Clinical Pharmacology and Pharmacometrics
Lixia Pei, PhD – Director, Oncology, Statistics and Decision Sciences

Acknowledgements: Paulo Paneque Galuzio, Mehrdad Javidi, Natalia A. Quijano Cardé

janssen | PHARMACEUTICAL COMPANIES OF Johnson & Johnson

Motivation for surrogate endpoints

- A surrogate endpoint is a measure that relates well with a **clinical endpoint of interest (CEI)** in assessing the effect of a new treatment
- Surrogate endpoints, if adequately established, can be measured earlier or with less variability, potentially helping to improve trial efficiency in terms of duration and cost
- Leveraging external aggregate-level data from multiple studies and compounds may be useful for surrogate marker exploration
  - *Consistency of the relationship between effect of treatment on the surrogate and the CEI across different situations may serve to increase one's confidence on the surrogate endpoint*

A great opportunity for cross-function collaboration

- Close collaboration between statistics and clinical pharmacology and pharmacometrics (MBMA group) – combined need for model development, data exploration and statistical interpretation
- From identification of question from the clinical team to delivering results, close communication and interaction between groups for data identification and discussion of results
- Numerous examples of close collaborations for decision making

Chandni Valiathan and John Maringwa - MBMA group, Clinical Pharmacology and Pharmacometrics
Lixia Pei - Statistics and Decision Sciences

For today's workshop …

- Quick review of available methods for surrogate marker selection
- Present a motivating example with aggregate-level and individual-level data by exploring challenges with currently available methods
- Introduce the Decision-Equivalence Metric (DEM) method
- Apply the DEM method to the motivating example to compare and contrast
- Summary and discussion

Traditional approaches to surrogate marker assessment

Summary of select set of common methods to assess surrogate markers

- Prentice's criteria-
  - *Set of equations/criteria that check that rejecting the null hypothesis (no treatment effect) on the surrogate endpoint implies rejecting the null hypothesis on the clinical endpoint of interest (CEI)*
- Freedman's proportion explained
  - *extending Prentice's criteria, calculate the proportion of the treatment effect explained by the surrogate marker*
- Li's proportion of treatment effect
  - *with fewer constraints than Prentice's criteria, calculate the contribution of the surrogate marker to the treatment effect as a proportion using fraction of weighted coefficients (measures the violation of Prentice's most stringent criterion)*
- R^2
  - *a measure of variance of CEI captured by the surrogate endpoint*

Prentice, R. L. (1989). Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in Medicine*, *8*(4), 431–440.

Freedman, L. S., Graubard, B. I., & Schatzkin, A. (1992). Statistical validation of intermediate endpoints for chronic diseases. *Statistics in Medicine*, *11*(2), 167–178.

Li, Z., Meredith, M. P., & Hoseyni, M. S. (2001). A method to assess the proportion of treatment effect explained by a surrogate endpoint. *Statistics in Medicine*, *20*(21), 3175–3188.

Summary of select set of common methods to assess surrogate markers - details

Prentice criteria

1. Treatment (Z) has a significant impact on the surrogate endpoint (S)
   α differs significantly from zero in $S_j = \mu_s + \alpha Z_j + \epsilon_{sj}$
2. Treatment (Z) has a significant impact on the Clinical endpoint of interest (CEI)
   β differs significantly from zero in $CEI_j = \mu_T + \beta Z_j + \epsilon_{CEIj}$
3. The surrogate endpoint (S) has a significant impact on the clinical endpoint of interest (CEI)
   γ differs significantly from zero in $CEI_j = \mu + \gamma S_j + \epsilon_j$
4. The full effect of treatment (Z) upon the clinical endpoint of interest (CEI) is captured by the surrogate (S)
   β_s should be "equal" to zero in $CEI_j = \tilde{\mu}_T + \beta_s Z_j + \gamma_Z S_j + \tilde{\epsilon}_{CEIj}$

Freedman Proportion Explained

$$PE = 1 - \frac{\beta_s}{\beta}$$

Proportion of Treatment effect

$$CEI \sim \boldsymbol{\beta_{20} + \beta_{21} Z + \beta_{22} S};$$

$$PTE = \frac{wt * \widehat{\beta_{22}}}{\widehat{\beta_{21}} + wt * \widehat{\beta_{22}}}$$

where $wt = \frac{\sum_{i=1}^{N} S_i * n_i}{\sum_{i=1}^{N} n_i}$

Prentice, R. L. (1989). Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in Medicine*, *8*(4), 431–440.
Freedman, L. S., Graubard, B. I., & Schatzkin, A. (1992). Statistical validation of intermediate endpoints for chronic diseases. *Statistics in Medicine*, *11*(2), 167–178.
Li, Z., Meredith, M. P., & Hoseyni, M. S. (2001). A method to assess the proportion of treatment effect explained by a surrogate endpoint. *Statistics in Medicine*, *20*(21), 3175–3188.

Motivating example

Diabetes: FPG is an accepted surrogate for HbA1c in clinical practice

Clinical Endpoint of Interest (CEI)

Hemoglobin A1c (HbA1c) (%)

is a validated clinical endpoint used as the basis for approval of drugs intended to treat diabetes mellitus, and validated to represent reduction of microvascular complications associated with diabetes mellitus.

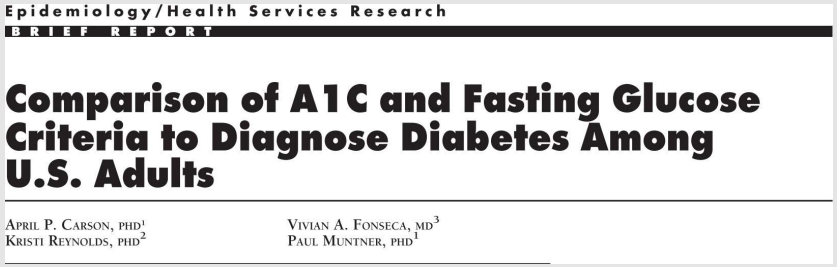
(3 months for HbA1c changes to stabilize)

Surrogate endpoint

Fasting plasma glucose (FPG):

The fasting plasma glucose (FPG) (mg/dl) test, measures the levels of glucose (sugar) in the blood upon an overnight fast. Used to screen for diabetes, it is a relatively simple, accurate, and inexpensive test that exposes problems with insulin functioning.

(4 weeks for FPG changes to stabilize)



Epidemiology/Health Services Research

BRIEF REPORT

Comparison of A1C and Fasting Glucose Criteria to Diagnose Diabetes Among U.S. Adults

April P. Carson, PhD[1]
Kristi Reynolds, PhD[2]
Vivian A. Fonseca, MD[3]
Paul Muntner, PhD[1]

Diabetes Care 33:95–97, 2010

- "6,890 participants without self-reported diabetes"
- "The results of the current study indicate the new recommendation by the International Expert Committee to use A1C to diagnose diabetes would result in the **same classification** as fasting glucose for **97.7% of U.S. adults**."

Aggregate-level data and individual-level data were used for exploration of methods

Aggregate-level data:

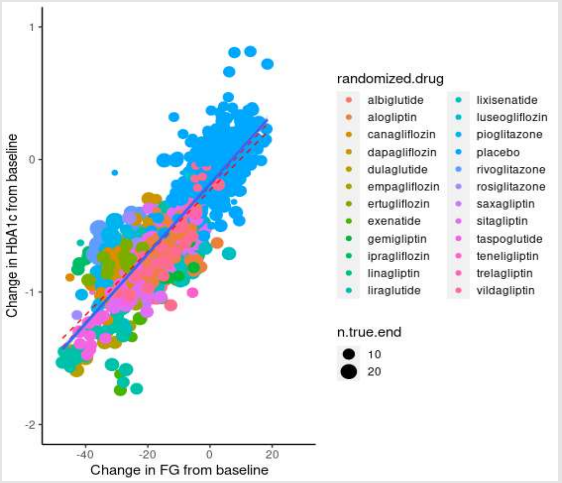
- Database created from systematic literature review for randomized, placebo controlled, phase 3 clinical trials in Type 2 diabetes investigating DPP4, GLP1, SGLT2 and TZD compounds
- Studies that reported both HbA1c and FPG at similar timepoint post 12 week (steady state) were included in the analysis
- Final analysis set: 167 trials for 24 treatments including placebo with total population of 46093 and 270 study arms.

Individual-level data:

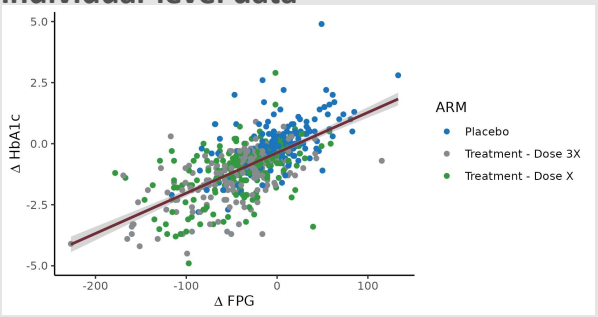
- Data from a randomized placebo-controlled phase 3 study in Type 2 Diabetes Mellitus studying 2 doses of drug for 26 weeks of treatment
- subjects randomized in 1:1:1 ratio
- Primary endpoint was HbA1c. FPG measures also available (data post 12 weeks)

Applying traditional methods for surrogate marker assessment shows poor performance of FPG as a surrogate marker for HbA1c (change from baseline)

Aggregate-level data



Individual-level data



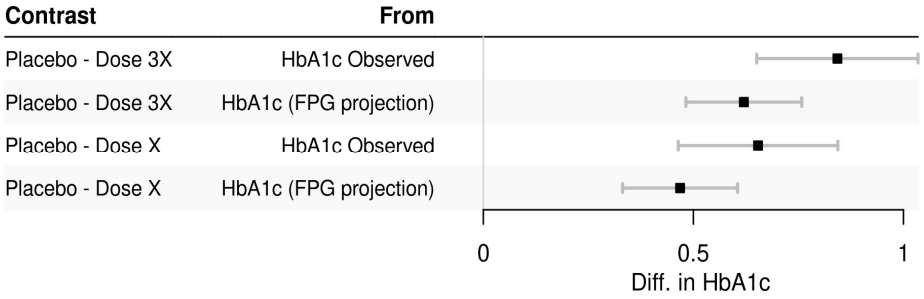
Evaluation criteria/parameters	Aggregate level data	Individual level data
Prentice 1st criterion: *treatment has significant impact on surrogate endpoint*		
Prentice 2nd criterion *: treatment has significant impact on true endpoint*		
Prentice 3rd criterion : *surrogate endpoint has significant impact on true endpoint*		
Prentice 4th criterion : *the full effect of treatment upon true endpoint is captured by the surrogate*		
Freedman Proportion Explained (PE)		
Proportion of treatment effect, Li method (PTE)		
R-sq		

Prentice, R. L. (1989). Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in Medicine*, *8*(4), 431–440.
Freedman, L. S., Graubard, B. I., & Schatzkin, A. (1992). Statistical validation of intermediate endpoints for chronic diseases. *Statistics in Medicine*, *11*(2), 167–178.
Li, Z., Meredith, M. P., & Hoseyni, M. S. (2001). A method to assess the proportion of treatment effect explained by a surrogate endpoint. *Statistics in Medicine*, *20*(21), 3175–3188.

The Decision-Equivalence Metric (DEM) approach

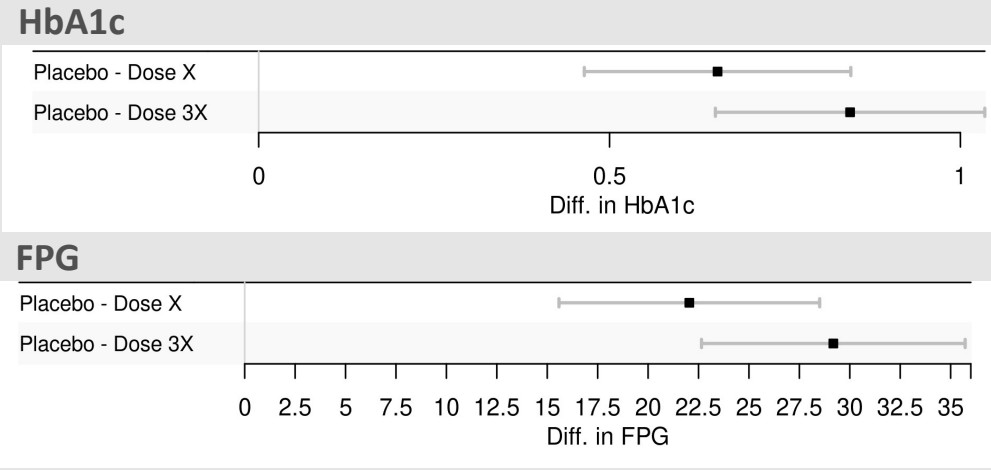
Exploring surrogate and CEI relationship for decision making

Similar treatment effect trend observed using HbA1c raw values (1st and 3rd rows) and HbA1c predicted from FPG (2nd and 4th rows)

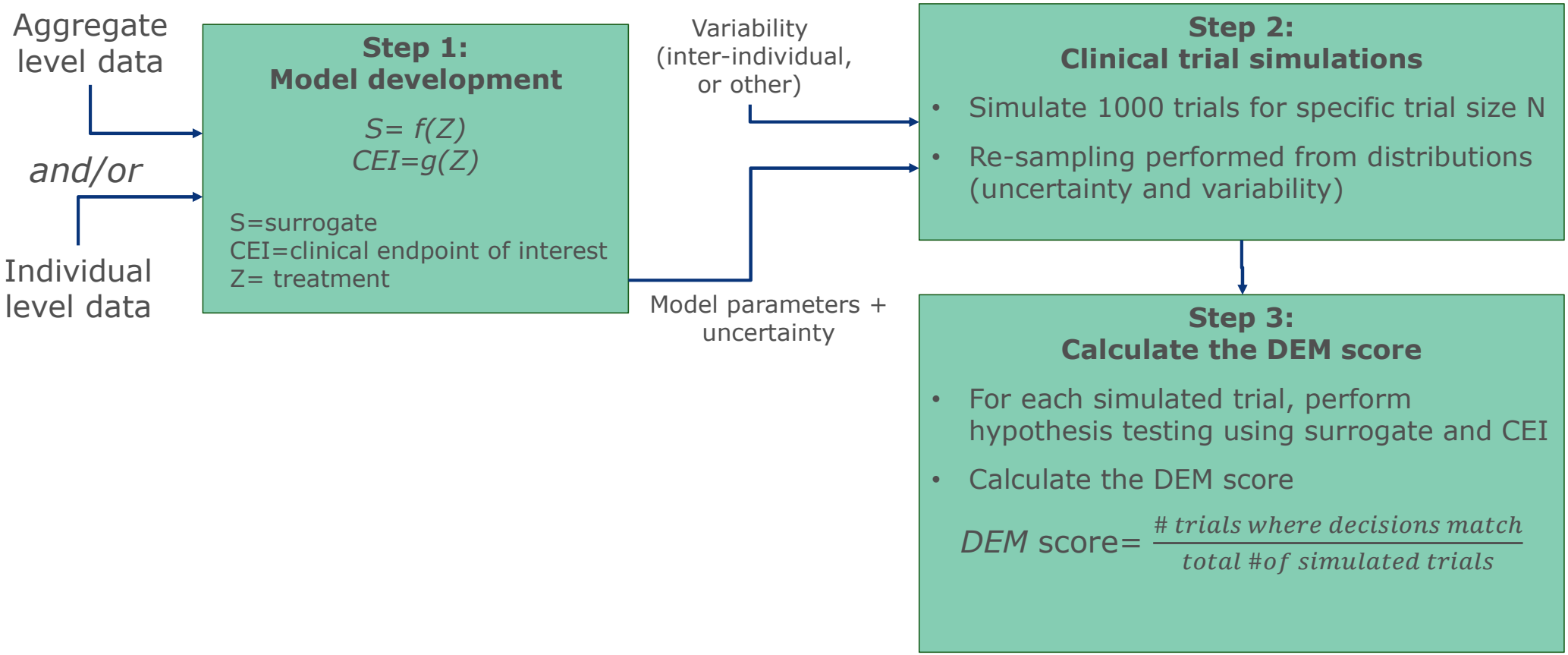


Explored decision making/hypothesis testing with surrogate endpoint rather than CEI

Similar conclusions of rejecting the null hypothesis would be achieved with both surrogate and CEI



The Decision-Equivalence Metric (DEM) Method



Applying the DEM method on the diabetes dataset yields a high DEM score validating FPG as a clinical surrogate marker

Aggregate level data

Step 1:
Model development

MBMA model developed across treatments and trials

$fpg_{avg} = f(Z);\ \ linear\ model$
$HbA1c_{avg} = g(Z); linear\ model$
Z ϵ {placebo, treatment}

$fpg_{ILD} \sim N(\widehat{fpg_{avg}}, \sigma_{fpg})$
$HbA1c_{ILD} \sim N(\widehat{HbA1c_{avg}}, \sigma_{HbA1c})$

Model parameters

Step 2:
Clinical trial simulations

- Simulate 1000 trials for specific trial size N=115 (median N of trials)
- Re-sampling performed from distributions (uncertainty and variability)

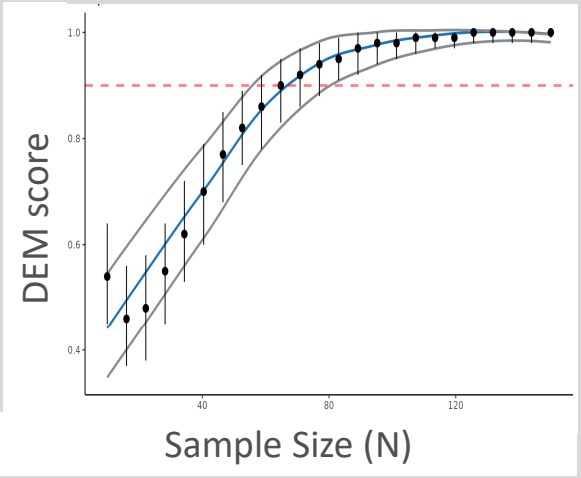
Step 3:
Calculate the DEM score

- Perform hypothesis testing on FPG and HbA1c values for each simulated trial

$$DEM\ score = \frac{\#\ trials\ where\ decisions\ match}{total\ \#of\ simulated\ trials}$$

$$DEM\ score = 0.988\ [0.925, 1.000]$$

The DEM method can also help identify the appropriate sample size for a surrogate marker



The DEM method provides a practical and general way of assessing and establishing a surrogate endpoint

Evaluation criteria/parameters	Aggregate level data
Prentice 1st criterion: *treatment has significant impact on surrogate endpoint*	Yes
Prentice 2nd criterion *: treatment has significant impact on true endpoint*	Yes
Prentice 3rd criterion : *surrogate endpoint has significant impact on true endpoint*	Yes
Prentice 4th criterion : *the full effect of treatment upon true endpoint is captured by the surrogate*	No
Freedman Proportion Explained (PE)	0.54
Proportion of treatment effect, Li method (PTE)	0.59
R-sq	0.71
DEM Score	0.98 (N=115)

Traditional methods fail to identify FPG as a surrogate to HbA1c

The DEM Method is general enough to include:

- most model types to describe surrogate or CEI
- different types of endpoints:
  - Continuous e.g., FPG
  - Binary e.g., ACR20
  - other types of endpoints are being explored
- aggregate-level or individual-level data

Summary, conclusions and discussion

Summary, conclusions and discussion

- Reviewed common methods currently used to assess surrogate markers for clinical endpoints of interest
- Described the DEM method that provides a practical framework to assess surrogate markers in a clinically relevant manner
- Where **current methods fail** to identify FPG as a reasonable surrogate for HbA1c, the **DEM method yields results that align with clinical observations** of surrogacy
- Additionally the **appropriate trial size** for a high DEM score was identified which aligned with the median trial size for the studies in the diabetes dataset
- DEM approach can be **applied to other indications and types of endpoints** e.g binary endpoints
- The **DEM method** provides a **practical, clinically relevant and generally applicable framework** to identify and assess surrogate markers

Thanks! Questions?


janssen
PHARMACEUTICAL COMPANIES OF
Johnson & Johnson