

Introduction

Rosa Gini

Head of Pharmacoepidemiology Unit, ARS Toscana
Vice President, VAC4EU

**Validate study variables to reduce
misclassification bias: recent tools and research needs**

HYBRID WORKSHOP

27 March 2025 - 14.30-18.30

ARS Toscana - Sala Rita Dioguardi, Villa La Quiete
via Pietro Dazzi 1 - Florence, Italy

Contents

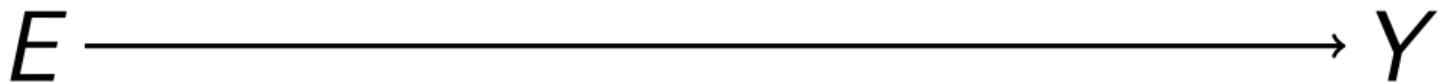
Introducing our heroes

Why is this important?

Understanding our heroes

What happens today?

The heroes of a (pharmaco)epi research question



The heros of our worskhop

$Y \longrightarrow M_Y$

Contents

Introducing our heroes

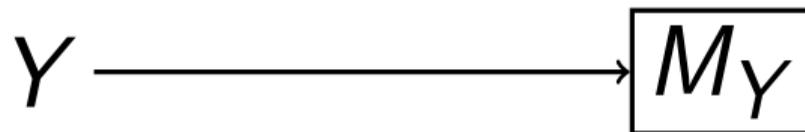
Why is this important?

Understanding our heroes

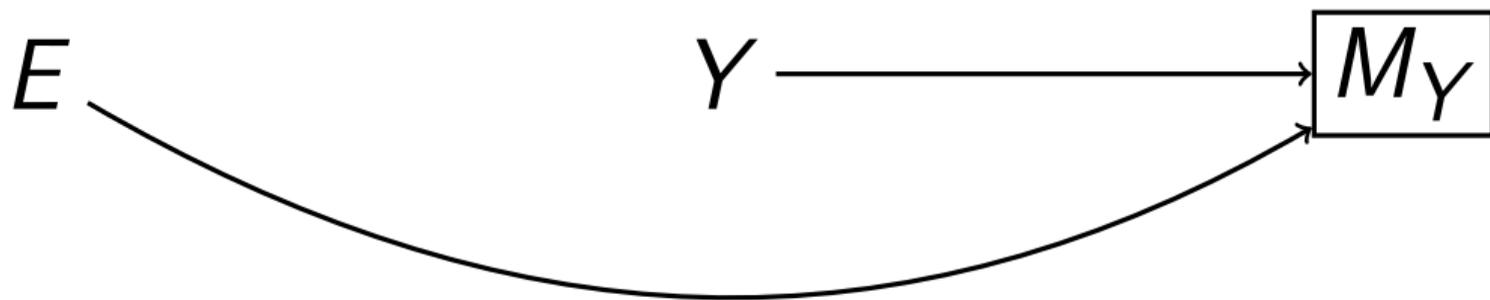
What happens today?

Biasing causal relations

E

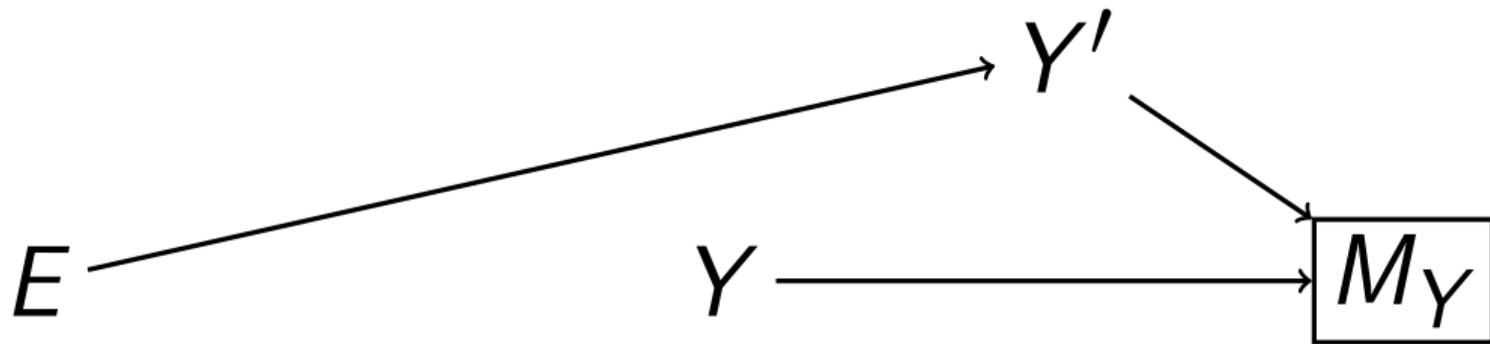


Biasing causal relations



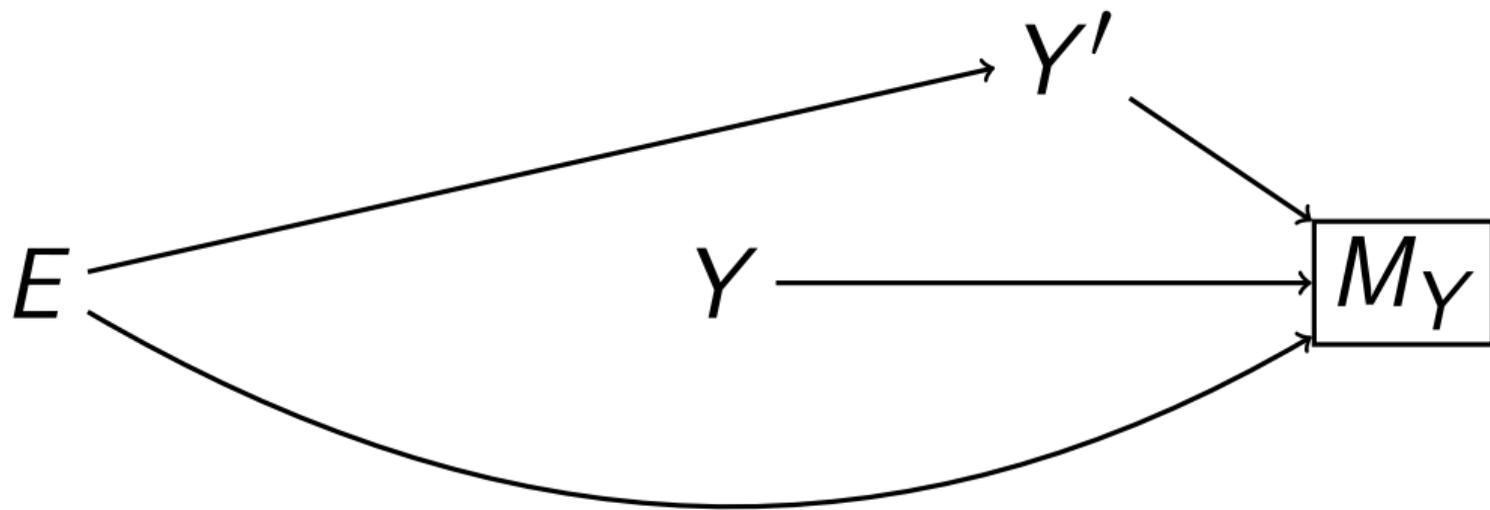
Differential
misclassification. . .

Biasing causal relations



... or association with other outcomes that interfere with M_Y (false positives, false negatives)

Biasing causal relations

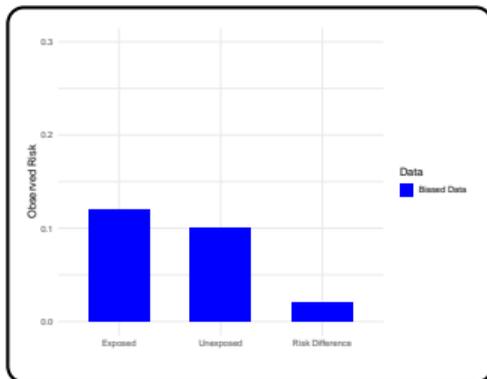


Biassing absolute measures of risk

If there are no false positives and sensitivity is non-differential, relative risks are unbiased but absolute risks are underestimated, which may affect decisions in public health

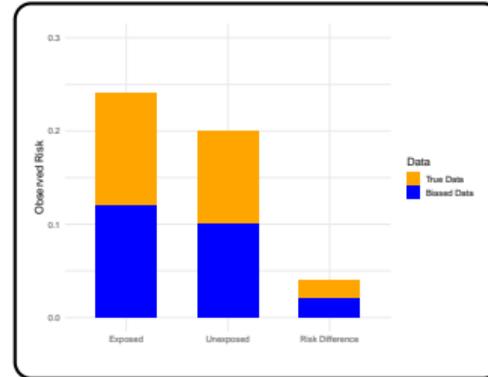
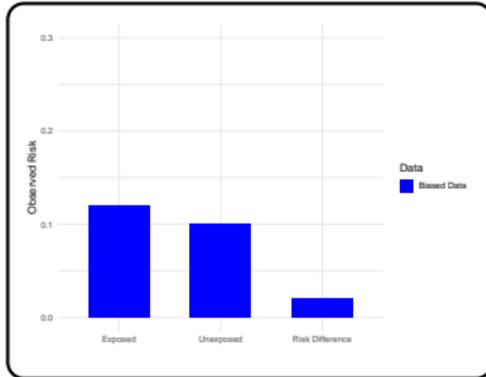
Biasing absolute measures of risk

If there are no false positives and sensitivity is non-differential, relative risks are unbiased but absolute risks are underestimated, which may affect decisions in public health



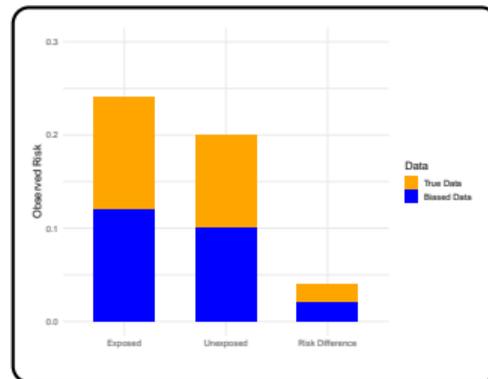
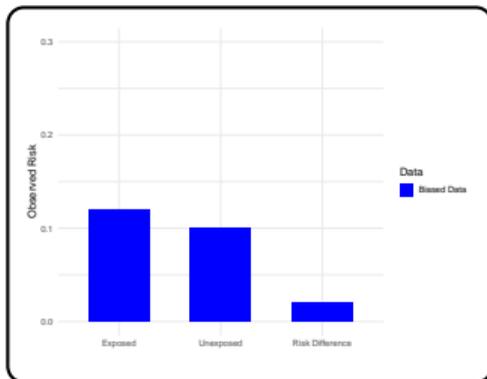
Biassing absolute measures of risk

If there are no false positives and sensitivity is non-differential, relative risks are unbiased but absolute risks are underestimated, which may affect decisions in public health



Biasing absolute measures of risk

If there are no false positives and sensitivity is non-differential, relative risks are unbiased but absolute risks are underestimated, which may affect decisions in public health



STROBE statement

Main results

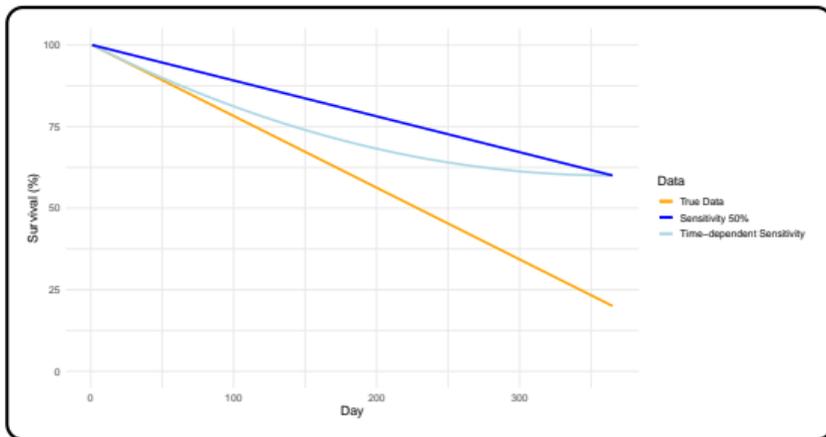
- 16 (a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included
- (b) Report category boundaries when continuous variables were categorized
- (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period



ARS TOSCANA
agenzia regionale di sanità

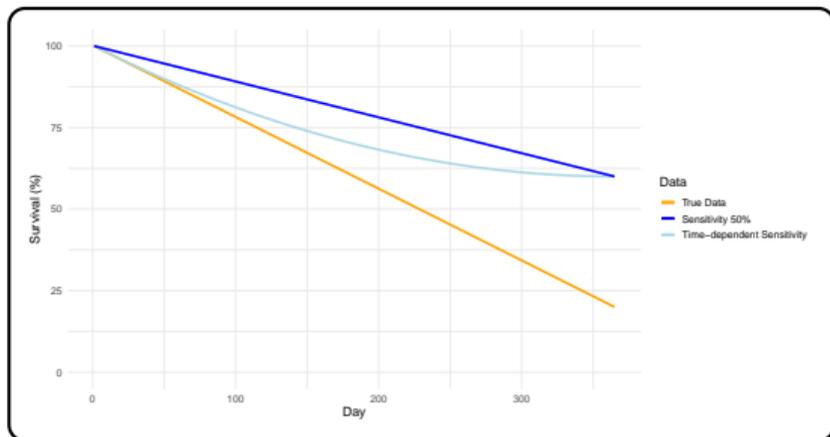
When time is involved

What about misclassification
across time?

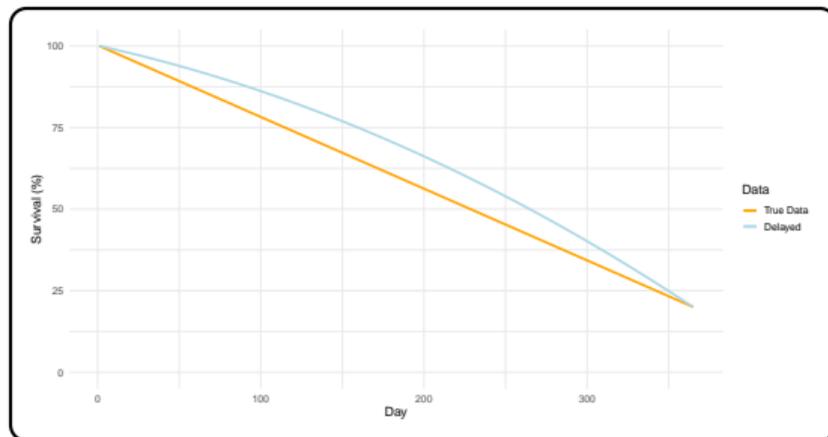


When time is involved

What about misclassification
across time?



What if detection is only delayed?



Contents

Introducing our heroes

Why is this important?

Understanding our heroes

What happens today?

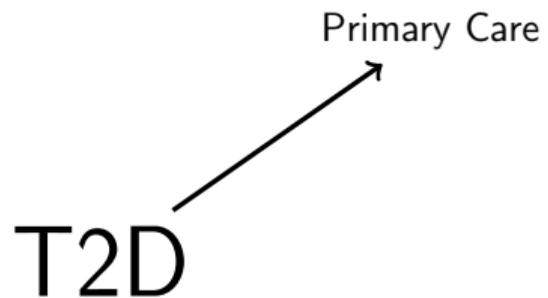
An example: type 2 diabetes

T2D

In our study
population type 2
diabetes is
developed

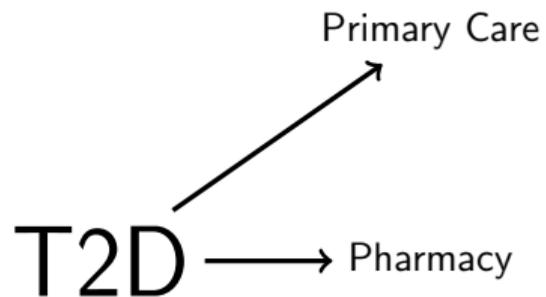
M_{T2D}

An example: type 2 diabetes



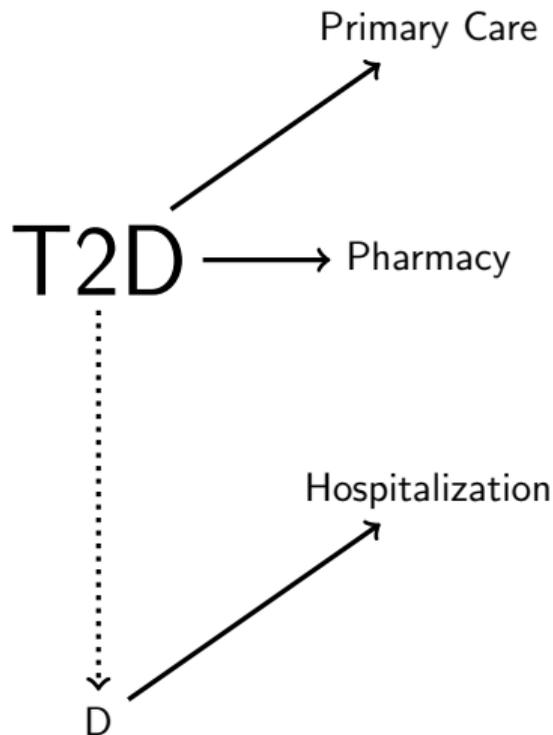
M_{T2D}

An example: type 2 diabetes



M_{T2D}

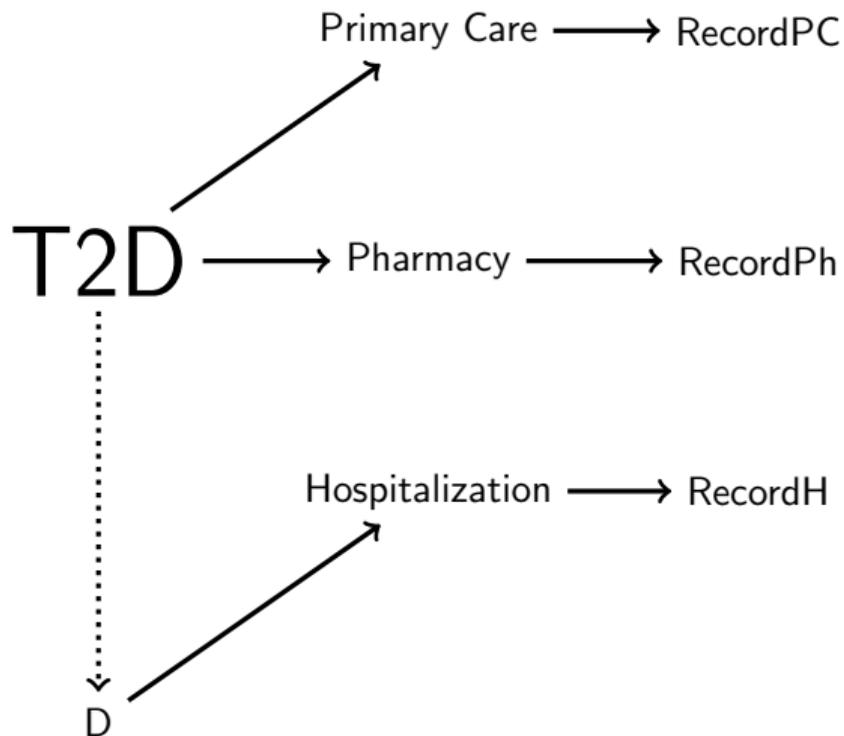
An example: type 2 diabetes



Disease causes access to primary care, and after a while use of antidiabetic medications. After more time: hospitalization, due to other diseases

M_{T2D}

An example: type 2 diabetes

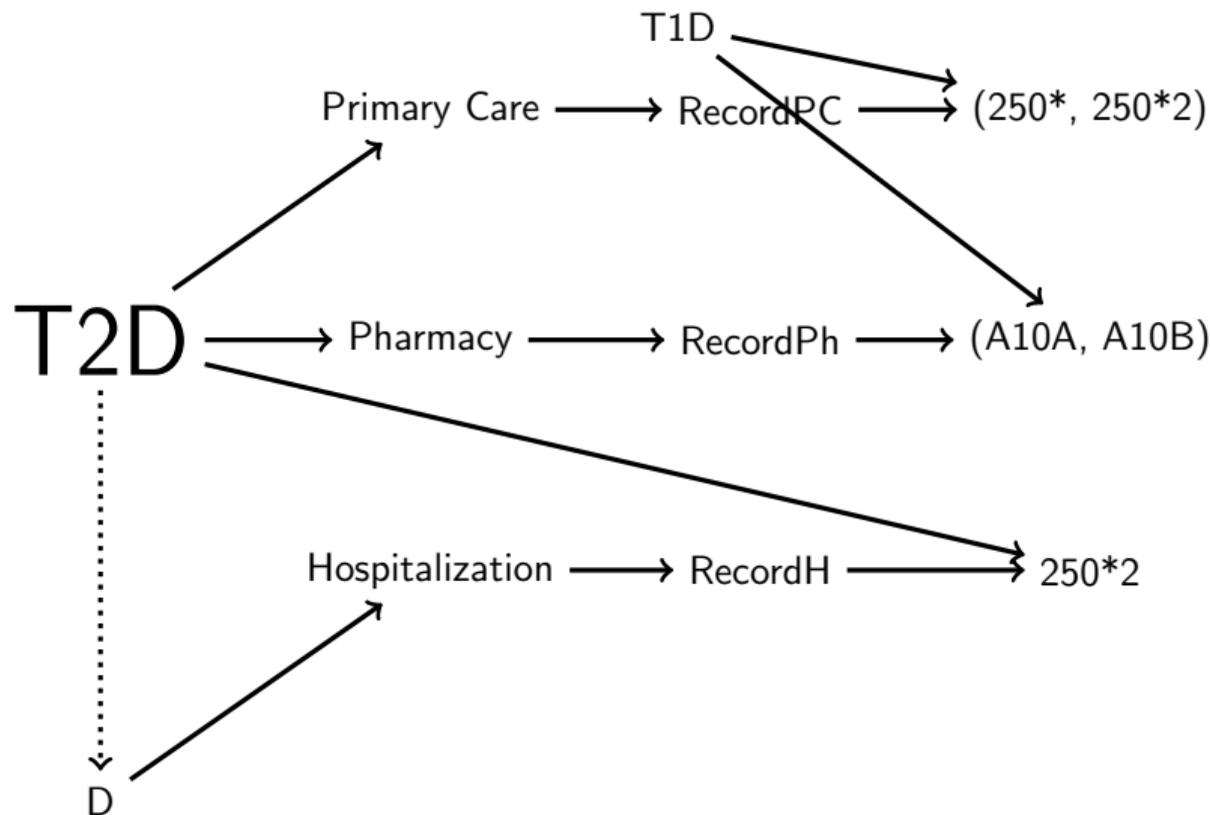


Such actions prompt records to exist in data banks

M_{T2D}

An example: type 2 diabetes

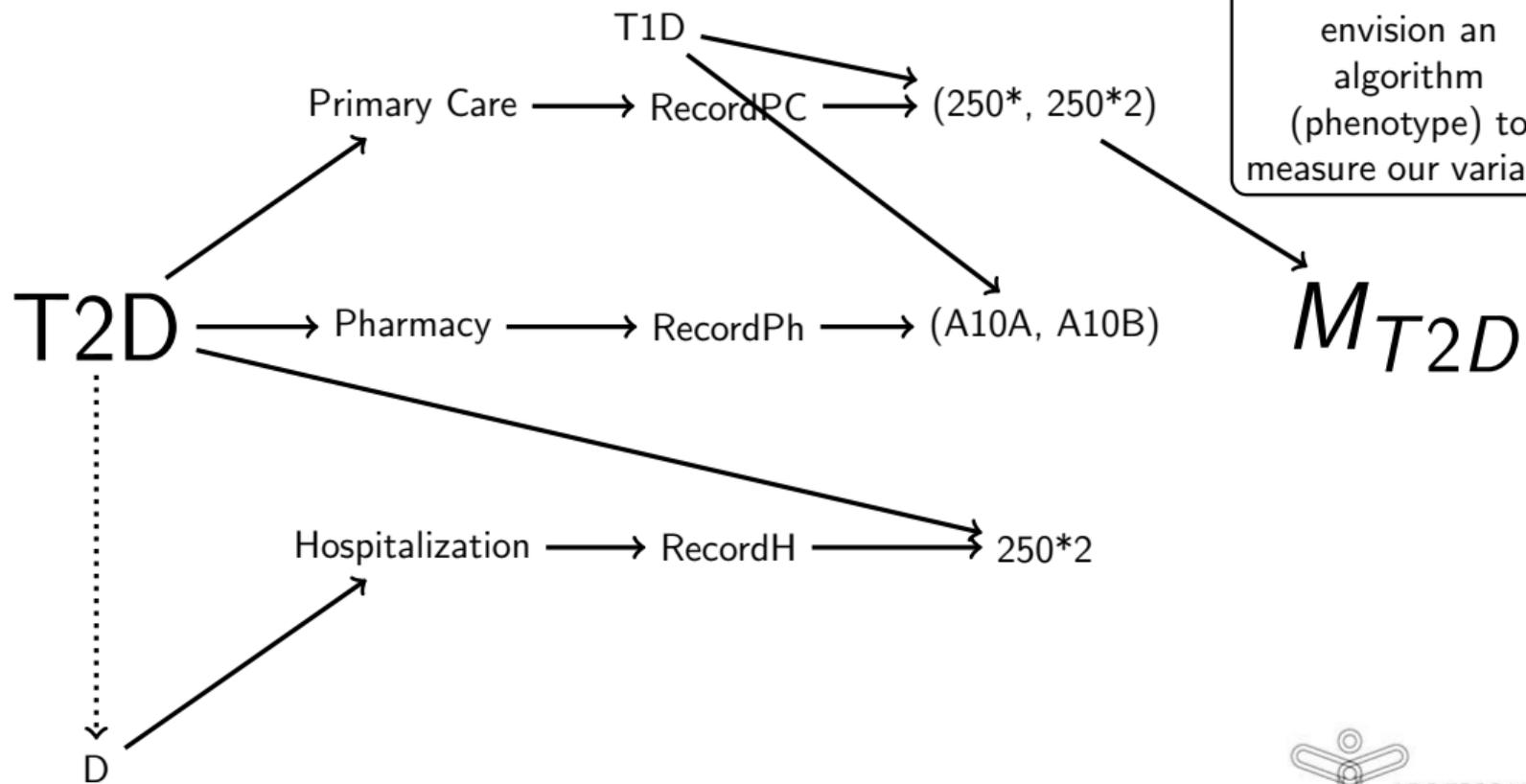
Content of records is populated



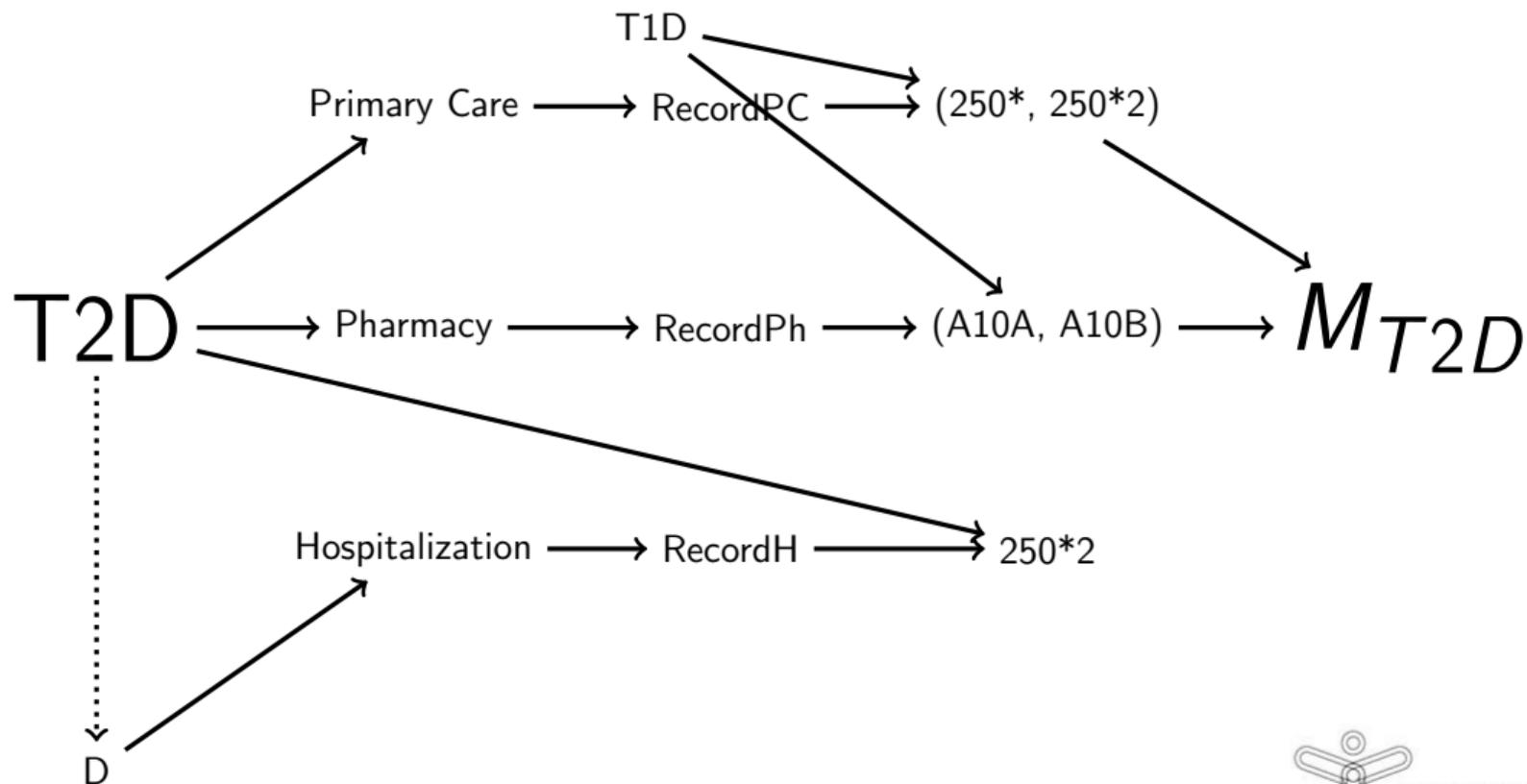
M_{T2D}

An example: type 2 diabetes

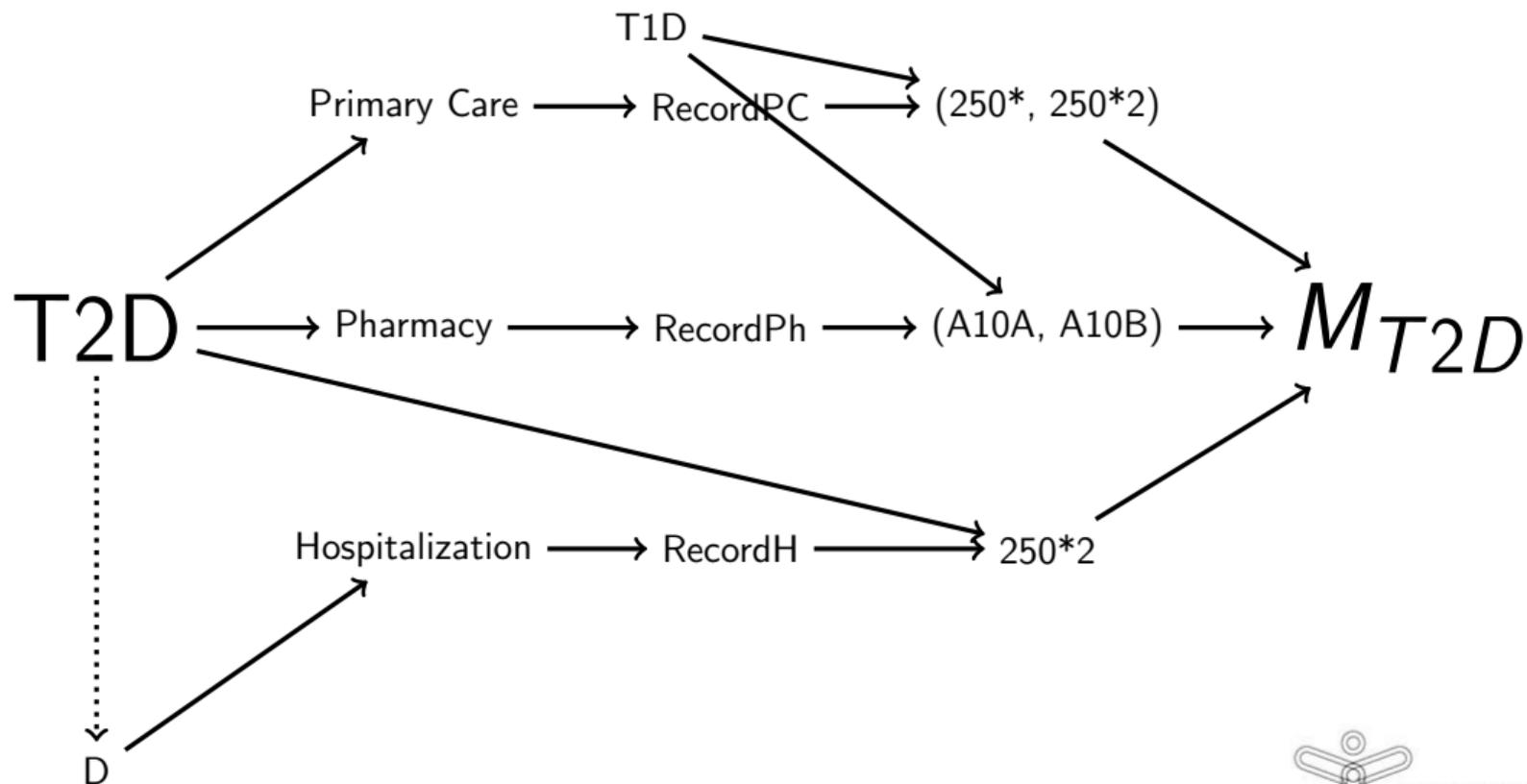
We get access to the data banks and envision an algorithm (phenotype) to measure our variable



An example: type 2 diabetes

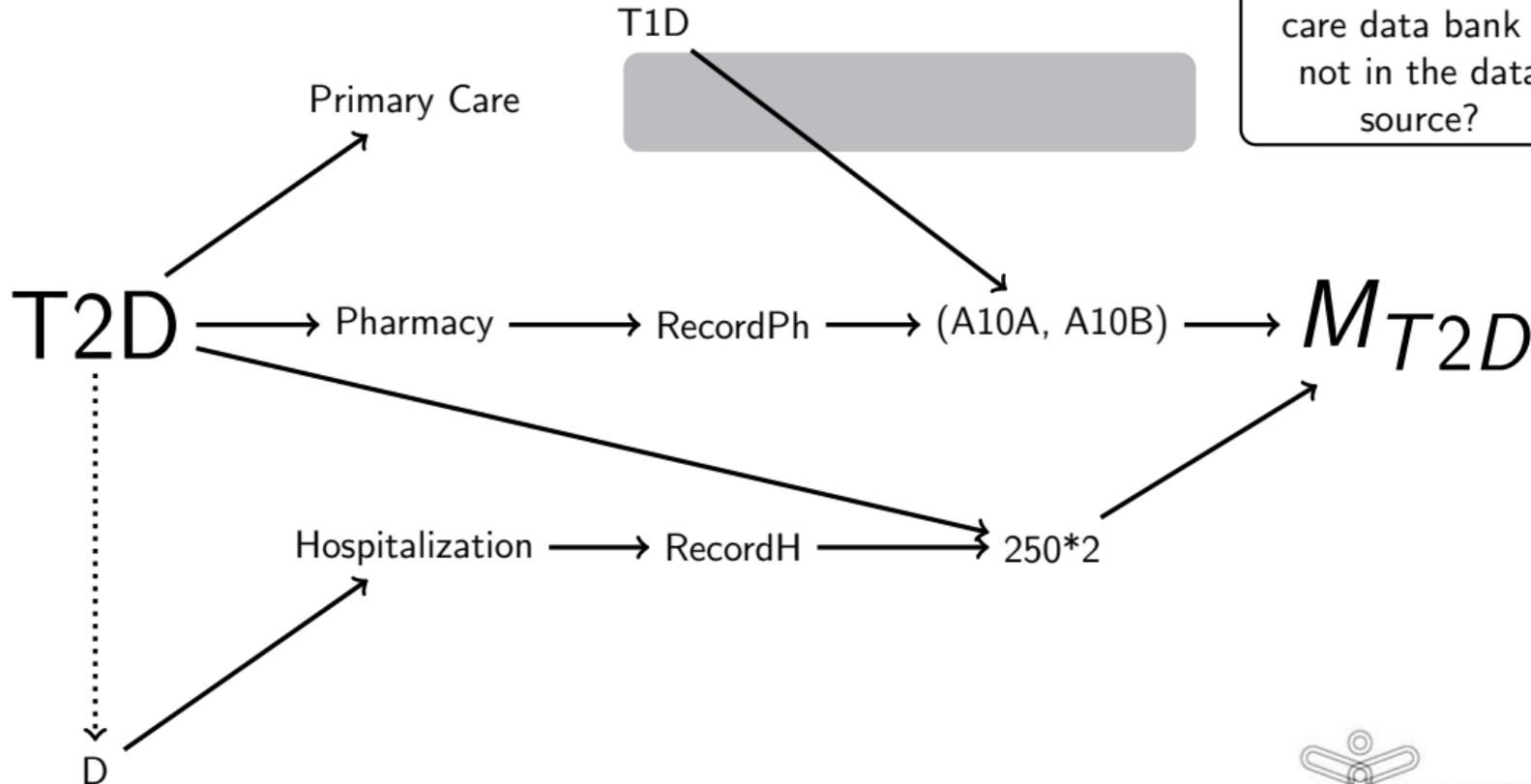


An example: type 2 diabetes

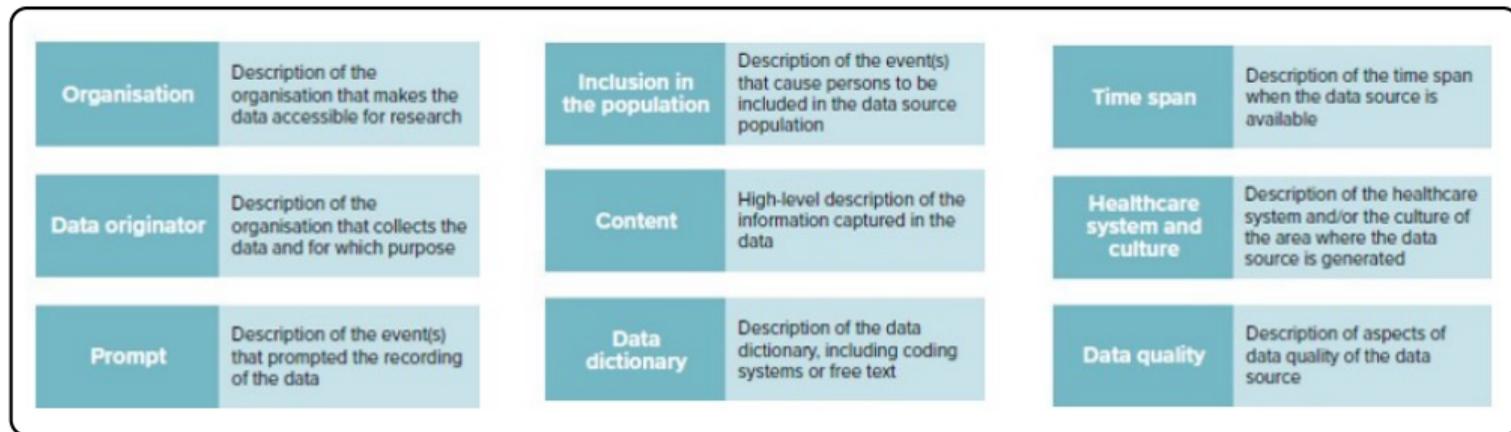


An example: type 2 diabetes

What if the primary care data bank is not in the data source?



The DIVERSE framework



Gini R, Pajouheshnia R, Gardarsdottir H, Bennett D, Li L, Gulea C, et al. Describing diversity of real world data sources in pharmacoepidemiologic studies: The DIVERSE scoping review. *Pharmacoepidemiology and Drug Safety*. 2024;33(5):e5787.

In general

Y

In our study
population a disease
occurs

M_Y

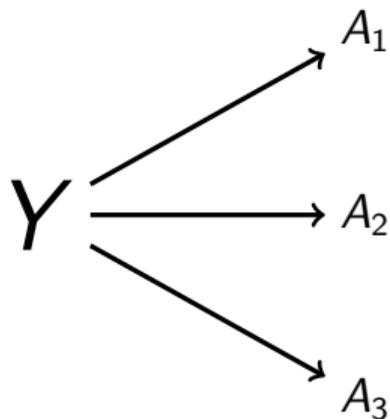
In general

Y

M_Y

Inclusion in
the population

In general



Multiple actions follow: access to hospital, access to primary care, access to emergency room, medicine dispensing, specialist encounter, death, ...

M_Y

Inclusion in the population

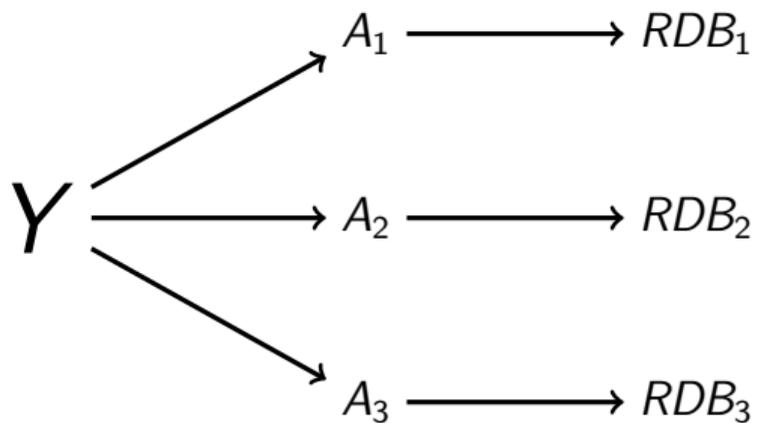


Healthcare system and culture



ARS TOSCANA
agenzia regionale di sanità

In general



Actions prompt
records to exist in
data banks

M_Y

Inclusion in
the population



Healthcare
system and
culture

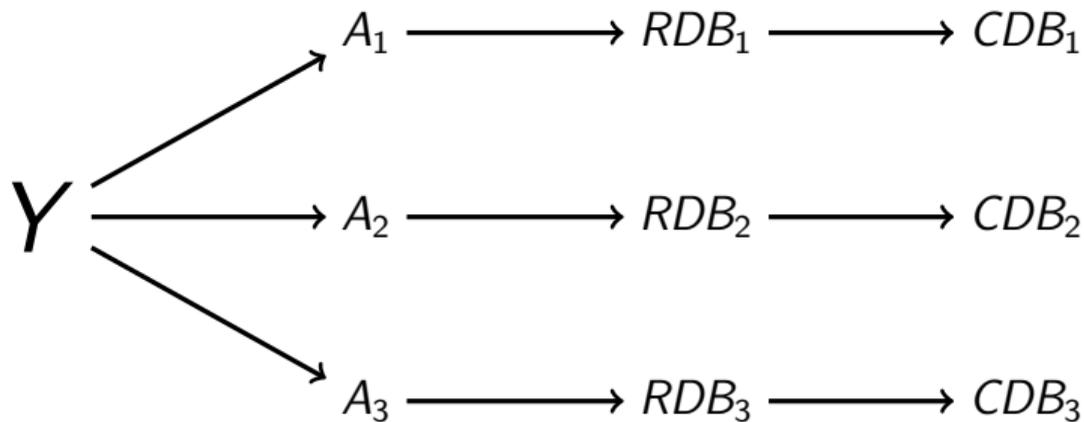
Prompt



ARS TOSCANA
agenzia regionale di sanità

In general

Content of records
is populated



M_Y

Inclusion in
the population



Healthcare
system and
culture

Prompt



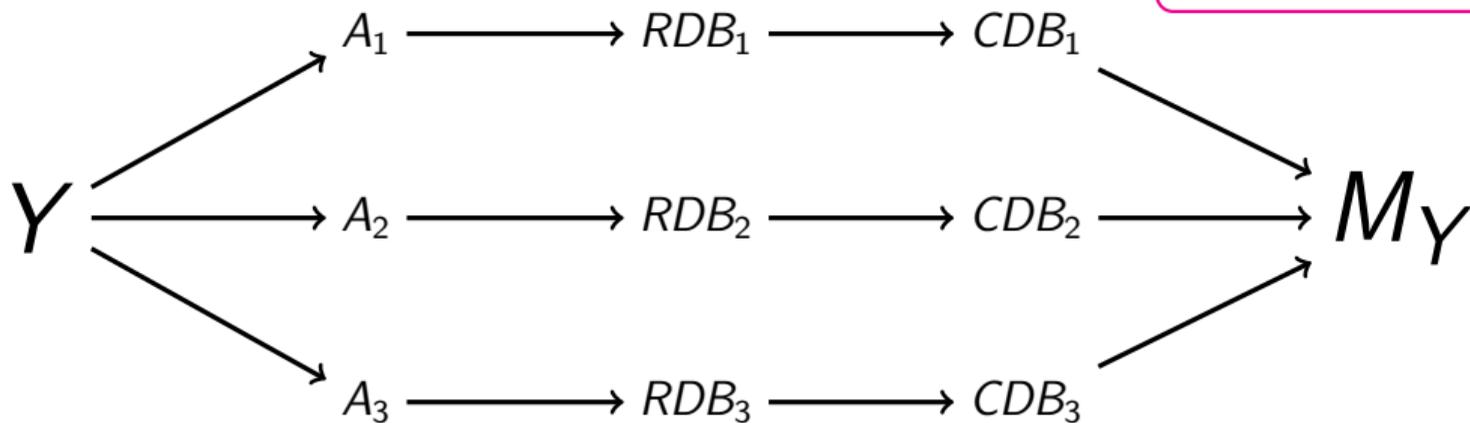
Content

Data originator



ARS TOSCANA
agenzia regionale di sanità

In general



We only enter the picture at the end with phenotype...

Inclusion in the population



Healthcare system and culture

Prompt



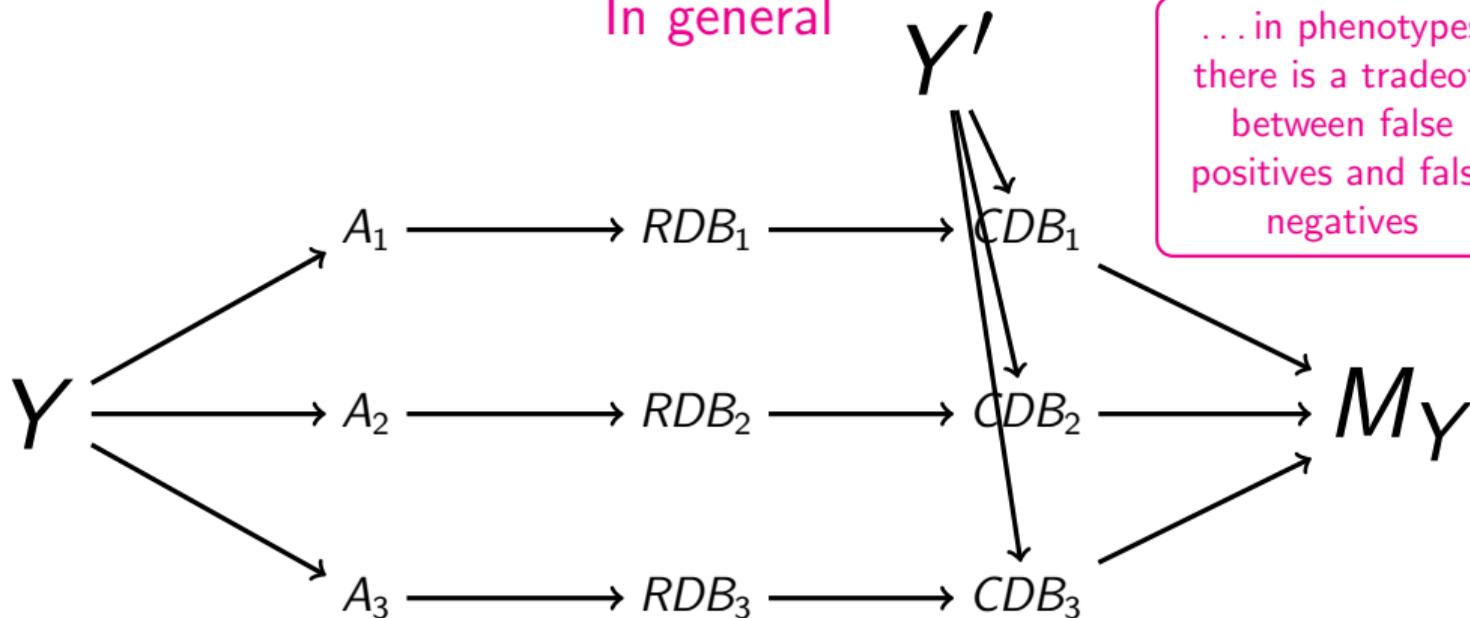
Content

Data originator



Organisation

In general



... in phenotypes there is a tradeoff between false positives and false negatives

Inclusion in the population

Healthcare system and culture

Prompt

Content

Data originator

Organisation

Contents

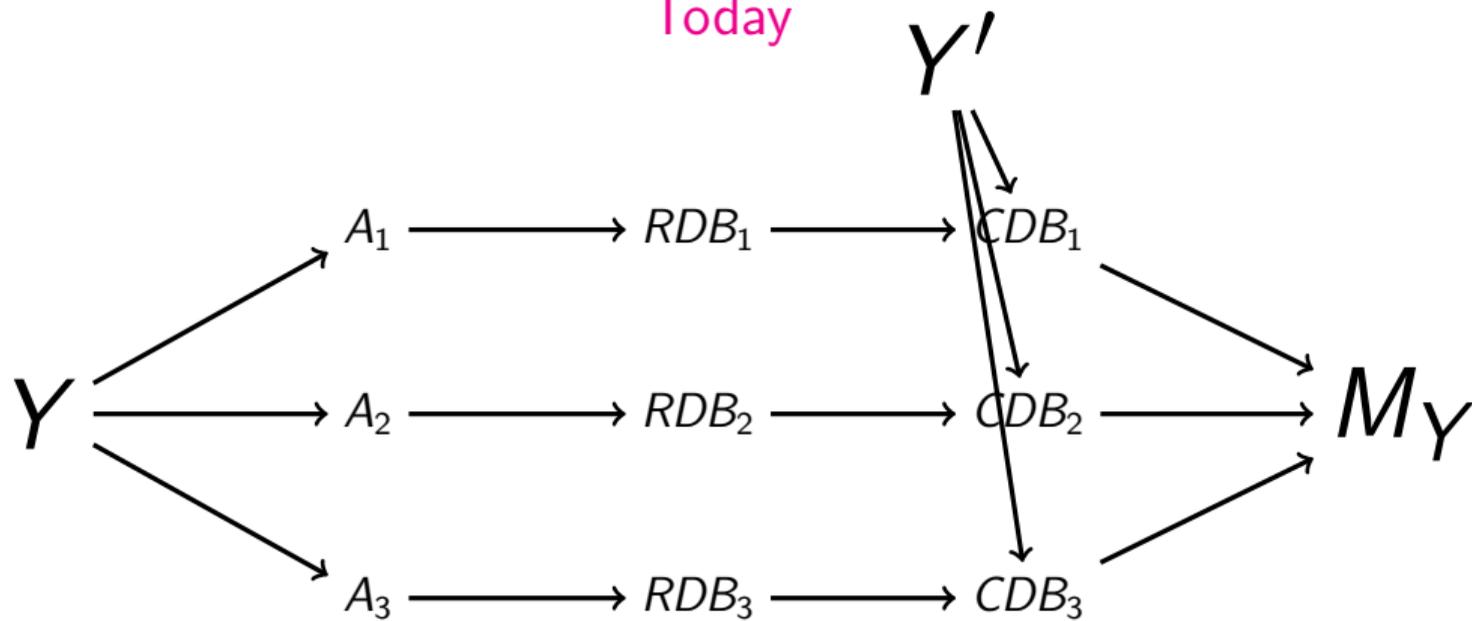
Introducing our heroes

Why is this important?

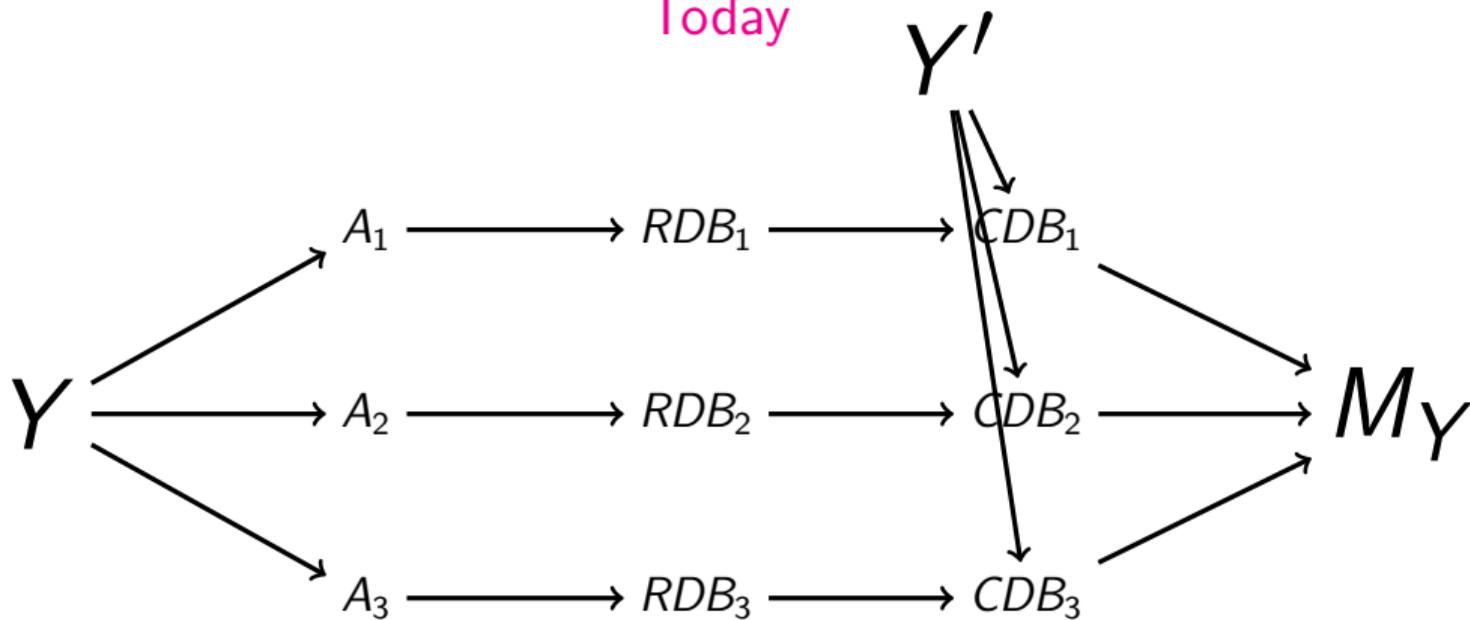
Understanding our heroes

What happens today?

Today



Today



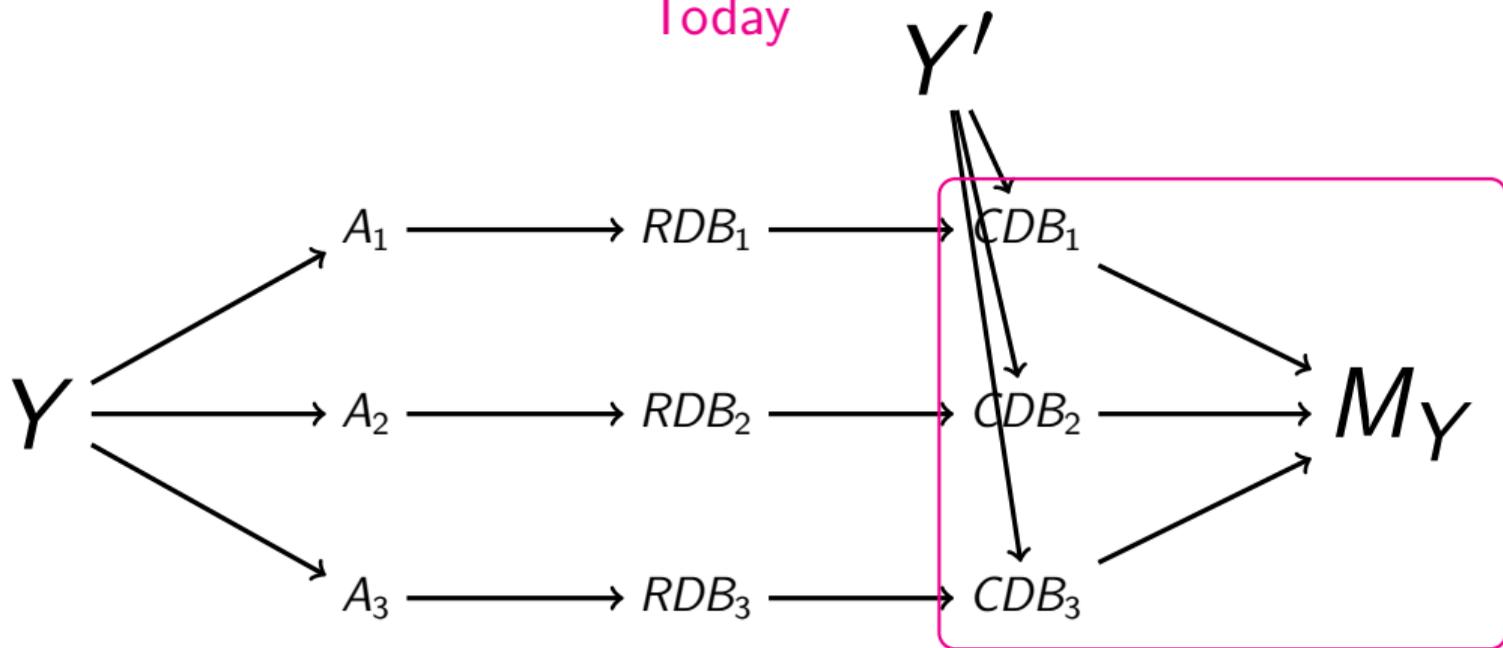
14.45

Recommendations and practice in validation studies and their use in studies. **Vincent Lo Re**



ARS TOSCANA
agenzia regionale di sanità

Today



How do we validate?

15.00

An ISPE guideline. **Vera Ehrenstein**

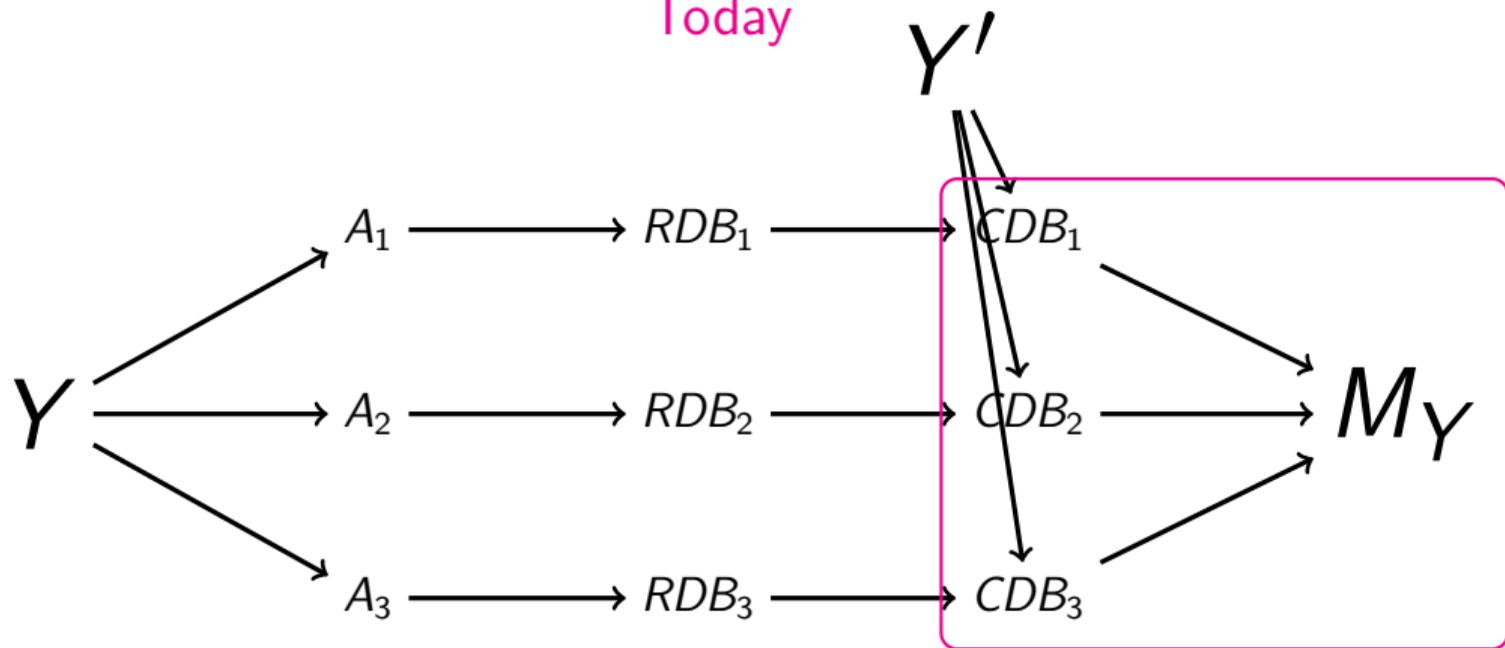
15.15

The VAC4EU validation pipeline. **Amirreza Dehghan Tarazjani**



ARS TOSCANA
agenzia regionale di sanità

Today



Advances in validation methodology

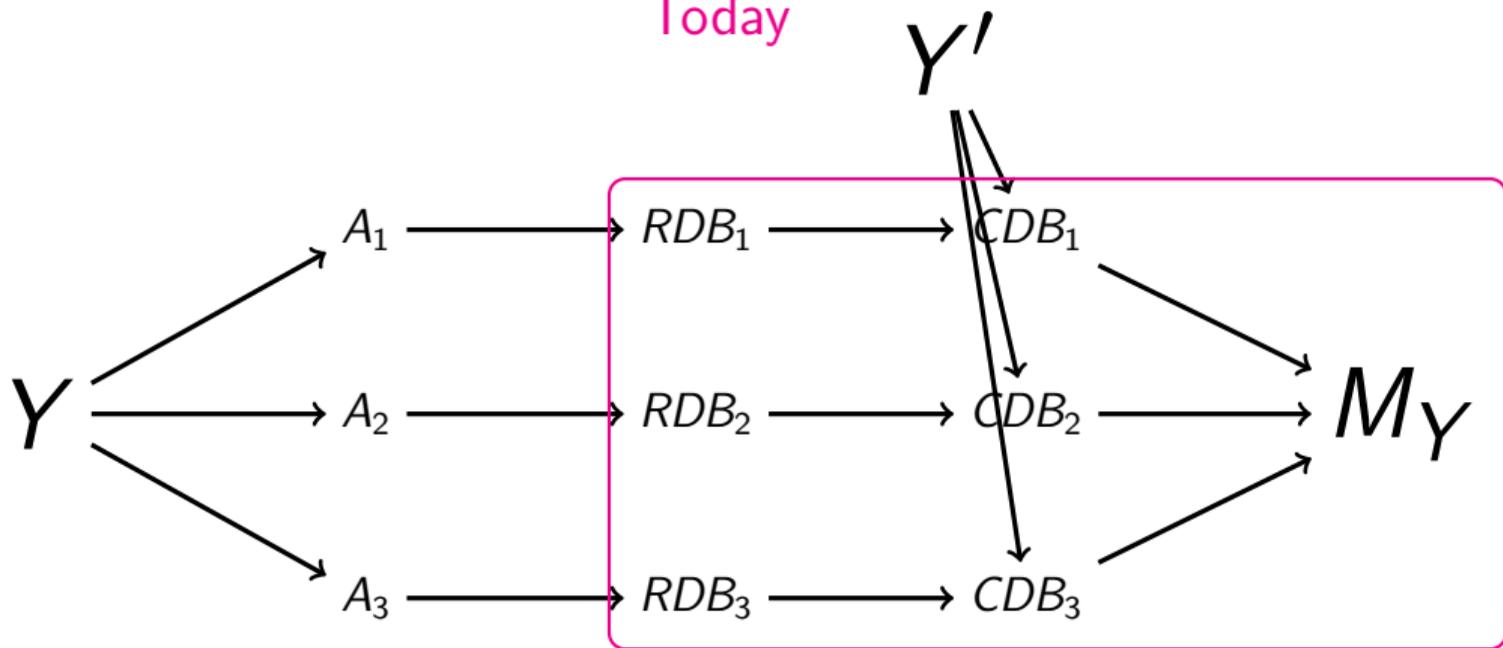
15.30

Interrelation between validity indices and validation in French administrative data.
Nicolas Thurin



ARS TOSCANA
agenzia regionale di sanità

Today



How can we leverage “screening” algorithms?

Advances in validation methodology

15.45

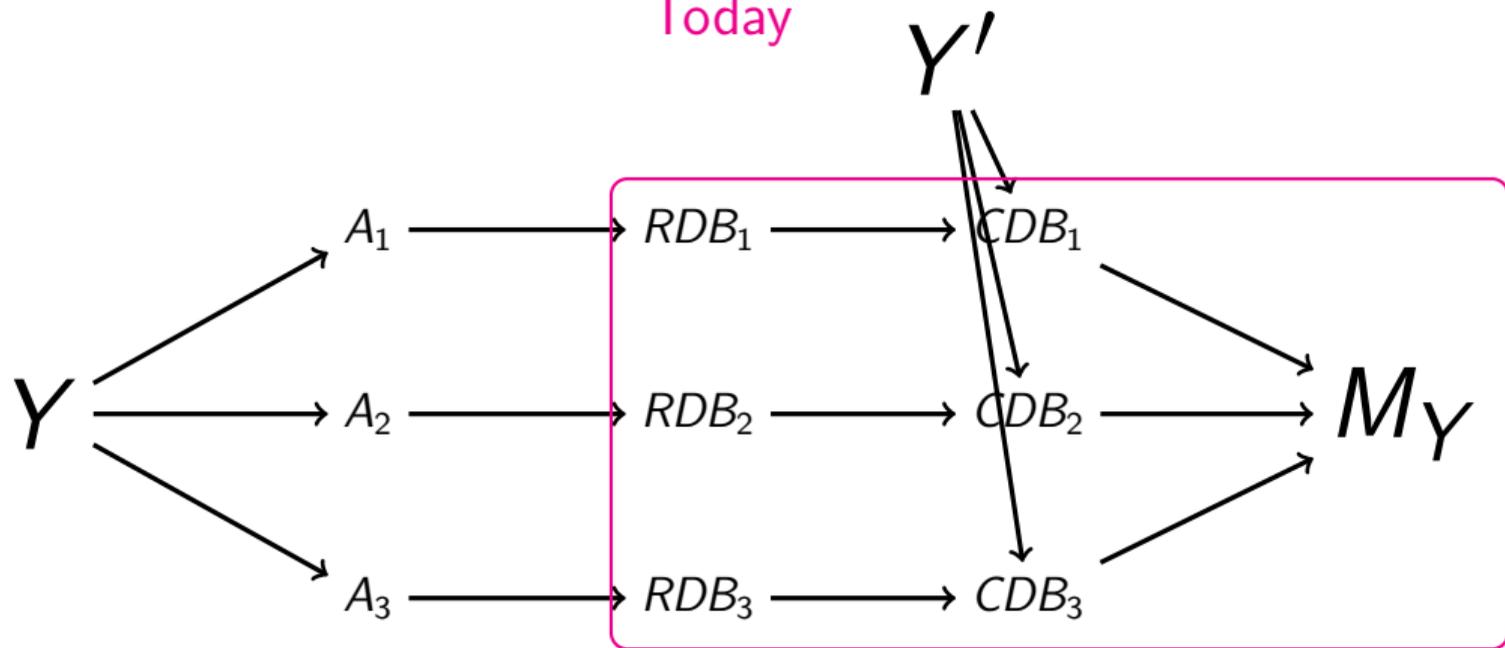
Screening algorithms to reduce underestimation and test for differential misclassification.

Giorgio Limocella



ARS TOSCANA
agenzia regionale di sanità

Today



... especially now that we have Artificial Intelligence tools?

Advances in validation methodology

16.00

The SeValid project. **Giulia Hyeraci**

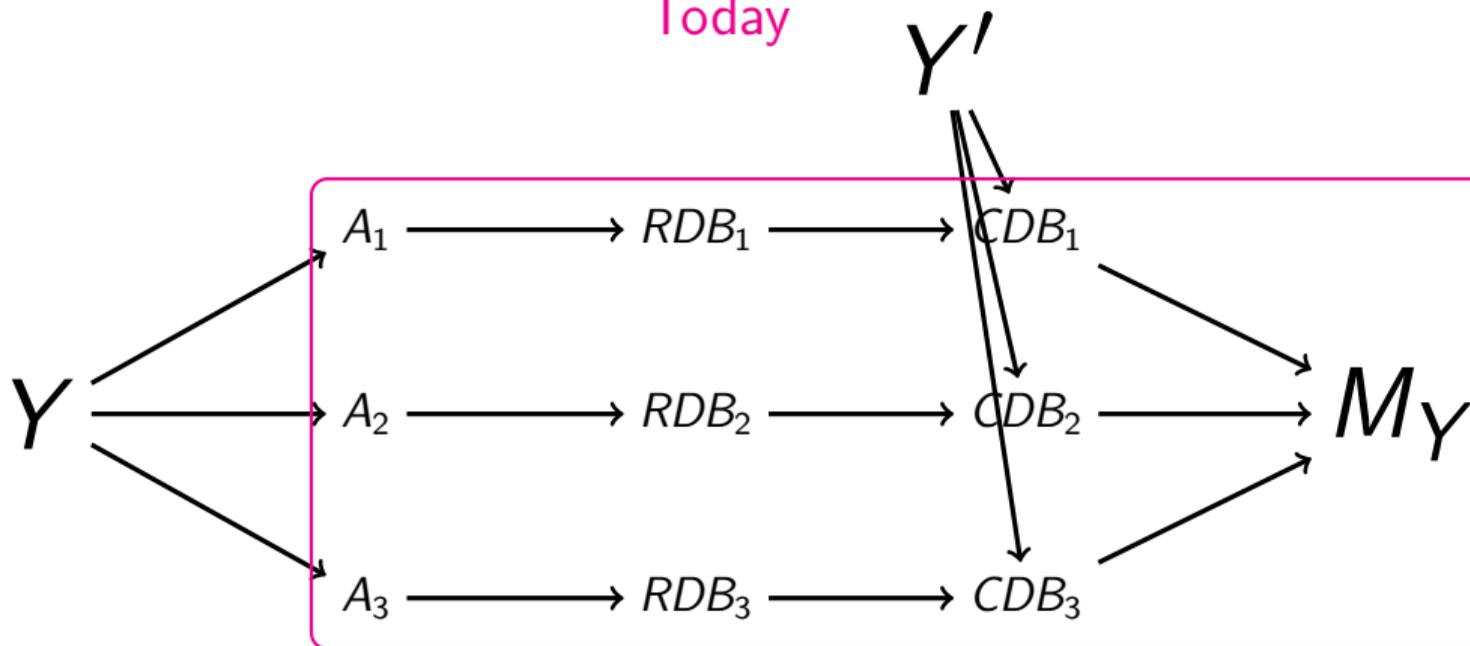
16.15

Contribution of Artificial Intelligence to validation. **Marco Lippi**



ARS TOSCANA
agenzia regionale di sanità

Today



Can we use ablation to understand the contribution of missing data banks?

Advances in validation methodology

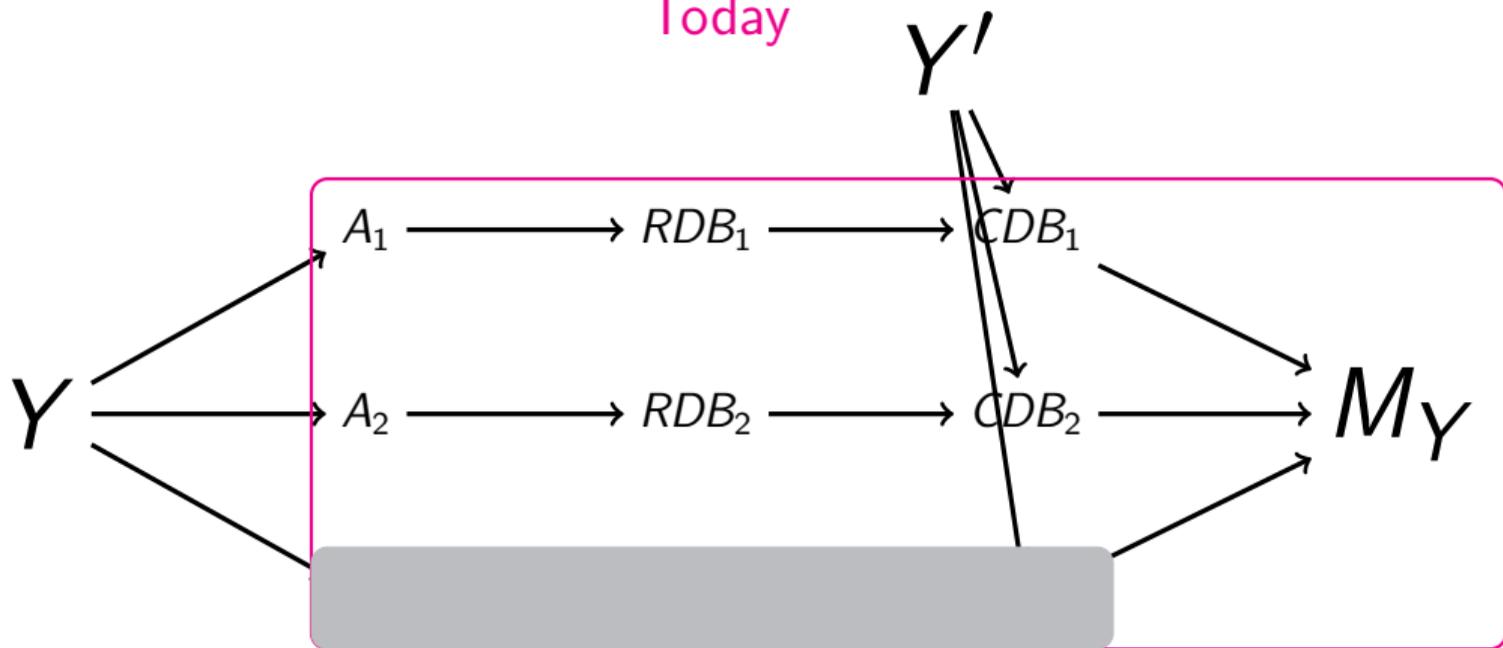
16.30

Ablation of prompts to estimate sensitivity of algorithms in a data source. **Giuseppe Roberto**



ARS TOSCANA
agenzia regionale di sanità

Today



Can we use ablation to understand the contribution of missing data banks?

Advances in validation methodology

16.30

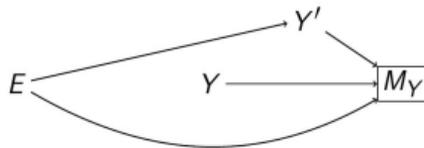
Ablation of prompts to estimate sensitivity of algorithms in a data source. **Giuseppe Roberto**



ARS TOSCANA
agenzia regionale di sanità

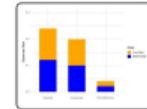
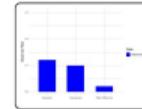
Today

Biasing causal relations



Biasing absolute measures of risk

If there are no false positives and sensitivity is non-differential, relative risks are unbiased but absolute risks are underestimated, which may affect decisions in public health



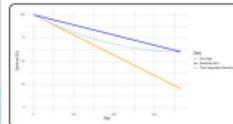
STROBE statement

- Main results
16. Use unadjusted estimates and, if applicable, confidence intervals and their precision (eg, 95% confidence interval). Make clear which confidence interval is being used and why they were included.
 17. Report complete numerator and denominator variables and categories.
 18. If relevant, consider presenting estimates of relative risk also include risk for a meaningful time period.

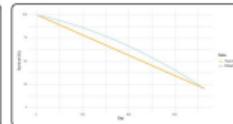


When time is involved

What about misclassification across time?



What if detection is only delayed?



Use of validation results in studies

17.00

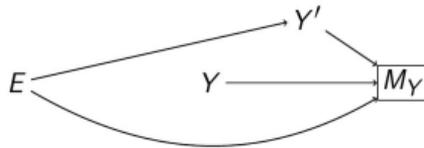
Quantitative bias analysis in practice. **Anna Schultze**



ARS TOSCANA
agenzia regionale di sanità

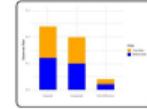
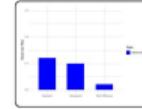
Today

Biasing causal relations



Biasing absolute measures of risk

If there are no false positives and sensitivity is non-differential, relative risks are unbiased but absolute risks are underestimated, which may affect decisions in public health



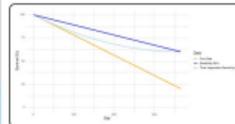
STROBE statement

- Main results
- 16. Use unadjusted estimates and, if applicable, confidence intervals and their precision (eg, 95% confidence interval). Make clear which confidence interval is being used and why they may be listed.
 - 17. Report complete numerator and denominator variables and corresponding N.
 - 18. If relevant, consider presenting estimates of relative risk also include risk for a meaningful time period.

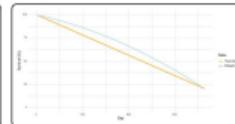


When time is involved

What about misclassification across time?



What if detection is only delayed?



Use of validation results in studies

17.15

Validation in BIFAP and application in studies: experience and challenges. **Elisa Martin Merino**



ARS TOSCANA
agenzia regionale di sanità

Role playing!



Xabi will play the role of an investigator writing a protocol and concerned about misclassification in his study; presenters will provide recommendations

Role playing!



Xabi will play the role of an investigator writing a protocol and concerned about misclassification in his study; presenters will provide recommendations



Ersilia and Robert will play the role of researchers developing a research plan; presenters will make requests

Role playing!



Xabi will play the role of an investigator writing a protocol and concerned about misclassification in his study; presenters will provide recommendations



Ersilia and Robert will play the role of researchers developing a research plan; presenters will make requests

17.30

Discussion. *Discussant:* **Xabier Garcia de Albeniz Martinez**

18.15

Conclusion. **Robert Platt, Ersilia Lucenteforte**

Let's kick off!

**Validate study variables to reduce
misclassification bias: recent tools and research needs**

HYBRID WORKSHOP

27 March 2025 - 14.30-18.30

ARS Toscana - Sala Rita Dioguardi, Villa La Quiete
via Pietro Dazzi 1 - Florence, Italy



VAC4EU



UNIVERSITÀ
DEGLI STUDI
FIRENZE

Dipartimento di Statistica,
informatica, Applicazioni
"Giuseppe Parenti"
C.so Italia 65/5 50132



ARS TOSCANA
agenzia regionale di sanità