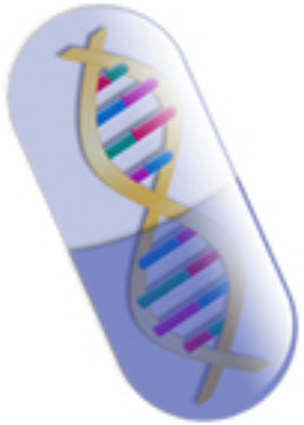


Privacy in Pharmacogenetics: An End-to-End Case Study of Personalized Warfarin Dosing

Presented By

Sharani Sankaran

Pharmacogenetics

Population Dataset

Learning Algorithm

$f(x)$

Trained Model

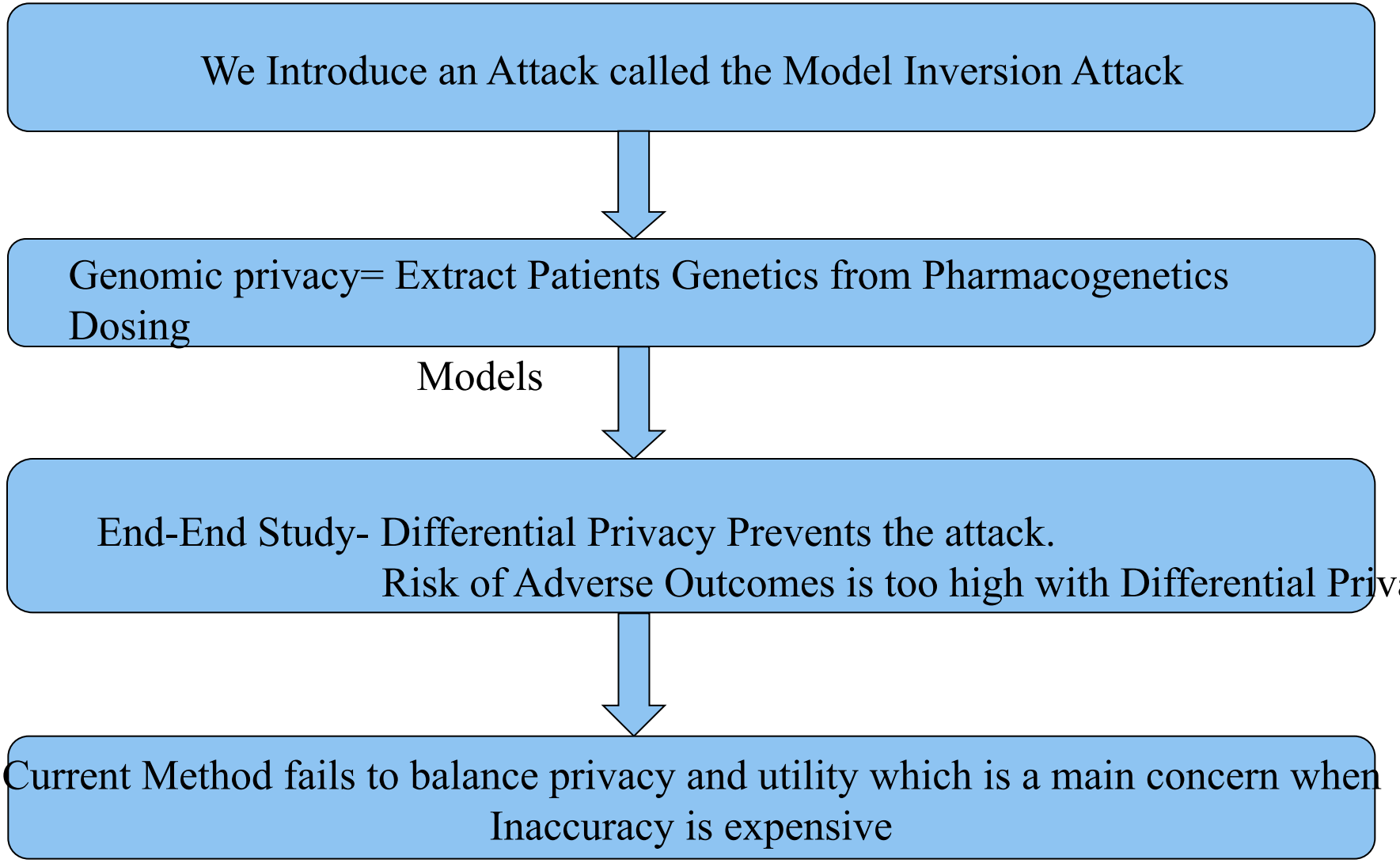
Genotype

Clinical Variables

$f(x)$

Trained Model

Medical Guidance


We Introduce an Attack called the Model Inversion Attack
Genomic privacy= Extract Patients Genetics from Pharmacogenetics Dosing
Models
End-End Study- Differential Privacy Prevents the attack.
Risk of Adverse Outcomes is too high with Differential Privacy
Current Method fails to balance privacy and utility which is a main concern when Inaccuracy is expensive

Warfarin Dosing

- Warfarin is the most popular anticoagulant drug in use today.
- Anticoagulants are used to prevent stroke and other clotting related incidents.
- Warfarin is one of the oldest and well studied targets in pahrmacogenetics.

- Warfarin is very difficult to prescribe doses for patients correctly.

- Low Dose
- Death

Embolism

Stroke

High Dose

Intracranial Bleeding

Death

Extracranial Bleeding

The IWPC Warfarin Model

Population Dataset → Learning Algorithm → Trained Model

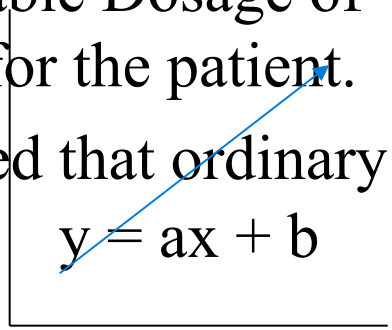
5700 patients from
21 sites in 6 countries, 4 continents

- Things Collected from each patient are
- Age
- Hieght

Patients Demographics,relevant parts of their medical history,cor orbidities,smoking status .

Independent variables

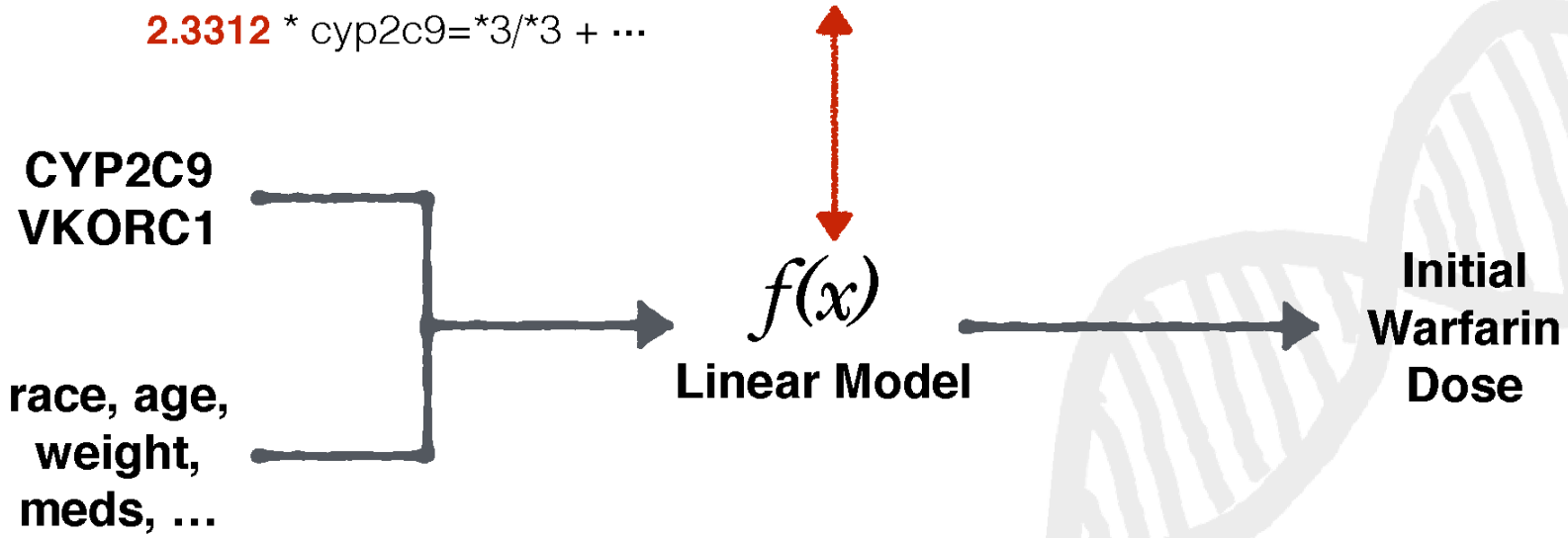
- weight
- Age
- Relevant Genotype : vkorc1,cyp2c9.
- These 2 aspects of their genotype that researchers previously found effect warfarin metabolism.
- Target outcome: Stable Dosage of Warfarin that achieved optimal therapeutic benefit for the patient.
- The IWPC confirmed that ordinary linear regression is the best learning algorithm



Pharmacogenetic Warfarin Dosing

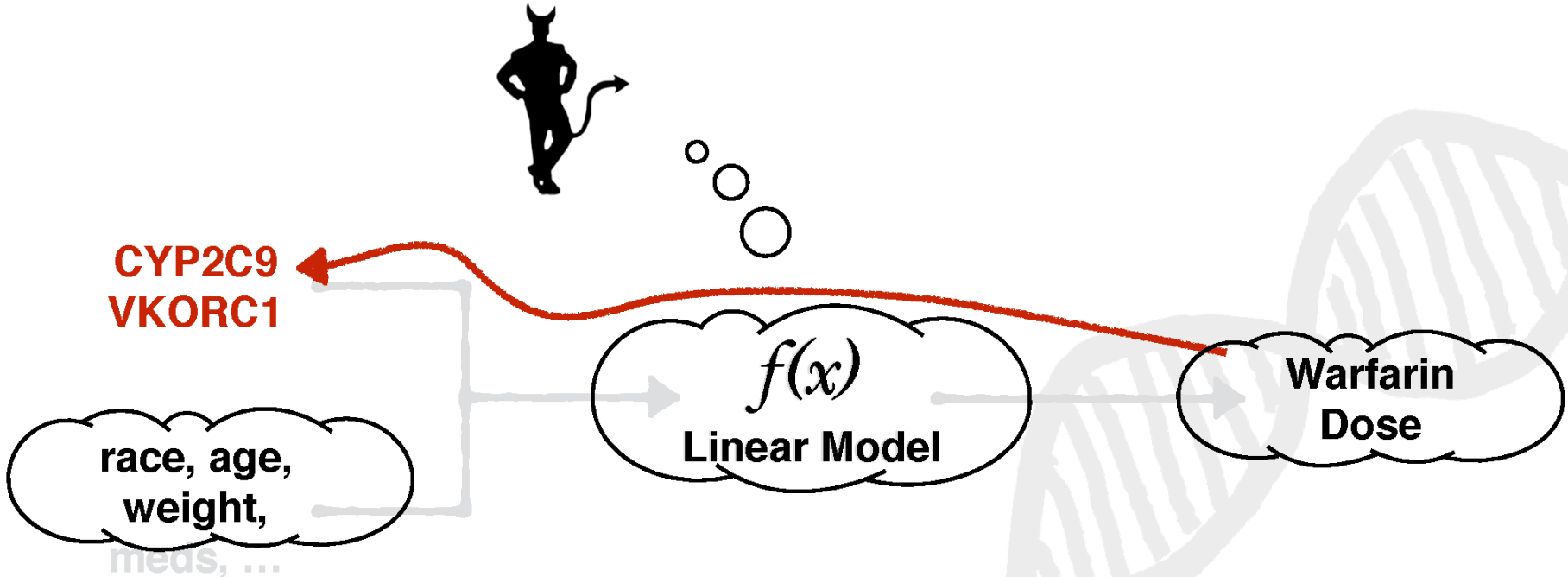
Population Dataset → Learning Algorithm → **Trained Model**

sqrt(dose) = **5.6044** + **0.2614** * age + **0.1092** * asian race - **0.2760** * black or african american - **0.8677** * vkorc1=A/G - **1.6974** * vkorc1=A/A - **1.9206** * cyp2c9=*2/*3 - **2.3312** * cyp2c9=*3/*3 + ···



Pharmacogenetic Privacy

age	height	weight	race	history	vkorc1	cyp2c9	dose
50-60	176.2	185.7	asian	cancer	A/G	*1/*3	42.0



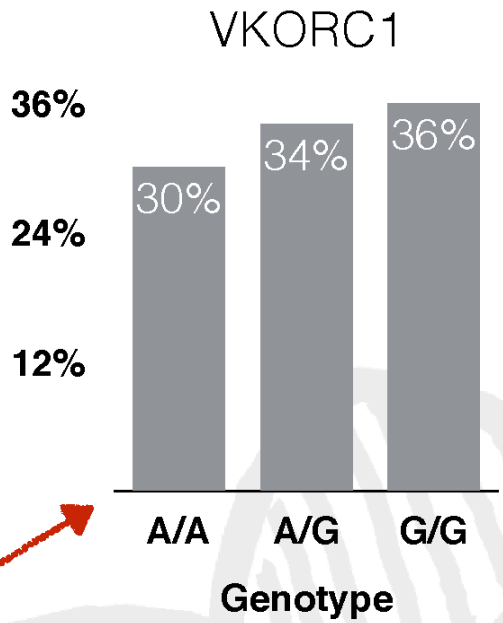
Model Inversion

basic demographics

stable warfarin dose

black-box access to model

marginal priors on patient distribution



... with better accuracy than the given "baseline" priors

Goal: infer the patient's genetic markers from this information

Our Model Inversion

1. Compute all values that agree with given information

$f(x)$

age	height	weight	race	history	vkorc1	cyp2c9	dose
50-59	176.53	144.2	white				42.0
50-59	176.53	144.2	white				42.0
50-59	176.53	144.2	white				42.0

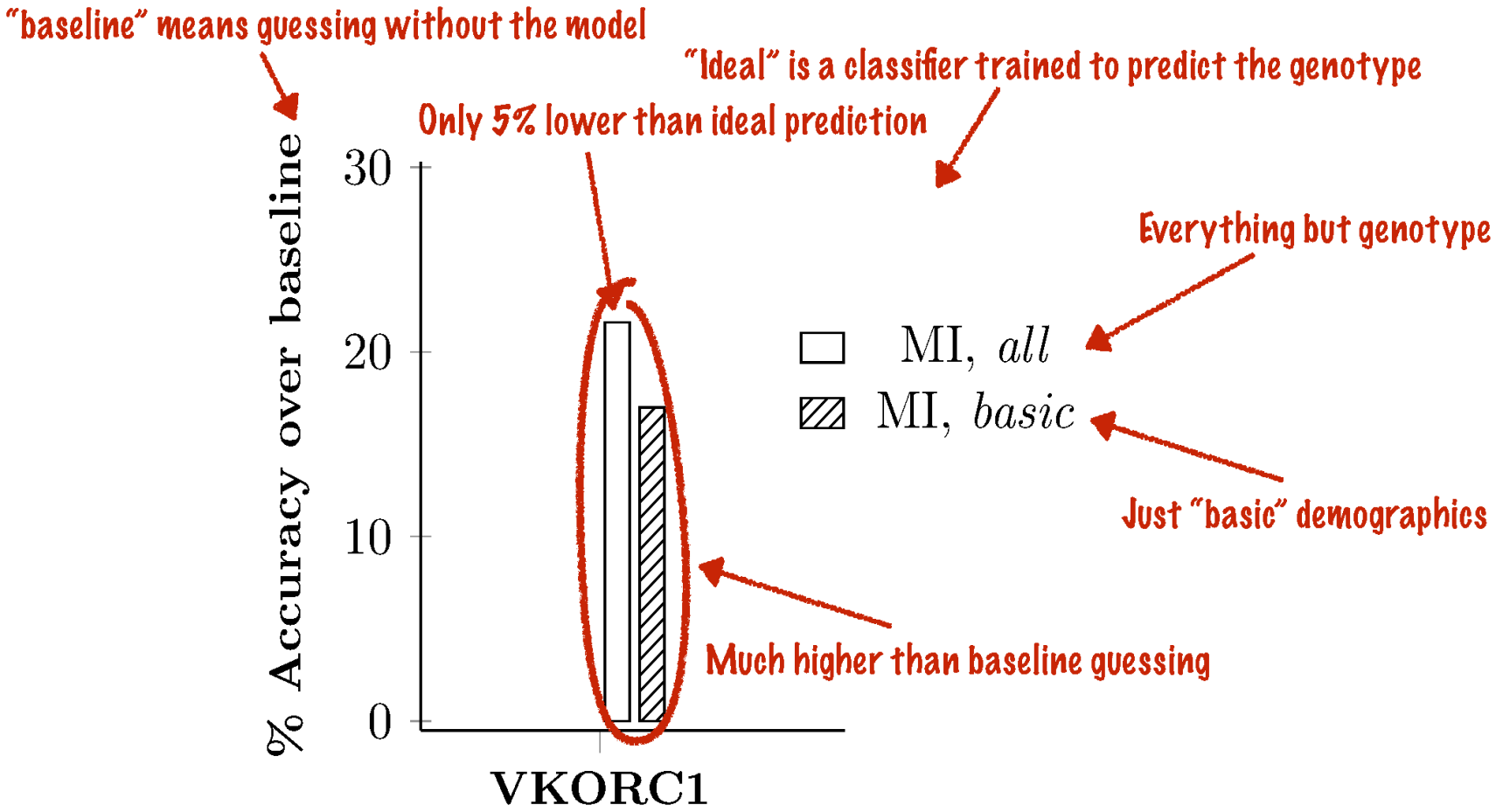
49.7	***p=0.23***
42.0	***p=0.75***
39.2	***p=0.01***

2. Find the most likely values among those that remain

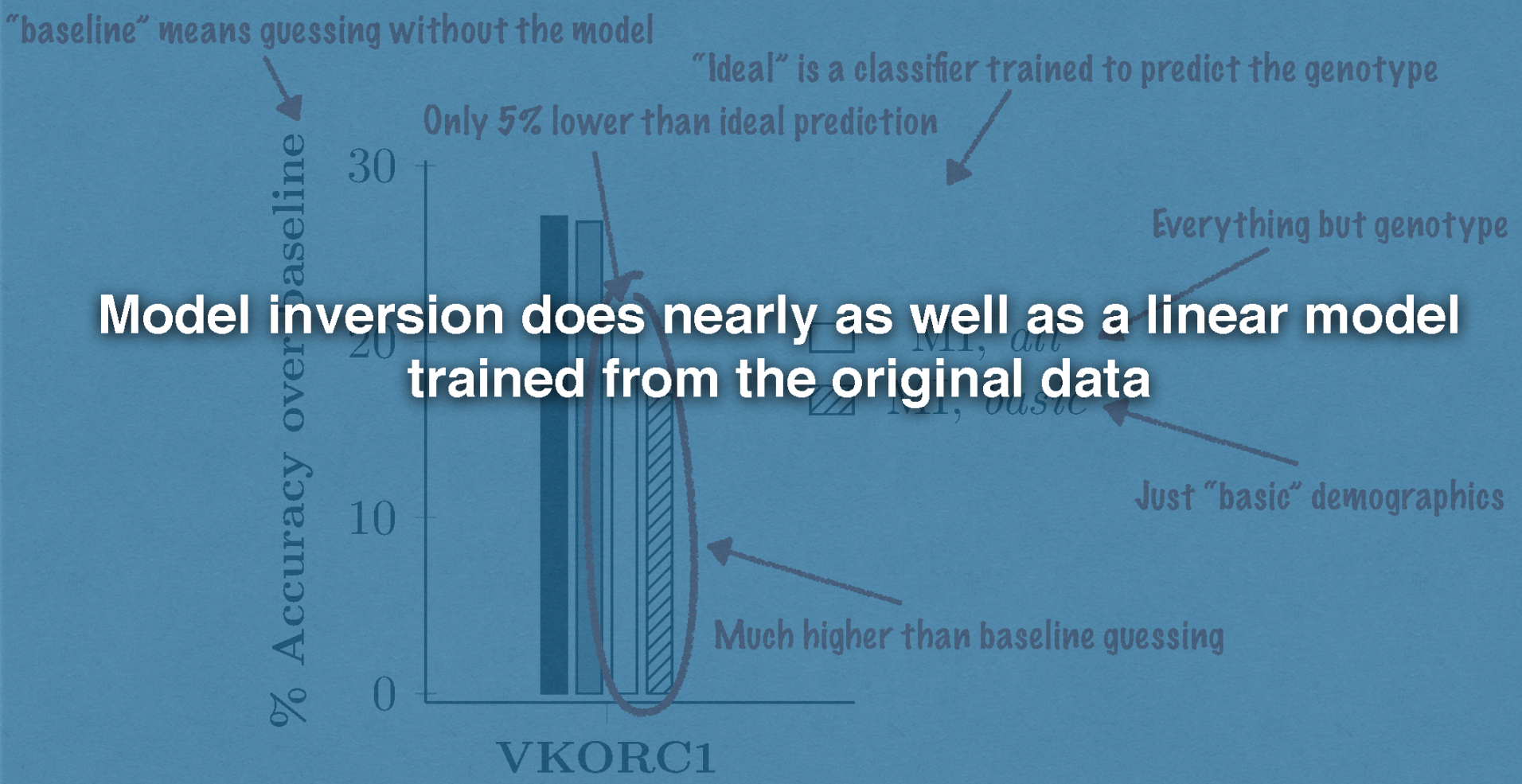
Use the marginal probabilities, model output to approximate this quantity

- The algorithm for computing the likelihood is optimal with the given information given that it minimizes the misprediction rate for these missing medical history ,genotypes

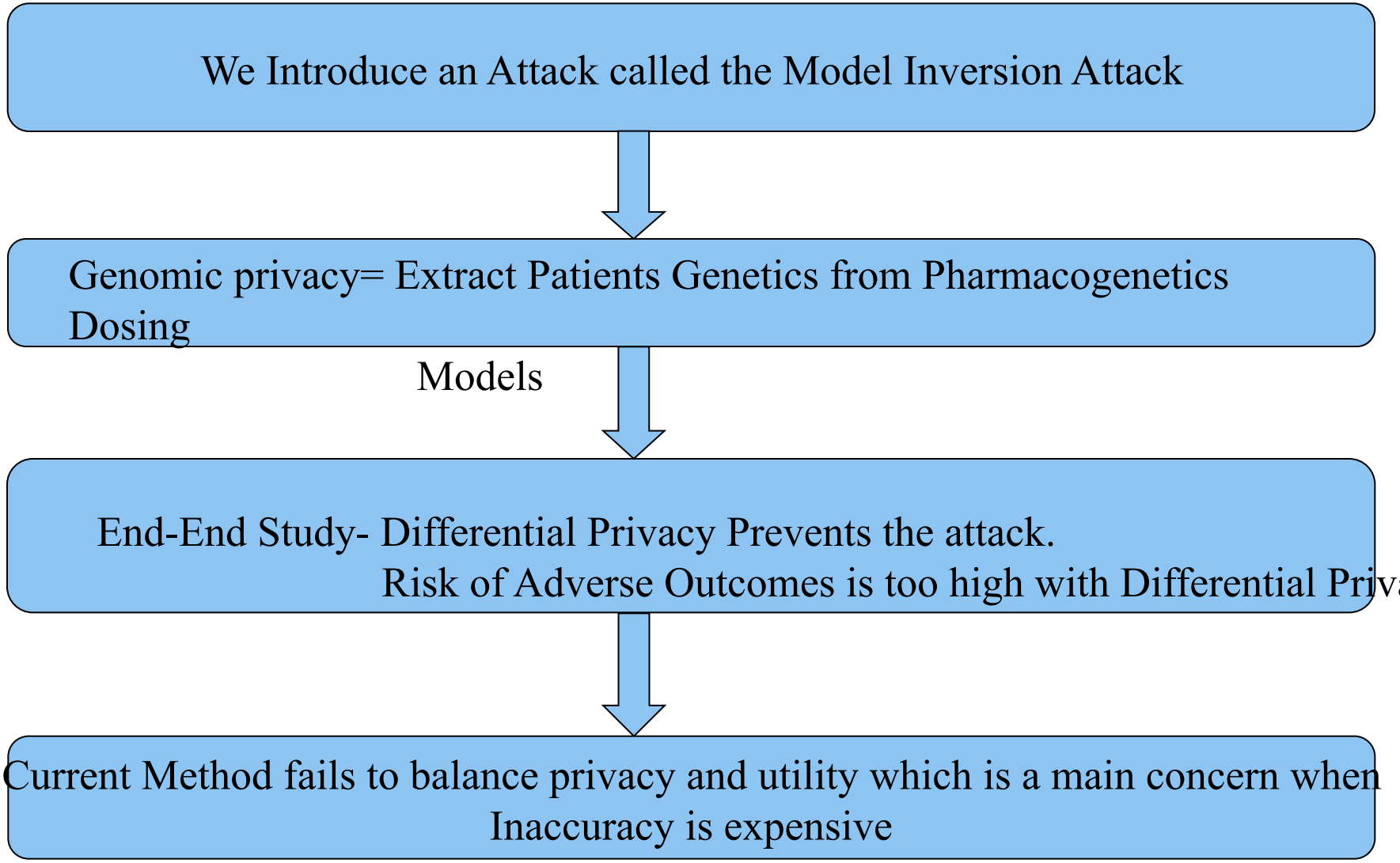
Results


"baseline" means guessing without the model
"Ideal" is a classifier trained to predict the genotype
Only 5% lower than ideal prediction
Everything but genotype
Just "basic" demographics
Much higher than baseline guessing
% Accuracy over baseline
30
20
10
0
MI, all
MI, basic
VKORC1

Results



Model inversion does nearly as well as a linear model trained from the original data


We Introduce an Attack called the Model Inversion Attack
Genomic privacy= Extract Patients Genetics from Pharmacogenetics Dosing
Models
End-End Study- Differential Privacy Prevents the attack.
Risk of Adverse Outcomes is too high with Differential Privacy
Current Method fails to balance privacy and utility which is a main concern when Inaccuracy is expensive

Differential Privacy

- Model Inversion is a problem so how to prevent it.
- We examine how to use differential privacy to prevent model inversion.
- A computation is differentially private if any output it produces going to be about as likely regardless of whether or not any particular individual row input to that computation.
- For D D' differing in one row
- Pr[K(D) = s] <=exp(e) *Pr[K(D')=s]
- Most Differential mechanism work by adding noise to their output in some capacity according to privacy budget
- There is also evidence of existing work that the attributes of virtual linear models are trained to be protected by adding the noise to the coefficients of those linear models.

Seeking a Remedy

Goal: see if a "reasonable" privacy budget solves the problem

End-to-End Study

Find budget that prevents model inversion

Evaluate risk of adverse events at these budgets

Private Linear Regression

[Zhang et al., VLDB 2012]

Private Histograms

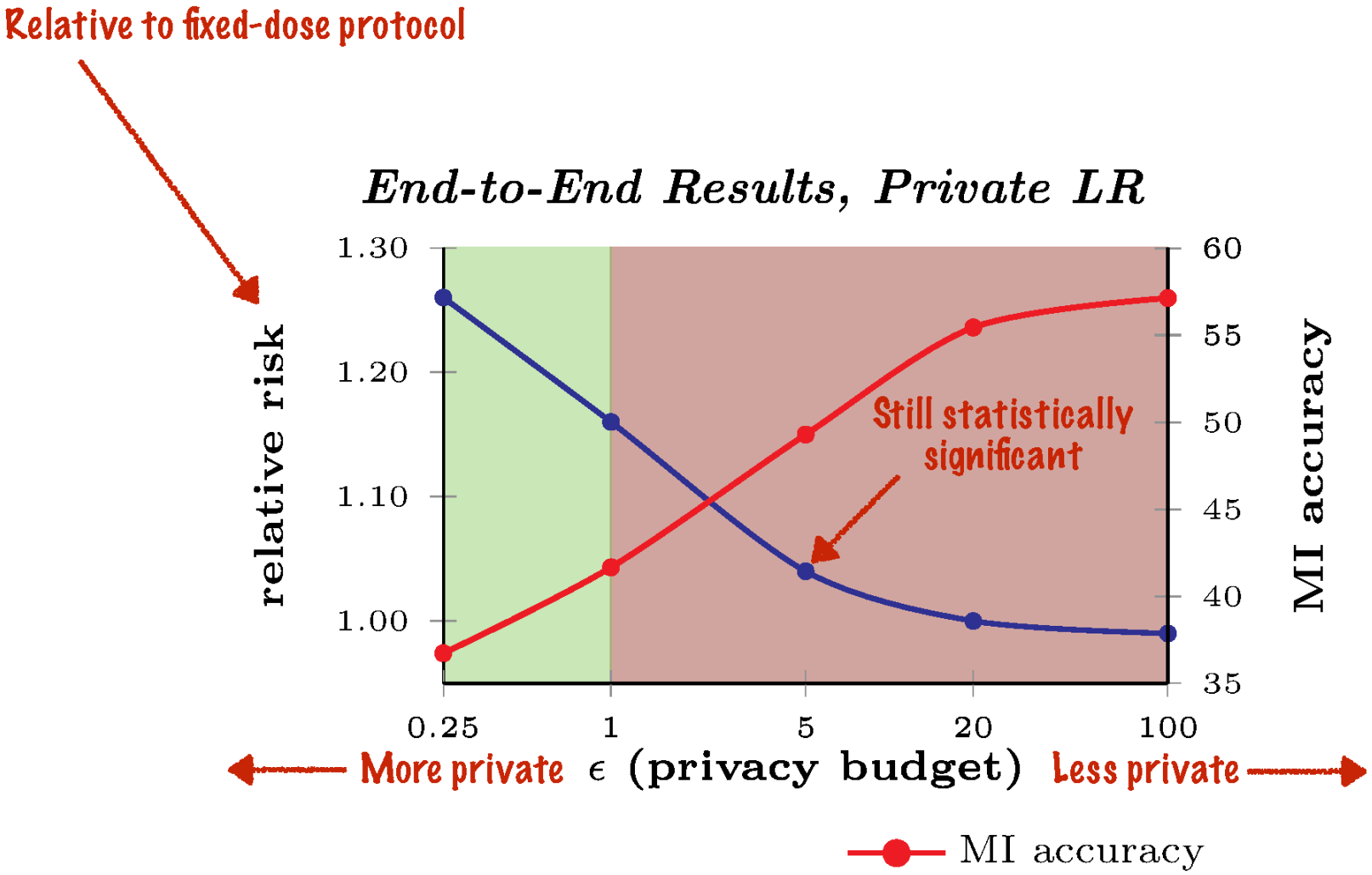
[Vinterbo, ECML-PKDD 2012]

Run model inversion experiments from before on DP models

Clinical Efficacy

Simulated Clinical Trials

Day 1	Days 1-2	Days 2-90	Days 3-90
Sample patient from IWPC validation set	standard fixed dose private model	Simulate body's response using PK/PD models (Hamberg et al., Clin. Pharm. Theory, 2007)	Defined in previous clinical trials


Relative to fixed-dose protocol
End-to-End Results, Private LR
relative risk
1.30
1.20
1.10
1.00
MI accuracy
60
55
50
45
40
35
Still statistically significant
0.25
1
5
20
100
More private
ϵ (privacy budget)
Less private
MI accuracy

Conclusion

- Current Method fails to balance privacy and utility which is main concern when Inaccuracy is expensive
- This paper did not observe that a privacy budget significantly prevented model inversion without introducing risk over fixed dosing.