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

Matthew Fredrikson, Eric Lantz,
Somesh Jha, David Page, Thomas Ristenpart

University of Wisconsin — Madison

Simon Lin, M.D.

Marshfield Clinic

This Talk

New Attack

We introduce an attack called *model inversion*

Genomic Privacy

Extract patients' genetics from *pharmacogenetic dosing models*

End-to-End Study

Differential privacy prevents the attack

Risk of adverse outcomes is too high with DP

Conclusion

Current methods fail to balance privacy and utility

This really matters when inaccuracy is expensive

Pharmacogenetics

Population Dataset

Learning Algorithm

$f(x)$

Trained Model

Genotype

Clinical Variables

Trained Model

Medical Guidance

Warfarin Dosing

Warfarin is the most popular anticoagulant in use today

Anticoagulants are used to prevent stroke and clotting-related incidents

Warfarin is one of the most well-studied targets in pharmacogenetics

100+ articles to date

The dangers of being wrong

Warfarin is notoriously difficult to dose correctly

Low dose ←

Death

Stroke

Embolism

High dose →

Intracranial bleeding

Extracranial bleeding

Death

ERANGGA
OL (AE)

: LRMP. R1/8012
h: 450 g (600 ml)
SEBELUM GUNA

lethrin (75/25) .... 0.11% w/w
rin .....0.07% w/w
.....99.82% w/w

r (M) Sdn. Bhd. (Co. No. 175141-H)
lan PJU 7/5, Mutiara Damansara,
Jaya, Selangor.
Faks:03-7719 1195

LAS IV

勿贮藏在靠近
食品或儿童
所接触的地方

KEEP
FOOD
CHILDREN

Jika berl
sila hubung
PUSAT RA
- 1800-88-
- 012-430

Tarikh

S IV

KE

KING KONG RAT KILLER

RACUN
TIKUS

殺鼠藥

UMPAN BERBUNGKAH (BB)

BACA LABEL SEBELUM GUNA

Kandungan Bersih :180 g

No. Pendaftaran. LRMP. R1/0020

Perawis Aktif: warfarin .....0.05% w/w
Perawis Lengai: .....99.95% w/w

Didaftarkan oleh:
TOHTONKU SDN. BHD.
186, Jalan Burma, (38311-A)
10350 Pulau Pinang.
Tel: 04-2287161

Jika berlaku **KERACUNAN**, sila hubungi: **PUSAT RACUN NEGARA**
-1-800-88-8099 (waktu pejabat) -012-4309499 (lepas waktu pejabat)

JAUHKAN DARIPADA
MAKANAN
DAN KANAK-KANAK

KELAS IV

勿貯藏在靠近
食品或兒童
所接觸的地方

The IWPC Warfarin Model

Population Dataset → Learning Algorithm → Trained Model

5700 patients from
21 sites in 6 countries, 4 continents

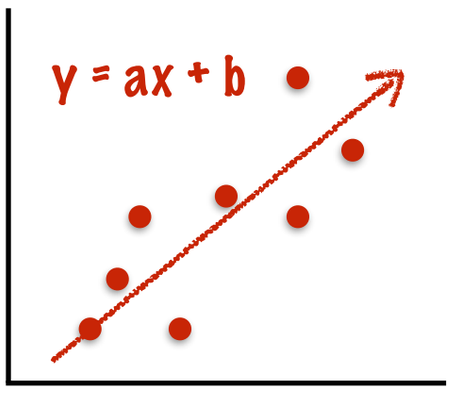
The IWPC Warfarin Model

Population Dataset → Learning Algorithm → Trained Model

age	height	weight	race	history	vkorc1	cyp2c9	dose

independent variables: patient demographics, medical history, comorbidities, smoking status, genotype

dependent variable (output): stable warfarin dose

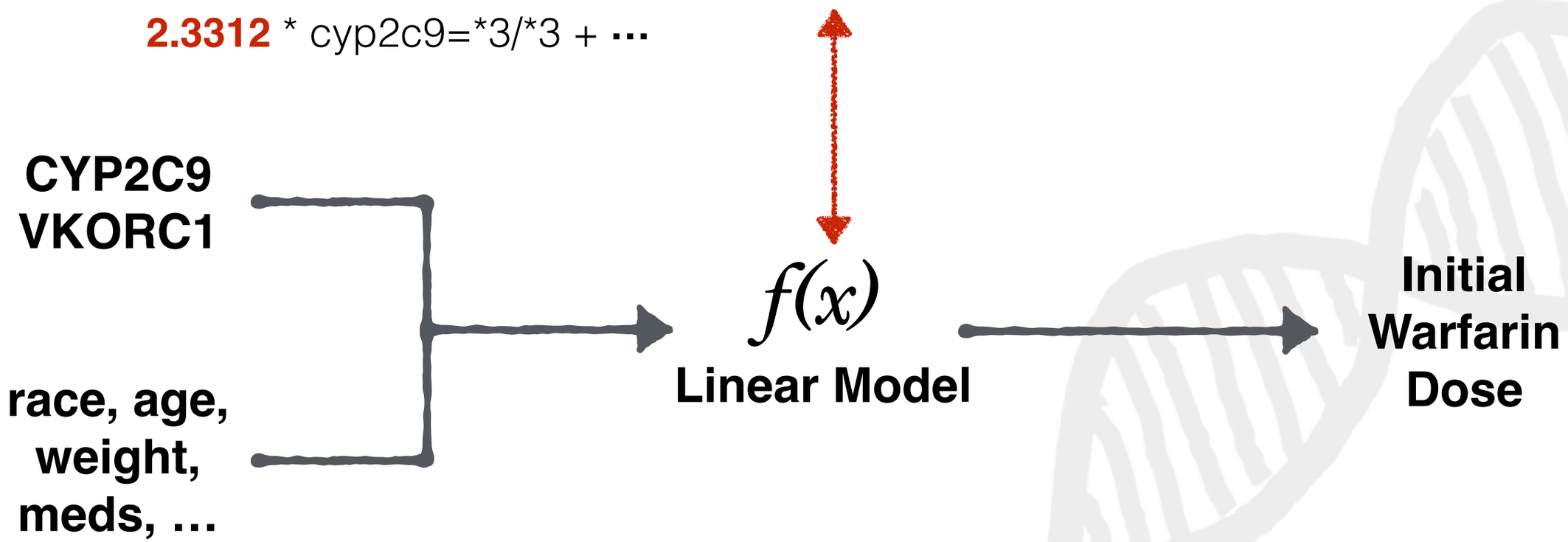


The IWPC found ordinary linear regression to be the best learning algorithm

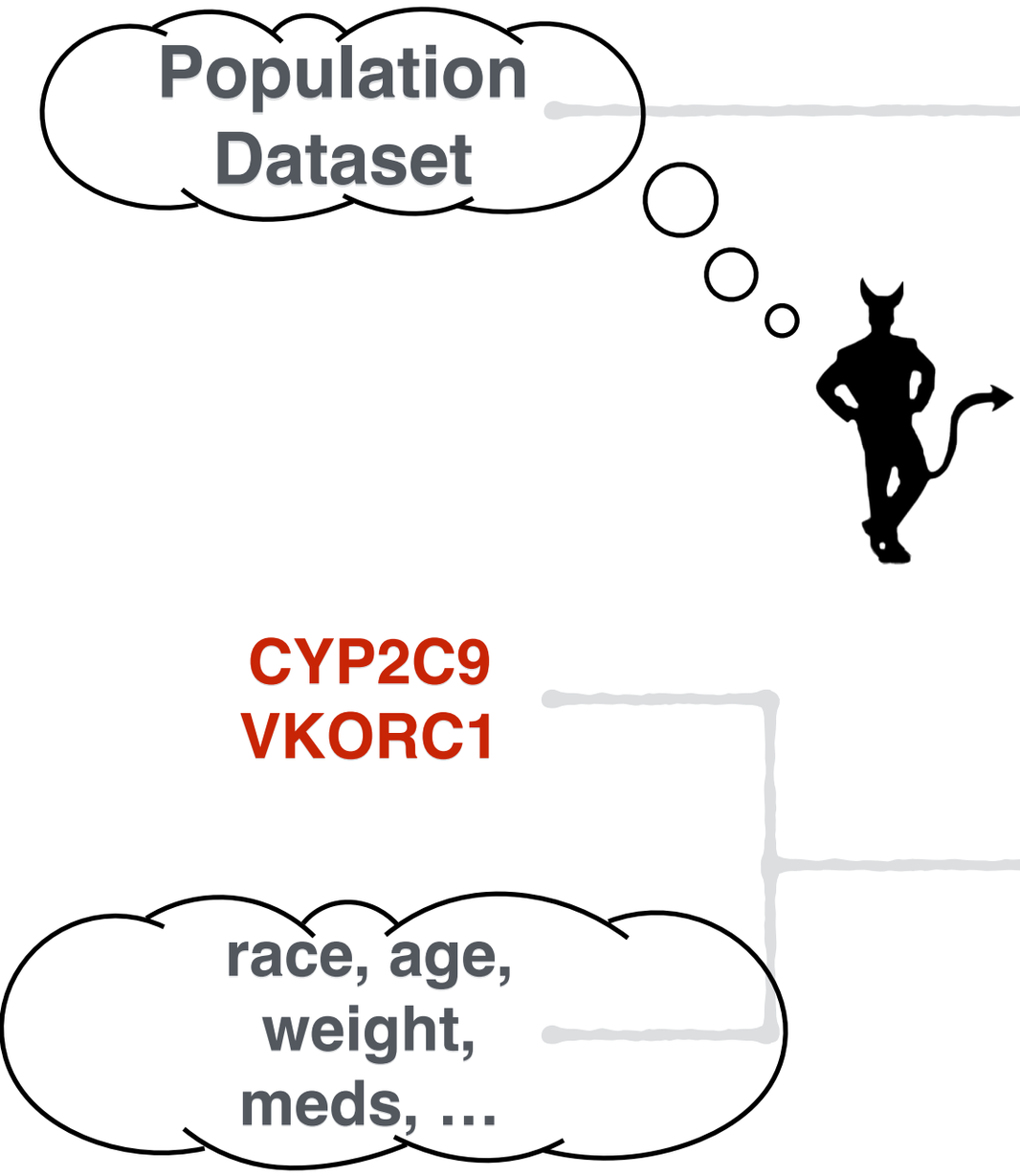
Pharmacogenetic Warfarin Dosing

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 + ⋯

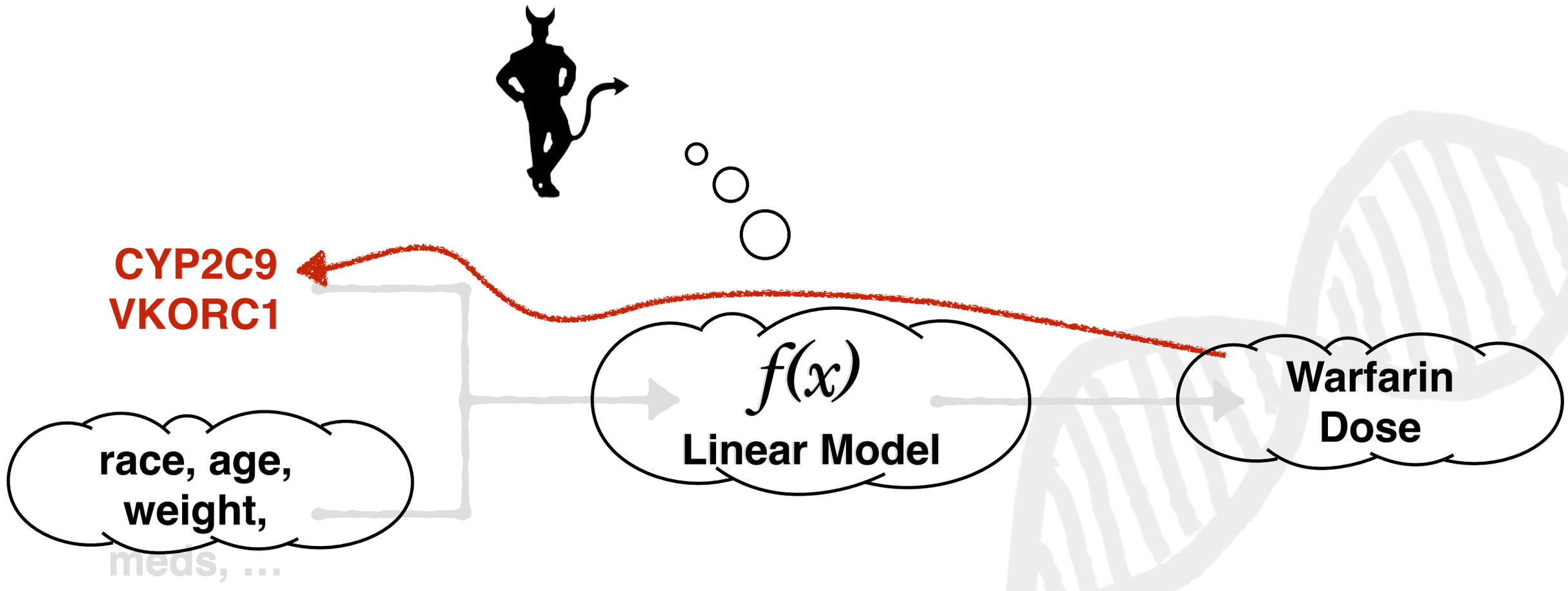


Pharmacogenetic Privacy



Pharmacogenetic Privacy

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



We call this attack *model inversion*

This Talk

New Attack

We introduce an attack called *model inversion*

Genomic Privacy

Extract patients' genetics from *pharmacogenetic dosing models*

End-to-End Study

Differential privacy prevents the attack

Risk of adverse outcomes is too high with DP

Conclusion

Current methods fail to balance privacy and utility

This really matters when inaccuracy is expensive

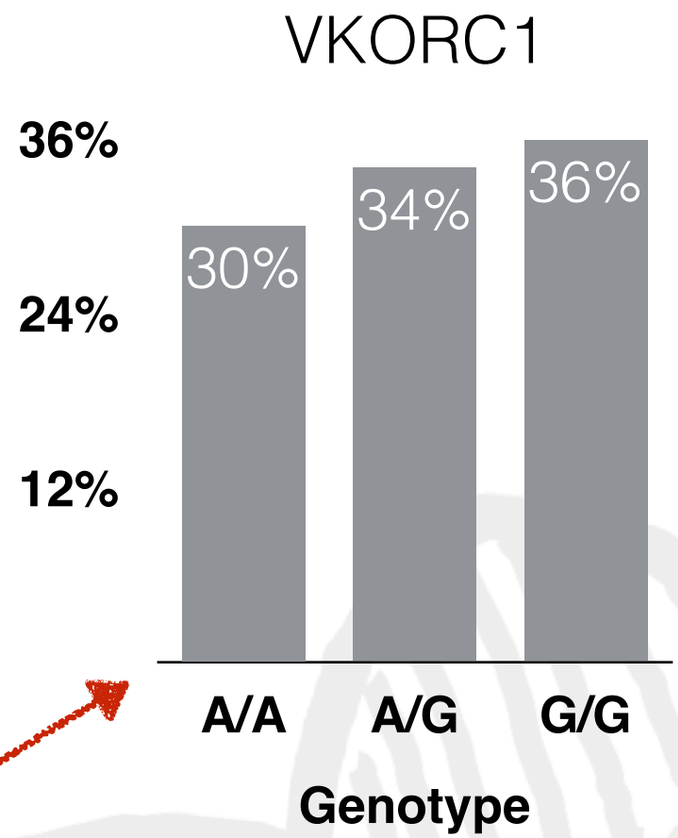
Model Inversion

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

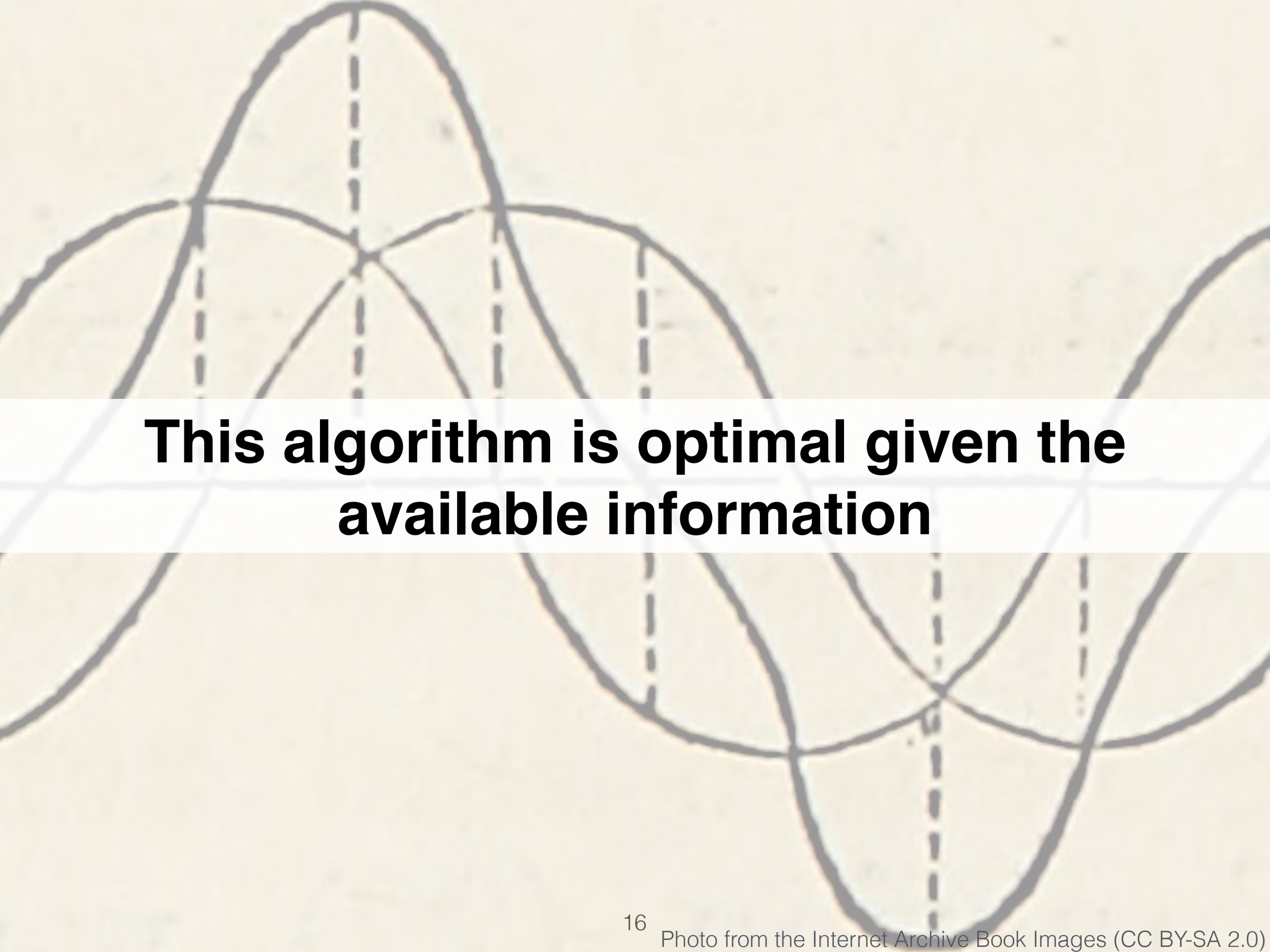
$f(x)$

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
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



This algorithm is optimal given the available information

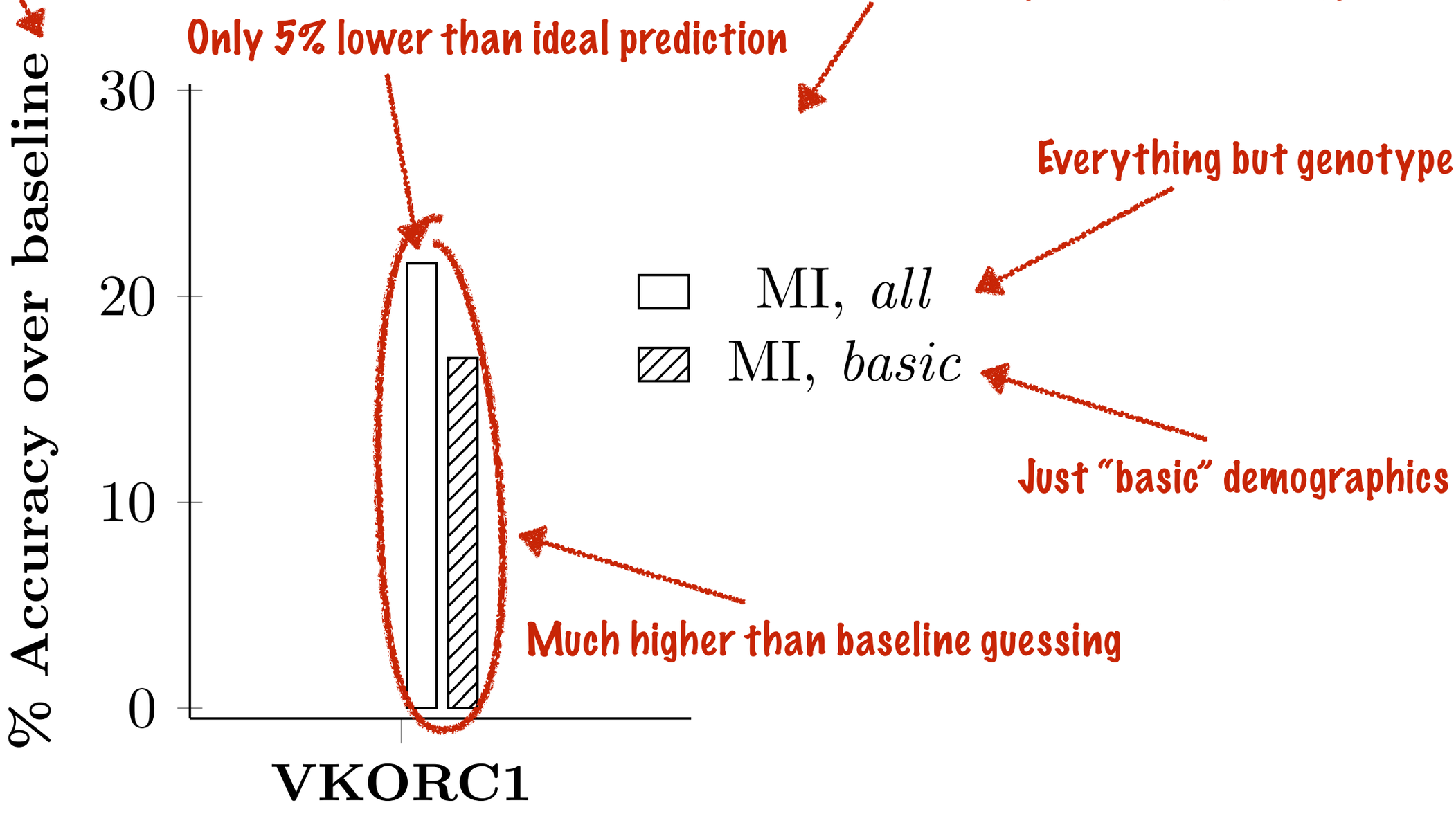
Results

"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

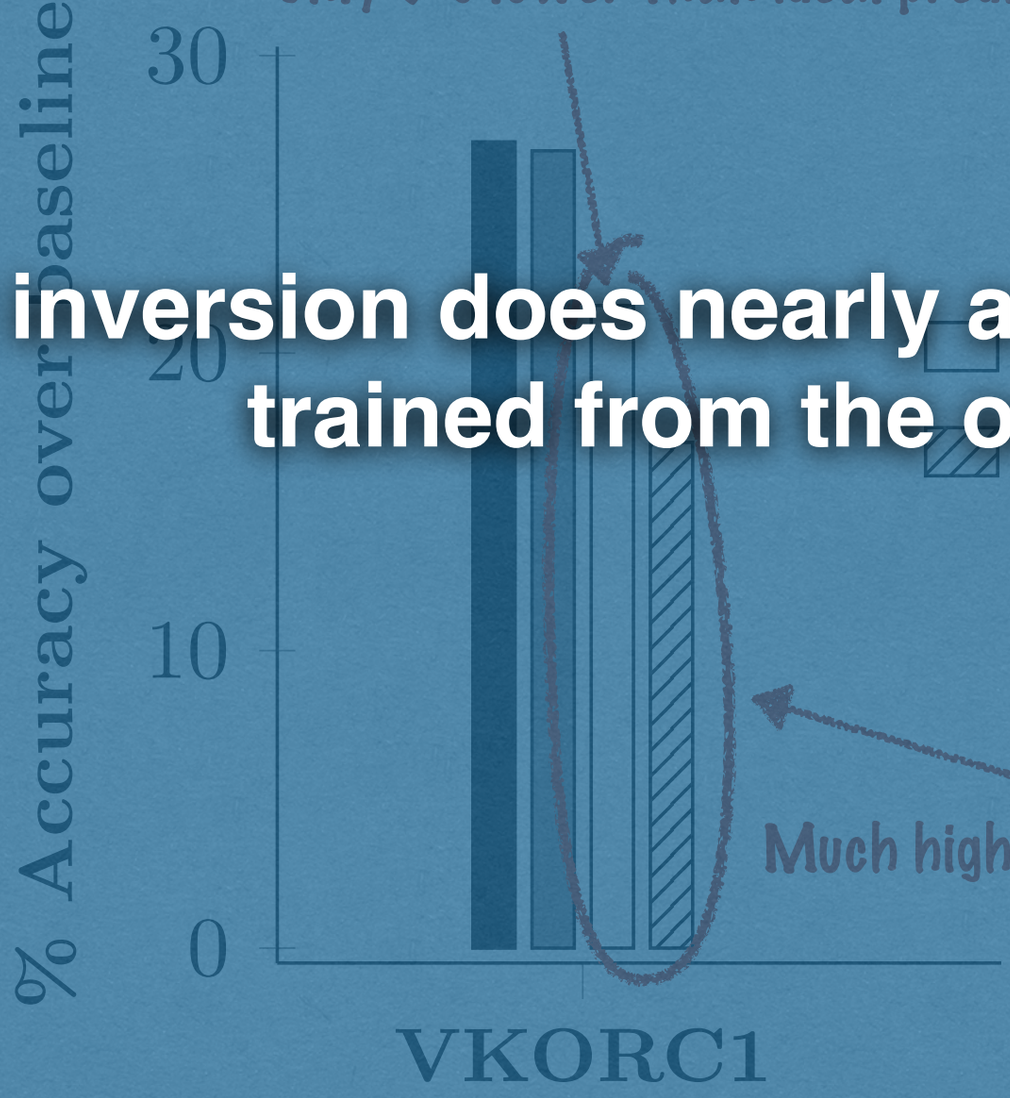
Results

"baseline" means guessing without the model

"Ideal" is a classifier trained to predict the genotype

Only 5% lower than ideal prediction

Everything but genotype



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

Just "basic" demographics

Much higher than baseline guessing

This Talk

New Attack

We introduce an attack called *model inversion*

Genomic Privacy

Extract patients' genetics from *dosing models*

End-to-End Study

Differential privacy prevents the attack

Risk of adverse outcomes is too high with DP

Conclusion

Current methods fail to balance privacy and utility

This really matters when inaccuracy is expensive

Seeking a Remedy

MI is a problem, so how can we prevent it?

We examine the use of differential privacy for preventing MI

For D, D' differing in one row,

Any output should be about as likely regardless of whether or not I am in the dataset

$$\Pr[K(D) = s] \leq \exp(\varepsilon) \times \Pr[K(D') = s]$$

Clean, provable guarantee

Most DP mechanisms "add noise" according to privacy budget

Evidence that this protects attributes in linear models
(Kasiviswanathan et al., SODA 2013)

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

End-to-End Study

Find budget that prevents model inversion

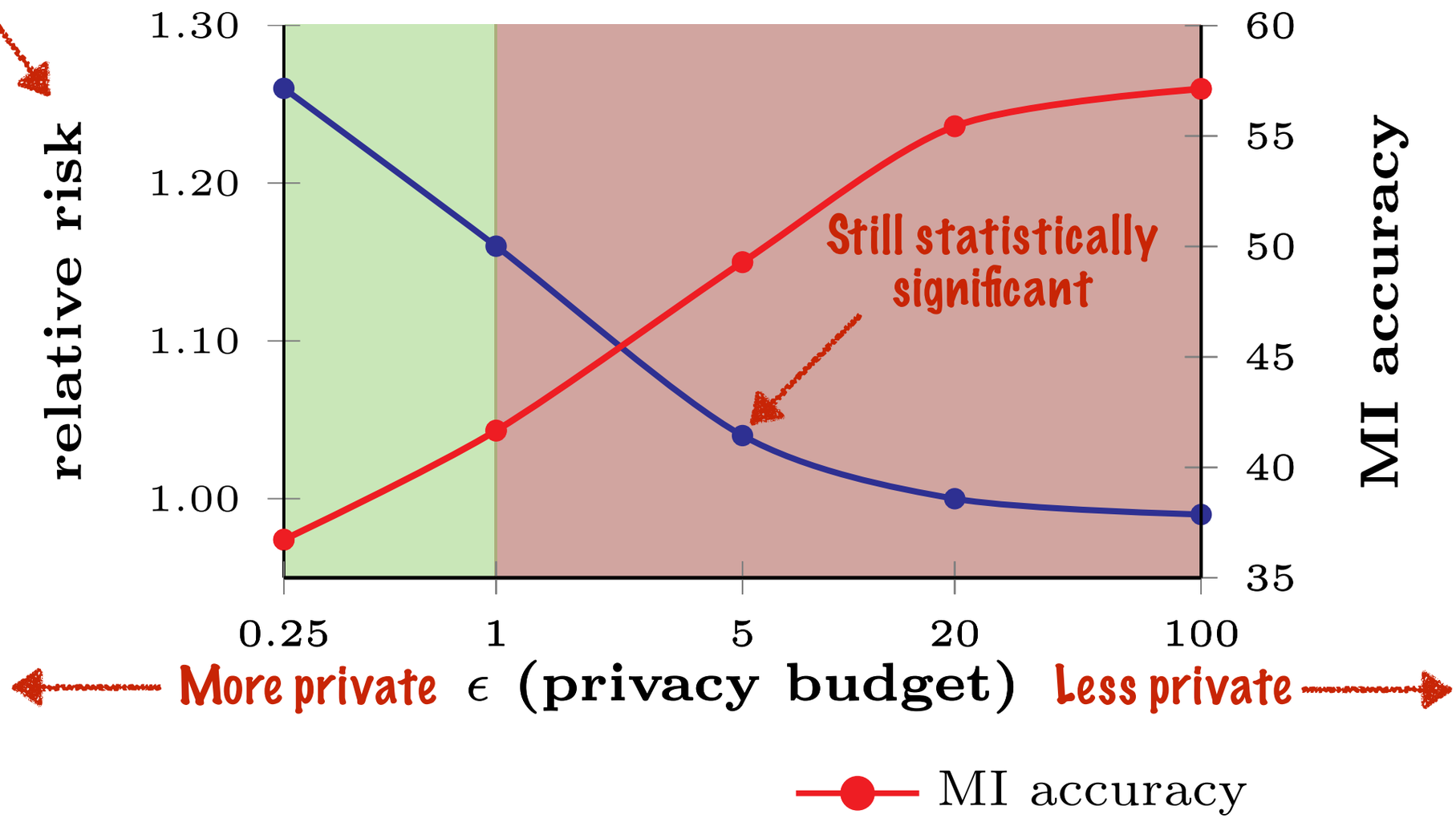
Evaluate risk of adverse events at these budgets

Simulate clinical trials to make this calculation

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



We did not observe a budget that significantly prevented model inversion, without introducing risk over fixed dosing.

Thanks!

New Attack

We introduce an attack called *model inversion*

Genomic Privacy

Extract patients' genetics from *pharmacogenetic dosing models*

End-to-End Study

Differential privacy prevents the attack

Risk of adverse outcomes is too high with DP

Conclusion

Current methods fail to balance privacy and utility

This really matters when inaccuracy is expensive