Measurement error and misclassification of covariates: Should we worry?

Veronika Deffner
TG 4 on measurement error and misclassification
of the STRATOS Initiative

DAGStat 2022, 31.03.2022

Errors during observations

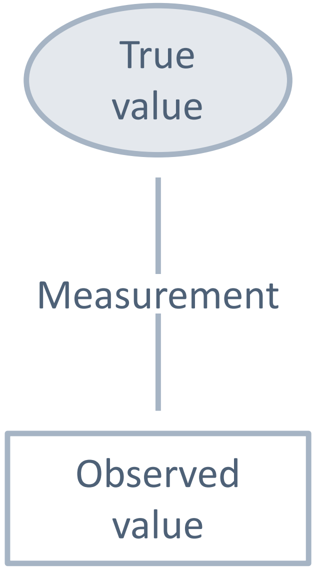
Simple questions …

- What did you eat yesterday?
- How tall are you?
- How high is your blood pressure?

… with a seemingly straightforward answer.

Two alternatives:

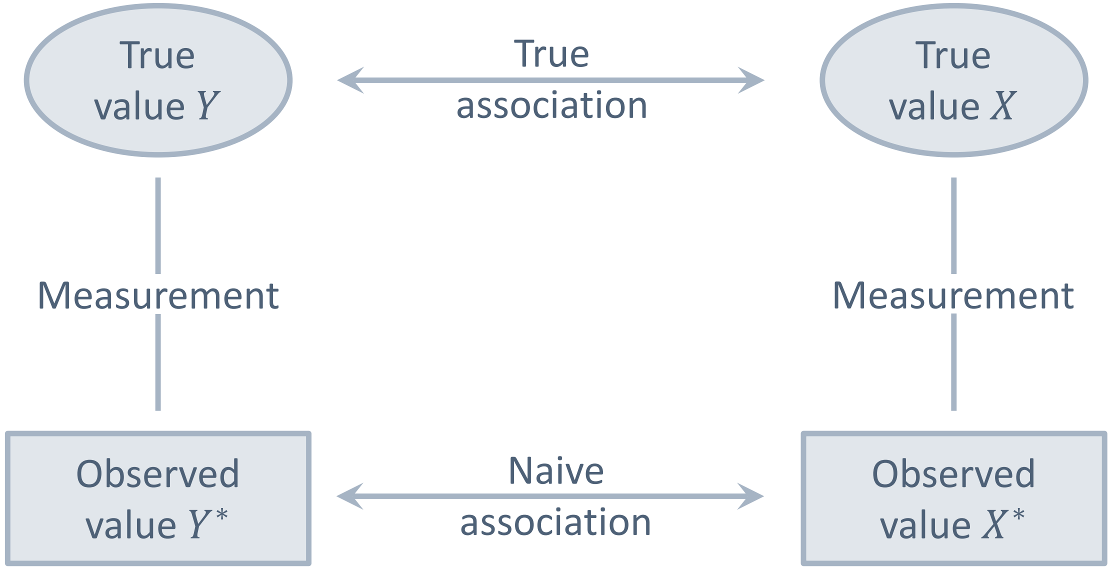
- Assumption: the observed value is exactly the same as whatever we are trying to measure
- Use a statistical method to account for measurement error in the analysis



Analyses with erroneous observations

True outcome model:

Observed outcome model:



STRATOS TG 4: Measurement error and misclassification

Aims

1. Increase the **awareness** of the problems caused by measurement error and misclassification in statistical analyses

2. Remove barriers to use **statistical methods** that deal with such problems

Activities

- Article on the **general relevance** of the topic (Wallace 2020)
- **Literature survey** about current practice (Shaw et al. 2018)
- **STRATOS guidance document:**

  Part 1 – Basic theory and simple methods of adjustment (Keogh et al. 2020)

  Part 2 – More complex methods of adjustment and advanced topics (Shaw et al. 2020)
- Presenting papers and workshops at **conferences**
- **Website:** http://www.stratostg4.statistik.uni-muenchen.de
- Interactive Shiny application **“MEM-Explorer”**

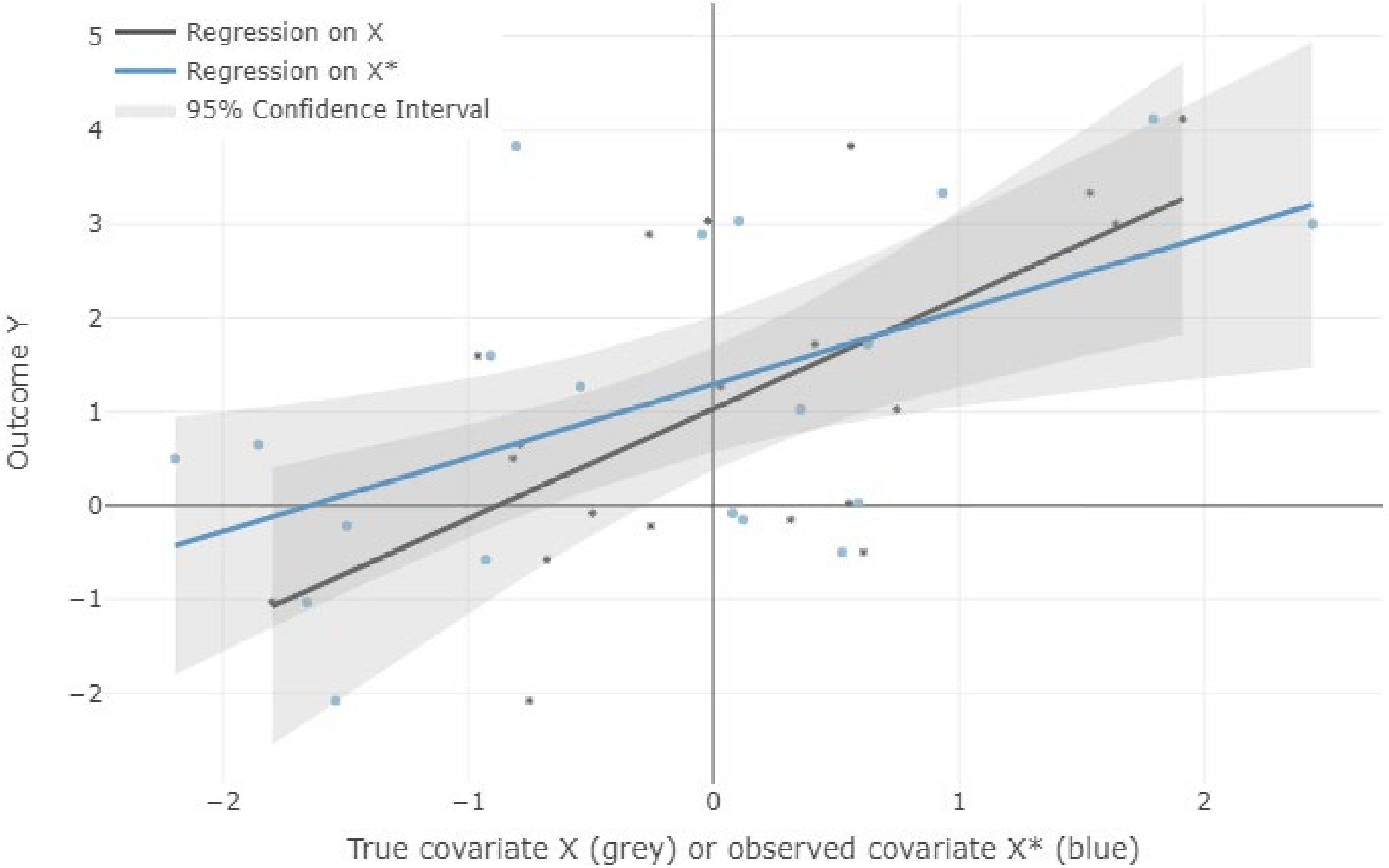
Impact on study results

Impact on …

- … the estimate of the **regression coefficient**
- … the **test of the null hypothesis** of no association
- … the **power** to detect the association

Depending on …

- … **error** (type, magnitude, etc.)
- … **model** (simple/multiple, linear, logistic, etc.)
- … **data** (variability, etc.)

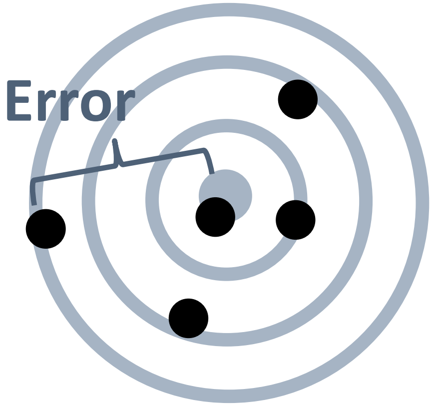


Classical measurement error

The observed value X^{*C} corresponds to the true value X, which is **overlaid with an error**, i.e. the observed value X^{*C} varies randomly around the true value X.

$$X^{*C} = X + U^C$$

X and U^C are assumed to be independent.
$Var(X^{*C}) > Var(X)$ if classical error is present.



<u>Example:</u>

X: average radon concentration in a uranium mine

X^{*C}: average of radon concentration measurements at different locations in the uranium mine

U^C: difference between X and X^{*C}

Berkson error

The observed value X^{*B} represents an **aggregated version** of the true value X, i.e. the true value X varies randomly around the observed value X^{*B}.

$$X = X^{*B} + U^B$$

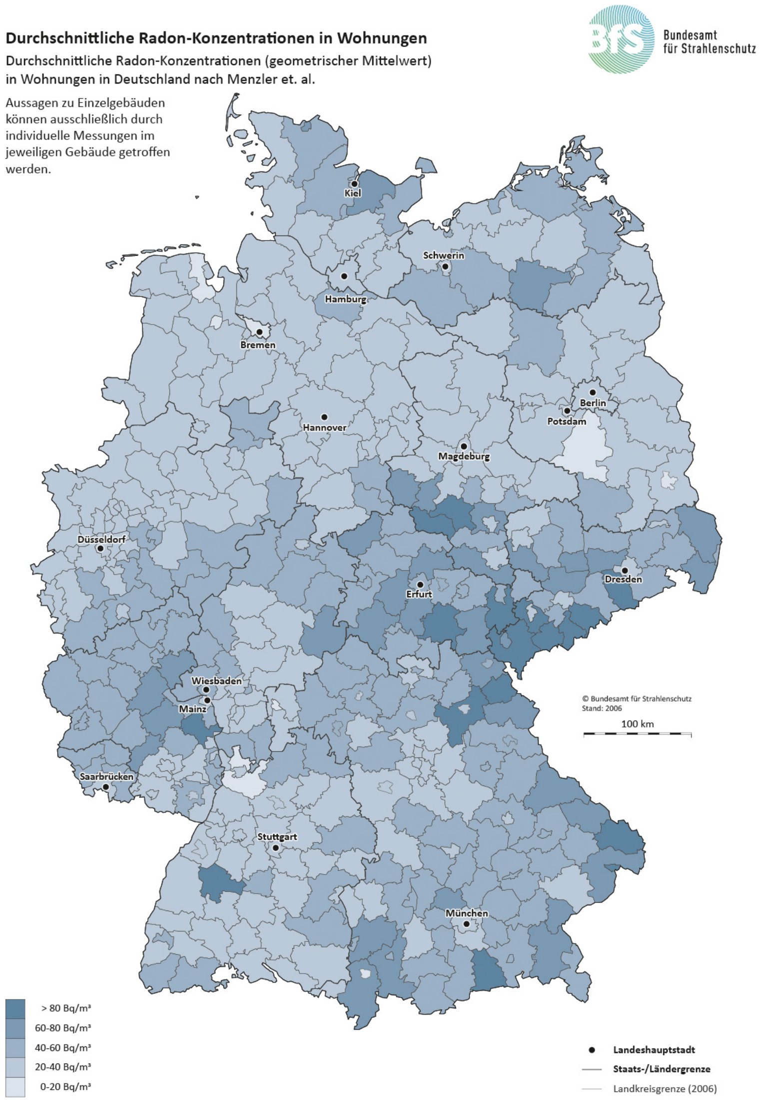
X^{*B} and U^B are assumed to be independent.
$Var(X^{*B}) < Var(X)$ if Berkson error is present.

Example:

X: indoor radon concentration in Neuherberg

X^{*B}: indoor radon concentration in the district of Munich

U^B: difference between X and X^{*B}



Average radon concentrations in dwellings in Germany (as of 2006)
Source: https://www.bfs.de/EN/topics/ion/environment/radon/maps/indoor.html (BfS, 2022)

Overview: effects of measurement error

		Nondifferential error		
Analysis	**Target**	**Classical**	**Linear**	**Berkson**
Single error-prone covariate regression	Regression coefficient	Underestimated	Biased in either direction	Sometimes unbiased[a]
	Test of null hypothesis	Valid	Valid	Sometimes valid[a]
	Power	Reduced	Reduced	Reduced[b]

[a]Unbiased and valid only when the Berkson error is independent of the other covariates in the model.
[b]The power of a test is only meaningful when the test of the null hypothesis is valid.

- **Linear measurement error**: $X^{*L} = \alpha_0 + \alpha_X X + U^L$
- **Differential error of X^***: the distribution of Y given X does not equal the distribution of Y given X and X^*, i.e. the measurement error contains extra information for a regression of Y on X.

Source: Keogh et al. 2020

Overview: effects of measurement error

		Nondifferential error			Differential error
Analysis	**Target**	**Classical**	**Linear**	**Berkson**	**Any**
Single error-prone covariate regression	Regression coefficient	Underestimated	Biased in either direction	Sometimes unbiased[a]	Biased in either direction
	Test of null hypothesis	Valid	Valid	Sometimes valid[a]	Invalid
	Power	Reduced	Reduced	Reduced[b]	Not applicable[b]
Regression with multiple error-prone covariates	Regression coefficients	Biased in either direction	Biased in either direction	Sometimes unbiased[a]	Biased in either direction
	Tests of null hypothesis	Invalid	Invalid	Sometimes valid[a]	Invalid
	Power	Not applicable[b]	Not applicable[b]	Reduced[b]	Not applicable[b]

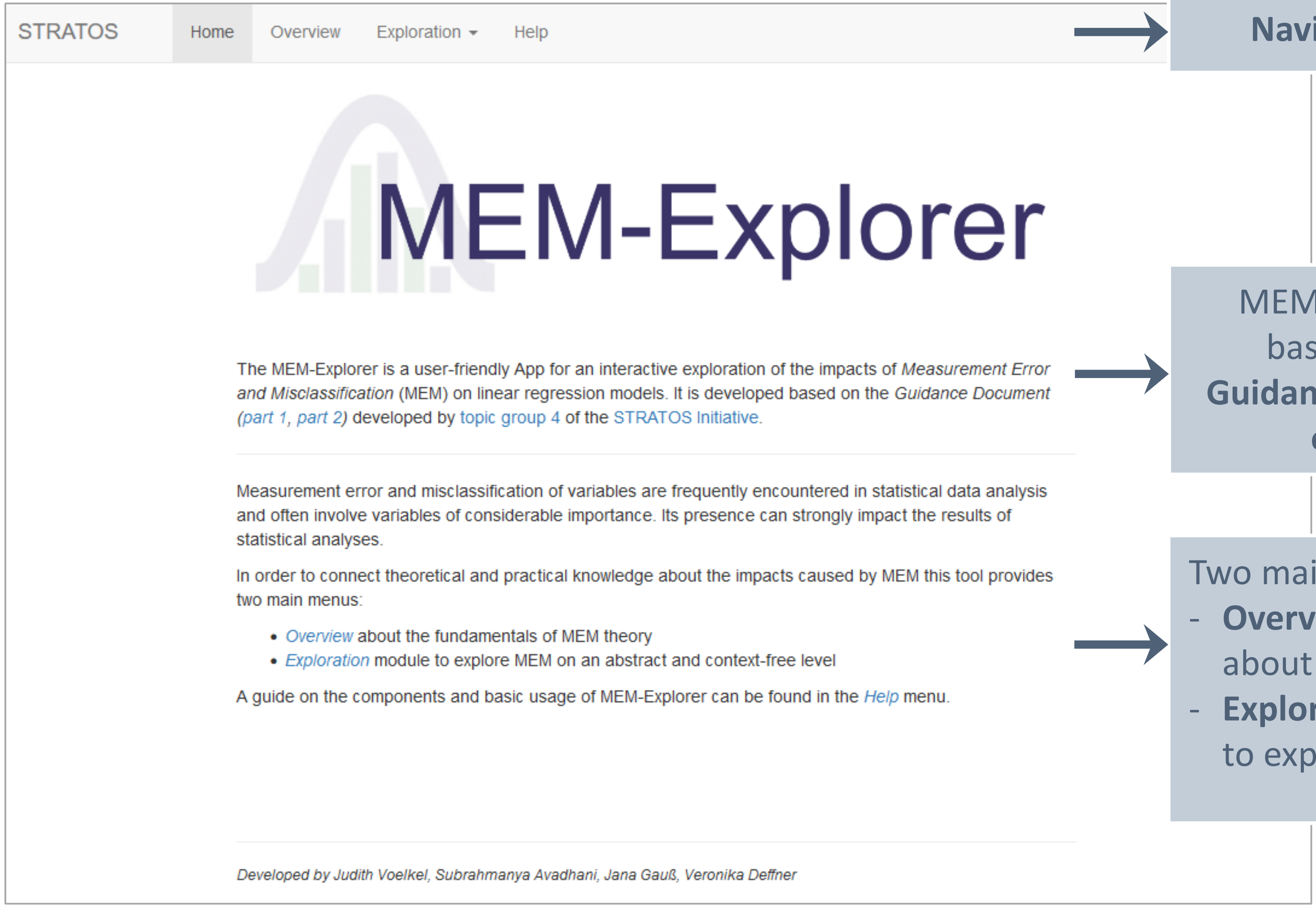
[a]Unbiased and valid only when the Berkson error is independent of the other covariates in the model.
[b]The power of a test is only meaningful when the test of the null hypothesis is valid.

Source: Keogh et al. 2020

Interactive exploration of measurement error and misclassification

MEM-Explorer:
Shiny app for interactive exploration of measurement error and misclassification

Find MEM-Explorer here:
https://mem-explorer.shinyapps.io/MEMExplorer-v5/



Navigation bar

MEM-Explorer is based on the **Guidance Document of TG 4**

Two main menus:
- **Overview** about theory
- **Exploration** module to explore MEM

MEM-Explorer - Overview

STRATOS Home Overview Exploration Help

Theoretical overview of Measurement Error and Misclassification

This page serves as an overview about the fundamentals of MEM theory and is mainly based on the guidance document (part 1, part 2) developed by STRATOS-Initiative.

Main Types of Error | Impact on Regression | Error Adjustment | Sources

This section describes the main types of error that occur in measurements during observational studies.

Terminology

Two seperate terms for errors in variables are distinguished:

- **measurement error** for error in continuous variables and
- **misclassification** for error in categorical variables.

Aim

The aim is to learn the regression relationship between a scalar outcome variable Y and covariates X.

Basic situation

X is measured with error, with the true value of X being unobserved. The error-prone observed variable is denoted by X^*.

Note: Mismeasurement can occur not only in covariates but also in the outcome variable Y. This is currently not considered in MEM-Explorer.

Requirements

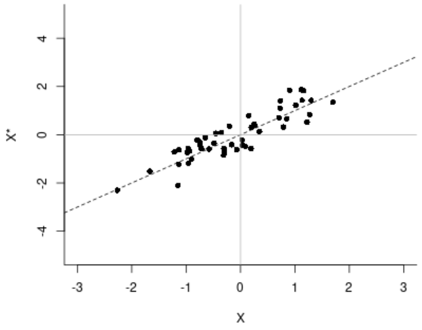
Learning the regression relationship betwen Y and X based on the observed covariates X^* requires knowledge about type and size of the error in X^*, i.e. the relationship of X and X^* has to be specified. This is called the **error model**.

Main Types of Error

Random vs. Systematic | Classical vs. Linear vs. Berkson | Additive vs. Multiplicative | Differential vs. Non-differential

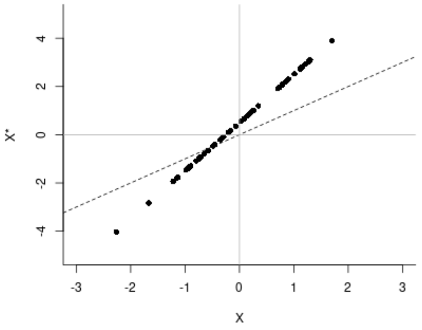
Random Error

Random or stochastic error occurs when X^* differs randomly from X. The following figure shows an example for random error.



Systematic Error

Systematic error occurs when X^* differs systematically from X. The following figure shows an example for systematic error with linear structure.



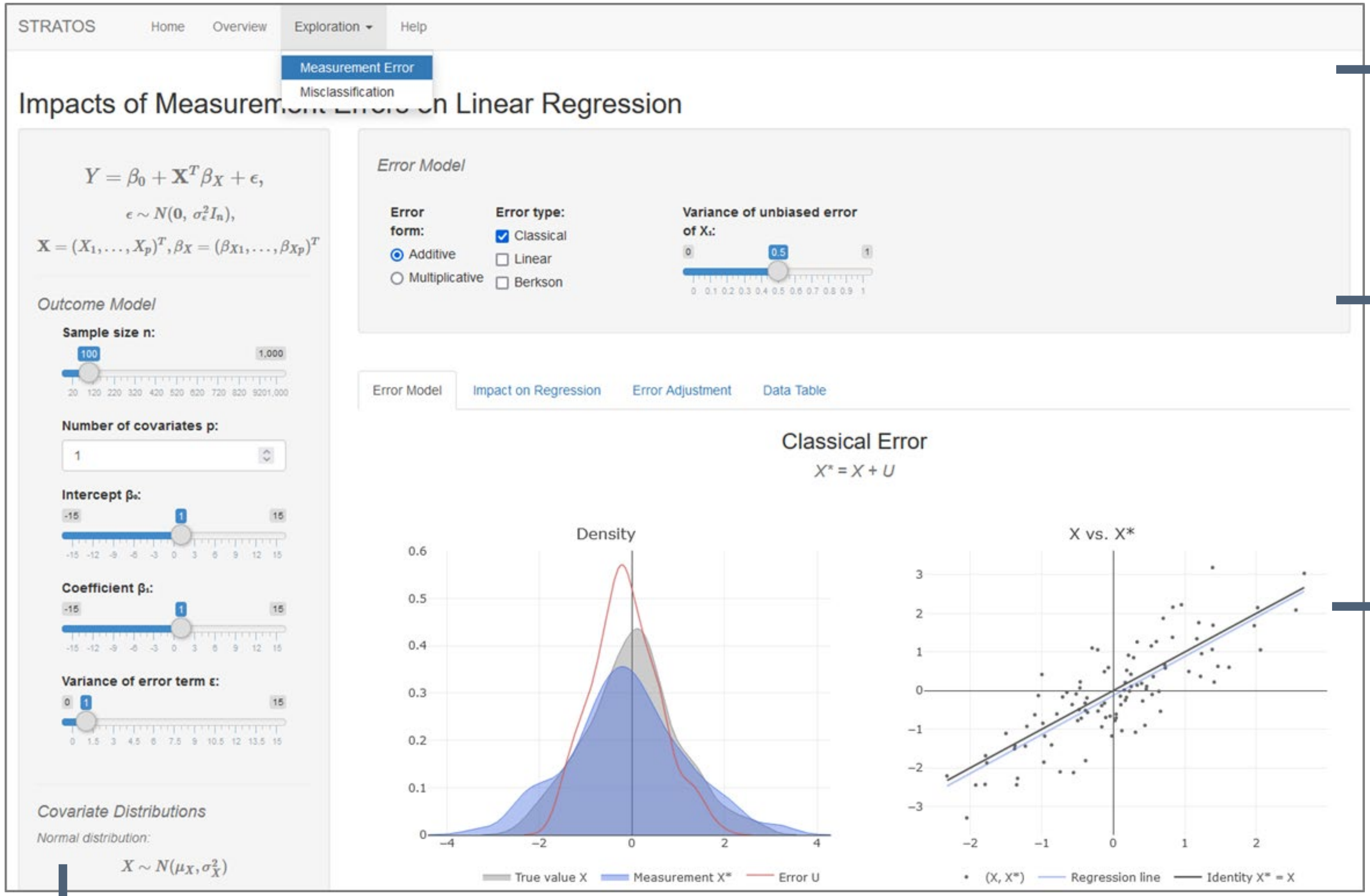
Sections:

- Main Types of error
- Impact on Regression
- Error Adjustment

Subsections for each section

Basics for each section

MEM-Explorer - Exploration



Exploration:

- **Measurement Error**
- **Misclassification**

Specification of **error model**

Subsections with output:

- Error Model
- Impact on Regression
- Error Adjustment
- Data Table

Specification of **sample size, outcome model** and **covariate distributions**

MEM-Explorer - Output

Error Model

Impact on Regression

Visualization of the impact of MEM on the regression model

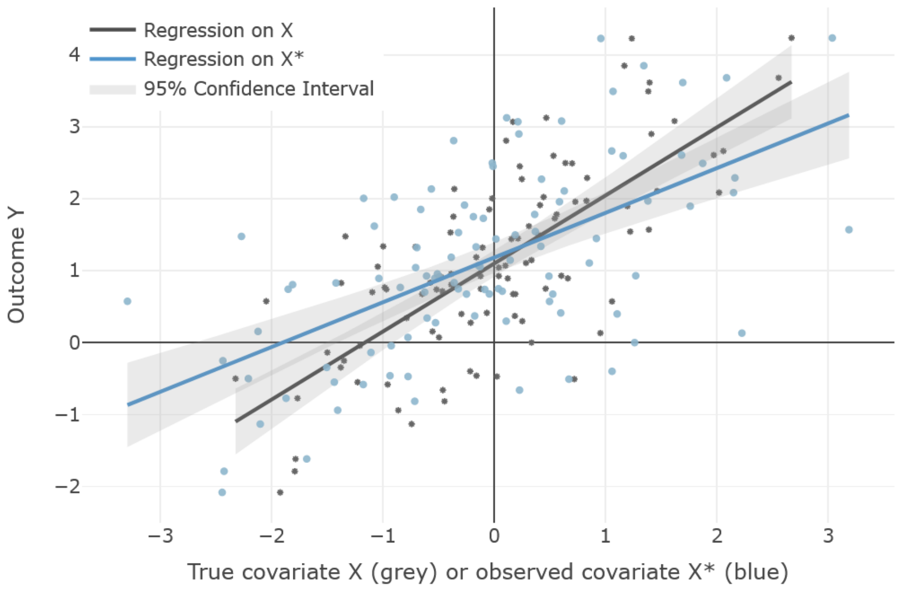
Error Adjustment

Data Table

Error Model | Impact on Regression | Error Adjustment | Data Table

Classical Error

$X^* = X + U$



	$Coeff_{Classical}$	$Coeff^*_{Classical}$	$\lambda_{Classical}$
Intercept	1.1	1.18	
X1	0.95	0.62	0.65

	$SE_{Classical}$	$SE^*_{Classical}$
Intercept	0.09	0.11
X1	0.09	0.09

	$p_{Classical}$	$p^*_{Classical}$
Intercept	< 0.001	< 0.001
X1	< 0.001	< 0.001

MEM-Explorer - Example

Sample size

$$n = 100$$

Outcome model

Simple linear regression model with

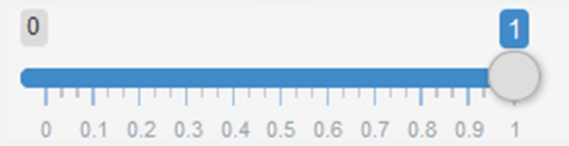
- coefficients $\beta_0 = 0$ and $\beta_1 = 1$
- error variance $\mathrm{Var}(\varepsilon) = 1$

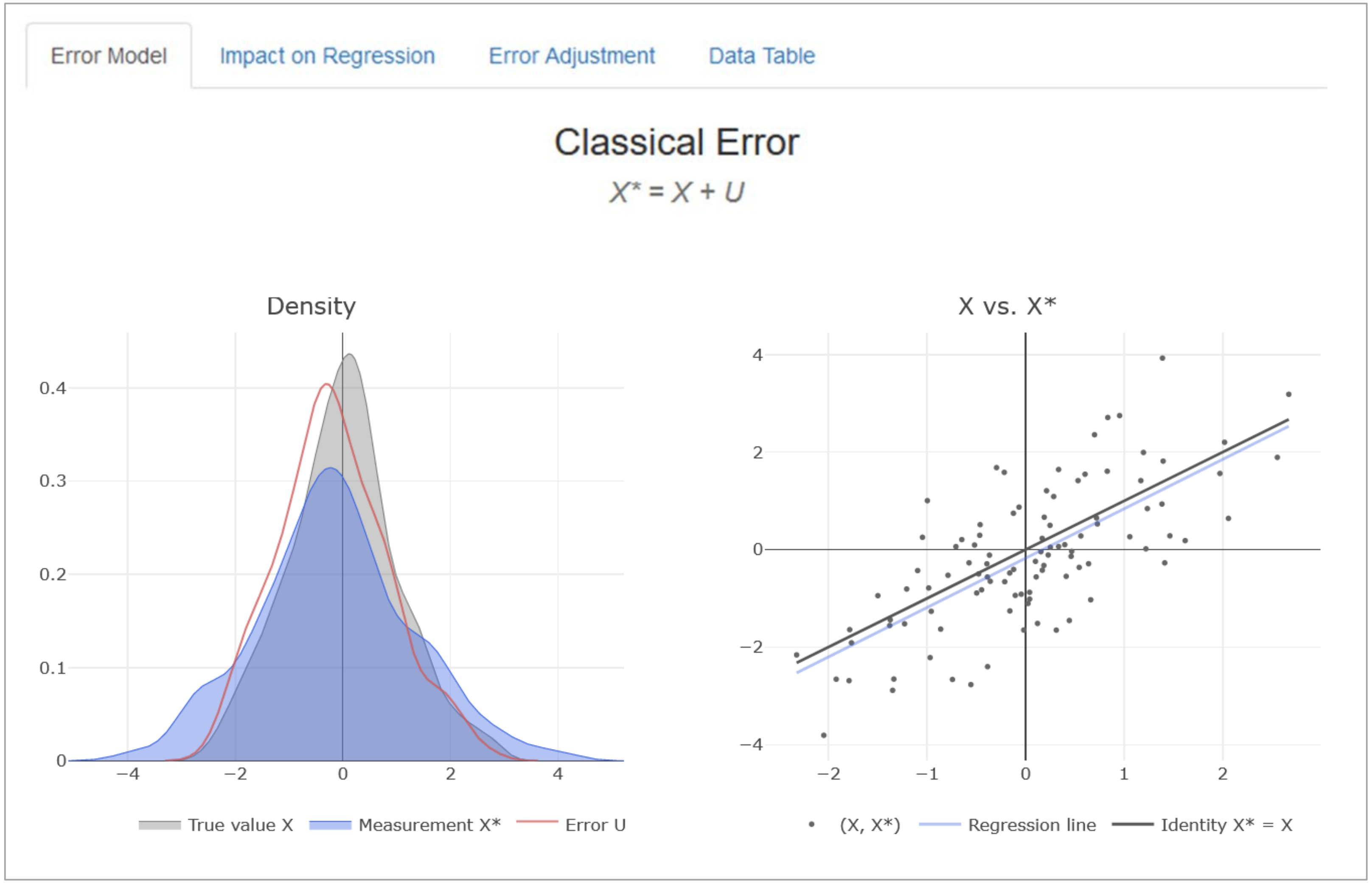
Covariate distribution

$$X \sim N(0,1)$$

Error model

- Additive, classical error
- Error variance





MEM-Explorer - Example

Sample size

$$n = 100$$

Outcome model

Simple linear regression model with

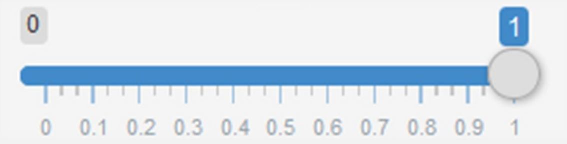
- coefficients $\beta_0 = 0$ and $\beta_1 = 1$
- error variance $\mathrm{Var}(\varepsilon) = 1$

Covariate distribution

$$X \sim N(0,1)$$

Error model

- Additive, classical error
- Error variance



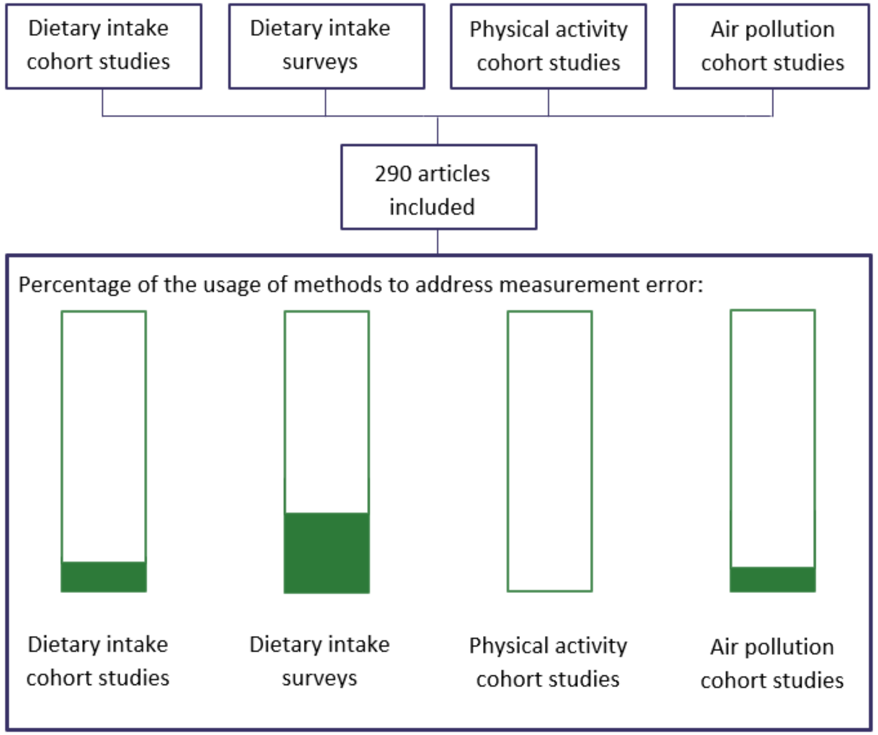
Literature survey of current practice (Shaw et al. 2018)

Aims:

- Assess current practice for addressing measurement error
- Identification of knowledge gaps

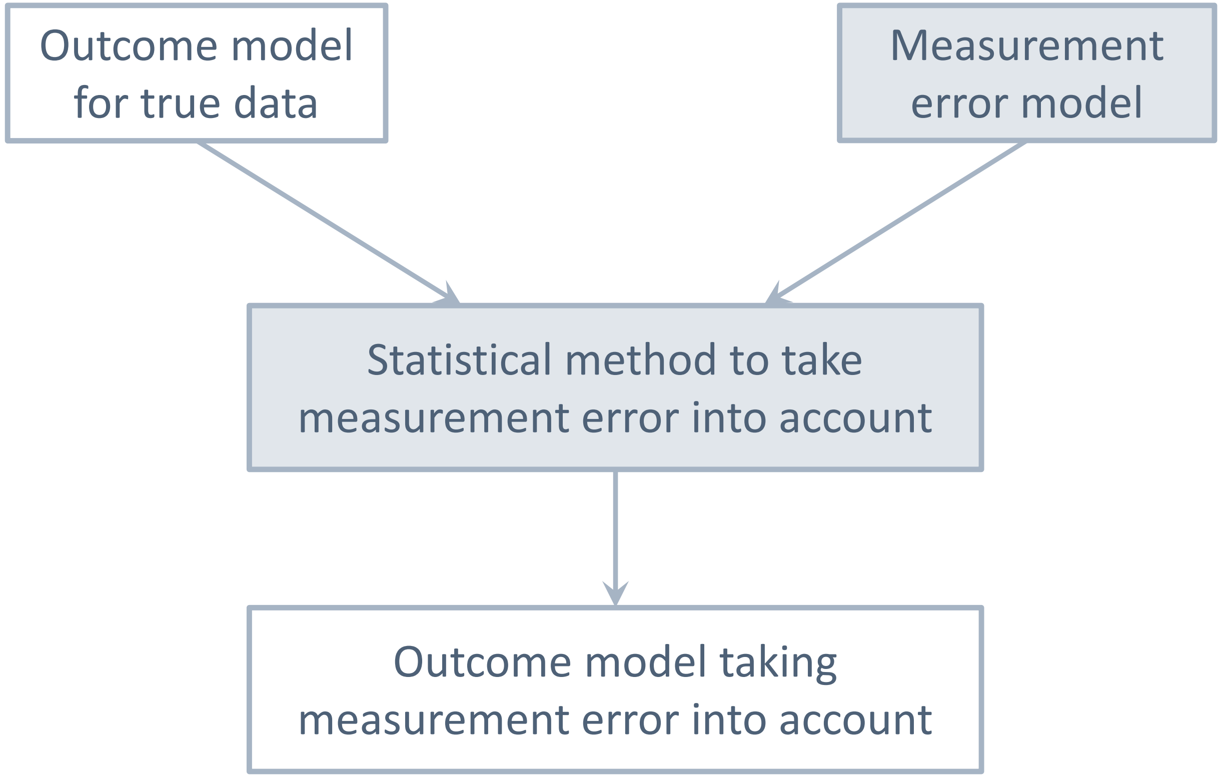
Results:

- Insufficient description of the measurement error
- Inadequate discussion of the impact of measurement error on the study results
- Several incorrect claims about the possible direction of the bias
- Rare use of methods which take measurement error into account
- Multiple error-prone exposures not acknowledged



Few published articles on observational epidemiologic studies include estimates that account for measurement error in exposure assessment.

Outcome model taking measurement error into account



Specific statistical methods are used to account for measurement error.

MEM-Explorer - Example

Sample size

$$n = 100$$

Outcome model

Simple linear regression model with

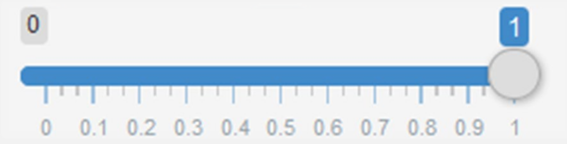
- coefficients $\beta_0 = 0$ and $\beta_1 = 1$
- error variance $\mathrm{Var}(\varepsilon) = 1$

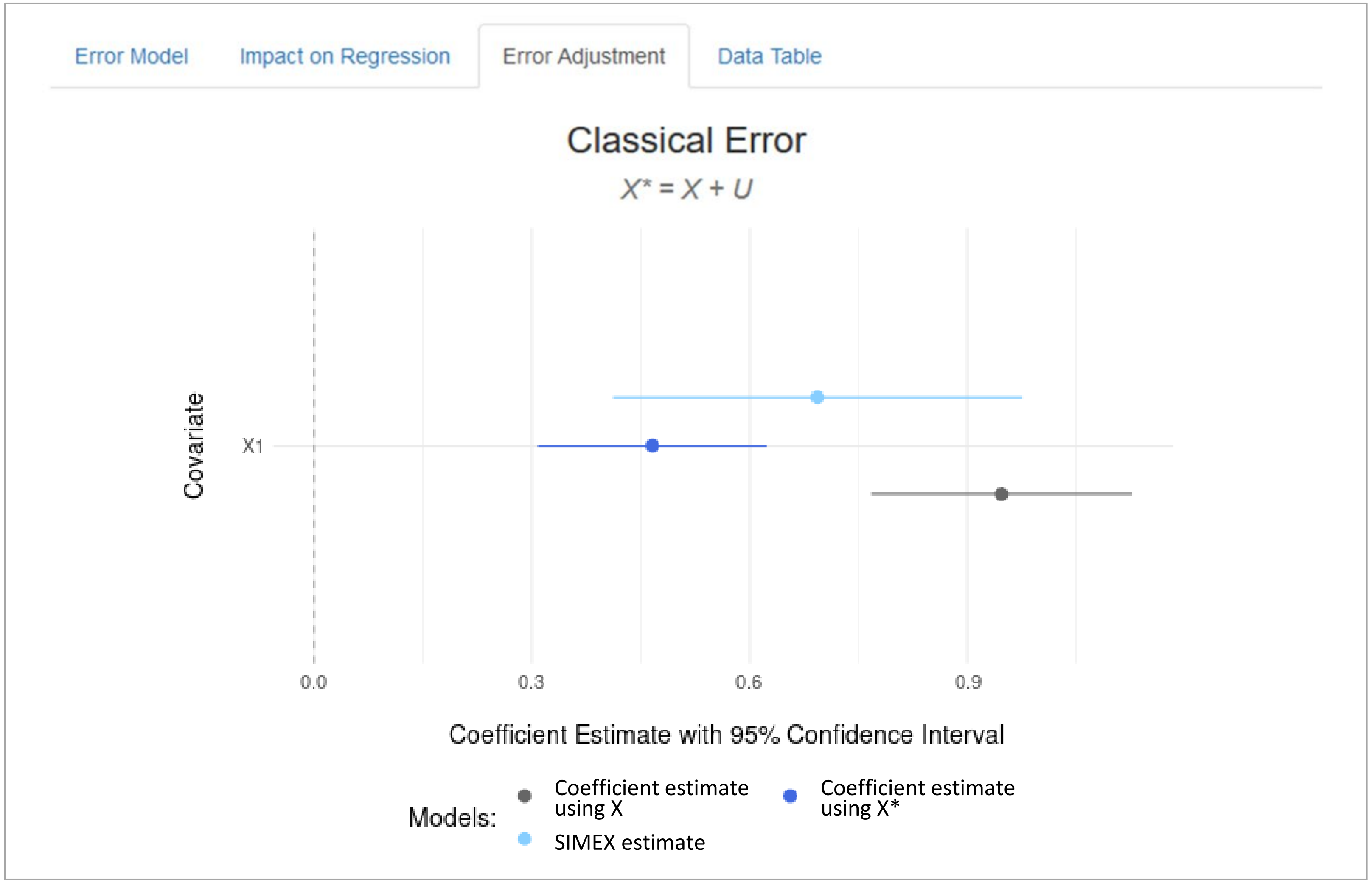
Covariate distribution

$$X \sim N(0,1)$$

Error model

- Additive, classical error
- Error variance





Key messages

- Few published articles on observational epidemiologic studies include estimates that account for measurement error. (Shaw et al. 2018)
- Accounting for measurement error is necessary because it can have an **impact** on study results.
- **Specific statistical methods** are used to account for measurement error. (Keogh et al. 2020, Shaw et al. 2020)
- **Additional information** about the magnitude of measurement error is required to account for measurement error.

Three steps of adequate treatment of measurement error and misclassification

1.

Consideration of potential measurement error at the **design stage** (sample size, ancillary studies)

2.

Explicit statement of **assumptions** regarding measurement error and exploration of its potential **impact** on the study results

3.

Application of analysis **methods** which take measurement error into account

Federal Office for Radiation Protection

STRATOS INITIATIVE

Legal Notice

Bundesamt für Strahlenschutz
Postfach 10 01 49
38201 Salzgitter

Tel.: +49 30 18333-0
Fax: +49 30 18333-1885
E-Mail: ePost@bfs.de

www.bfs.de

Contact for questions

Dr. Veronika Deffner
vdeffner@bfs.de
030/18333-2251