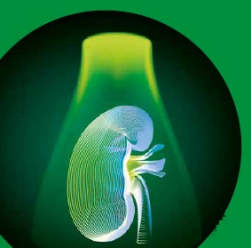


La metodologia statistica applicata al diabete

Gian Paolo Fadini

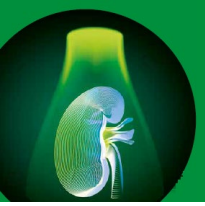
Professor of Endocrinology

University of Padova
Veneto Institute of Molecular Medicine

Disclosures

Il prof. Gian Paolo Fadini dichiara di aver ricevuto, negli ultimi 2 anni, compensi in qualità di relatore o per partecipazione a board scientifici da parte delle seguenti Aziende portatrici di interessi in ambito sanitario:

- Abbott
- AstraZeneca
- Boehringer
- Lilly
- Mundipharma
- Novo Nordisk
- Sanofi



Assunzione

Esclusi gli RCTs, la metodologia statistica nella ricerca clinica in diabetologia è rilevante soprattutto nell'ambito della…

Comparative effectiveness analysis

… ovvero la ricerca clinica comparative su dati routinari

(spesso big data sanitari)

Comparative effectiveness research

Innovative designs in randomized controlled trials	Causality in observational studies	Diagnostic tests, biomarkers and screening
Stratified or precision medicine	Big data & routinely collected data	Meta-analysis of randomised trials
Meta-analysis of diagnostic studies	Network meta-analysis	Evidence synthesis

Causality in observational studies

Randomized controlled trial

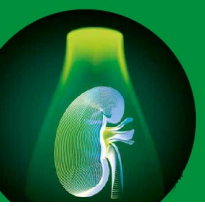
Lancet. 2019 Jan 19;393(10168):210-211

Comment

Real-world studies no substitute for RCTs in establishing efficacy

We live in the real world, so it is reasonable to expect that data collected from the real world should help identify effective therapies.

*Hertzel C Gerstein, John McMurray, Rury R Holman
Population Health Research Institute, McMaster University and Hamilton Health Sciences, Hamilton, ON L8S4K1, Canada (HCG); University of Glasgow, BHF Cardiovascular Research Centre, Glasgow, UK (JM); and Diabetes Trials Unit, University of Oxford, Oxford, UK (RRH)
gerstein@mcmaster.ca



Today's and tomorrow's aims

Today

Use, effectiveness & safety
In broader populations in routine practice

Challenge or confirm RCTs

Tomorrow

Use, effectiveness & safety
In broader populations in routine practice

Complement or substitute RCTs

Comparative effectiveness research

Caratteristiche imprescindibili

- Quesito scientific ilevante
- Definizione della o studio
- Endpoint ben def
- Valutazione della ti
- Uso dei migliori r

(Retrospective) data sources

Clinical charts
- Anthropometric data,
- Blood exams
- Complications
- Therapies
- … … …

Administrative data
- Delivered services
- Dispensed drugs
- Exemption codes
- Discharge codes
- Causes of death
- … … …

Clinical charts
- Adverse events
- Hypoglycaemia
- Compliance
- Adherence, satisfaction
- … … …

Administrative data
- All clinical data
- … … …

(Retrospective) data sources

Anthropometric data
Blood exams
Therapies
Comorbidities
Complications
… … …

+

Delivered services
Hospital discharge codes
Dispensed drugs
Causes of death
Costs
… … …

Issues of data quality

Types of error

- Instrument / analytical error
- Measure (e.g, SBP 140 in place of 138 mmHg)
- Reporting error (e.g. unit of measure)
- Coding error (e.g. wrong field)
- Extraction error (especially if manual)
- Identification error (es. database linking)
- … … …

Error cascade

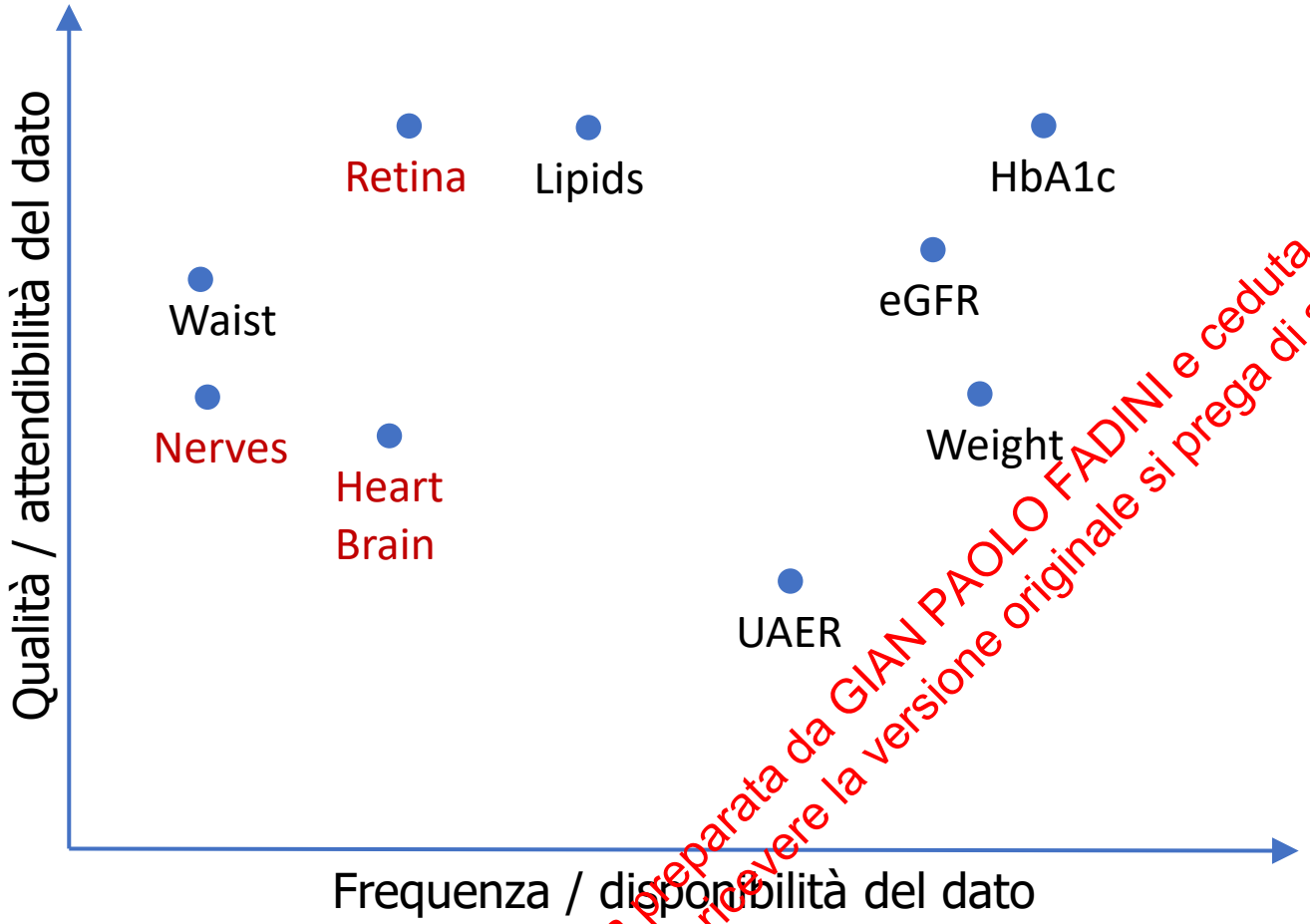
Issues of data quality

Clinical data → Clinical Repository → Study database

What determines data quality?

- *Why data are collected*
  - Clinic
  - Research
- *How data are collected*
  - Paper
  - Electronic
    - Direct input
    - Manual input
- *How data are extracted*
  - Manual extraction
  - Automatic extraction

Data quality in diabetes clinics



- Laboratory data have good availability and quality
- Categorical data can be used only when structured (e.g. ICD codes)
- Free text almost unusable

ML processing of free text

scientific reports

Deep-learning-based natural-language-processing models to identify cardiovascular disease hospitalisations of patients with diabetes from routine visits' text

Alessandro Guazzo[1], Enrico Longato[1], Gian Paolo Fadini[2], Mario Luca Morieri[2], Giovanni Sparacino[1] & Barbara Di Camillo[1,3]✉

Original dataset

Removed words

Processed dataset

Biases

- Confounding by indication

  Trattamenti diversi prescritti a pazienti diversi nel mondo reale

- Confounding by unmeasured variables

  Es. attività fisica, dieta, livello socio-economico, compliance, …

- Residual confounding

  Variabili non misurabili (es. attitudine del medico e del paziente nei confronti della malattia e della terapia)

Il solo metodo per <u>eliminare</u> questi fattori confondenti è la randomizzazione !!!

What works better?

Strategy A or Strategy B?

Strategies can be:

- 2 drugs
- 2 diets
- 2 surgical approaches
- 2 diagnostic tests
- 2 models of care
- …

Problems

- Only observational (big) data
- NO randomization
- Data not collected for research purposes
- Under free-living conditions

Doctors tend to use different «strategies» for different patients

↓

Group A and group B will be different

Approach to channelling bias

1) Adjust / Correct for differences between groups

→ Multivariable adjustment and similar procedures

2) Create a quasi-experimental condition

→ Propensity score matching and other PS-based approaches

JOURNAL OF THE AMERICAN COLLEGE OF CARDIOLOGY
VOL. 69, NO. 3, 2017

ISSN 0735-1097/$36.00
PUBLISHED BY ELSEVIER
http://dx.doi.org/10.1016/j.jacc.2016.10.060

Comparison of Propensity Score Methods and Covariate Adjustment

Evaluation in 4 Cardiovascular Studies

Markus C. Elze, PhD,[a,b] John Gregson, PhD,[a] Usman Baber, MD, PhD,[c] Elizabeth Williamson, PhD,[a] Samantha Sartori, PhD,[c] Roxana Mehran, MD, PhD,[c] Melissa Nichols, PhD,[d,e] Gregg W. Stone, MD, PhD,[d,e] Stuart J. Pocock, PhD[a]

Issues with multivariable regression

Problem

- Assumes linearity
- Co-linearity
- Too many variables increase C.I. reducing precision

→ **Too many assumptions !**

Note(s)

- Use splines
  - But how many splines and where?
- Spare variables
  - But which to sacrify?
- But sample size is large(r) because all cases are retained in the analysis.

Propensity score

- Probability of being assigned to a group (or the other) given covariates
- Ranges from 0 to 1
- Calculated from a logistic multiple regression
  - *Dependent variable* = group
  - *Independent variables* = all available variables

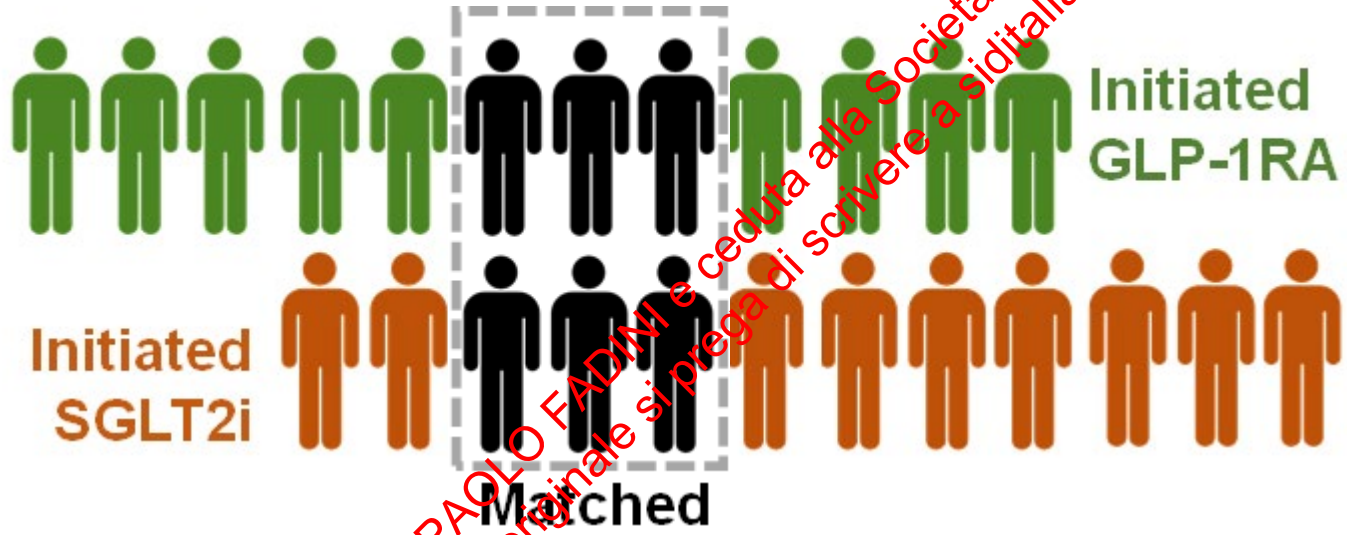
Propensity score matching

- Matches patients 1:1 (or 2:1, 3:1 etc…) on the same PS (within a certail tolerance, called caliper)

- When successfull, creates two groups of patients who are identical (on average) for all variables used to calculate PS

Propensity score matching

Patients differed for most clinical characteristics



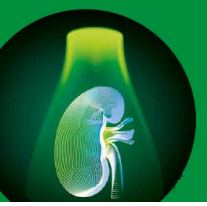
Matching created two cohorts with identical baseline clinical characteristics

N = 4298
Initiated SGLT-2i

N = 4298
Initiated GLP-1RA

Advantages of PSM

- Yields «identical» groups for all matching variables in most cases
- Patients have the same probability of being assigned to a group (pseudo-randomized , quasi-experimental setting)
- Can be done 1:1, 1:2, 1:3, ecc..
- Easy to be analyzed, presented and understood
- Very popular!



Issues with PSM

Problem	Note(s)

• Reduces sample size
  • When there is poor common support
	• Play with caliper or discard case studies with limited common support
	  • Cannot be applies always?
	  • When populations are highly different, is MVA any good, too?

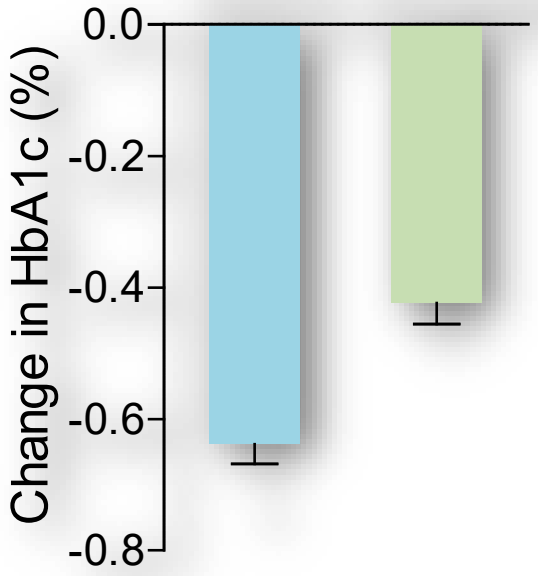
• Groups are similar only on average	• Resembles the trial populations?
• How to deal with matched pairs	• Paired or unpaired analysis?
• Use substitution?	• No, poorly clinically acceptable

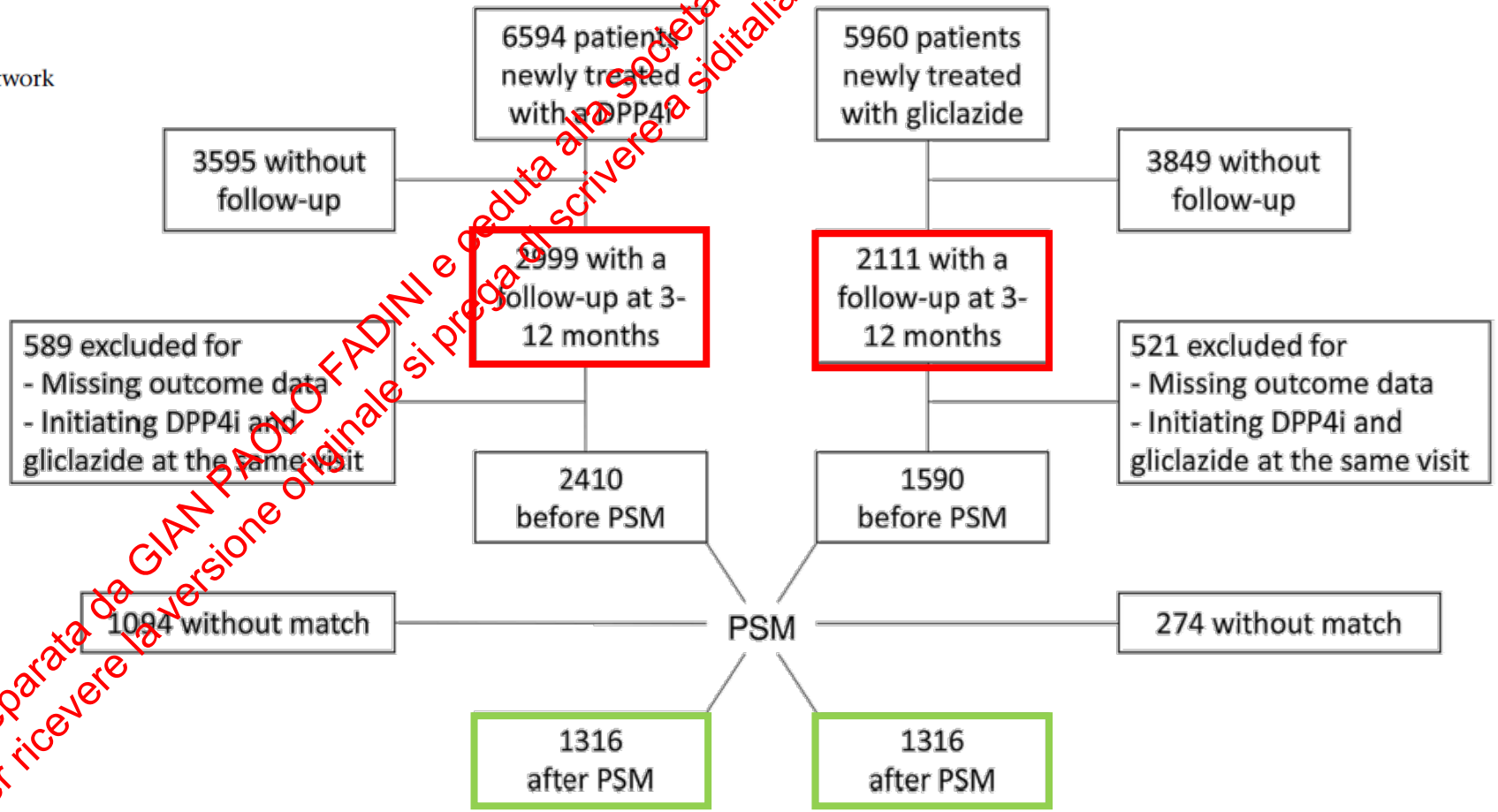
Diabetes Ther (2018) 9:1477–1490
https://doi.org/10.1007/s13300-018-0452-y

ORIGINAL RESEARCH

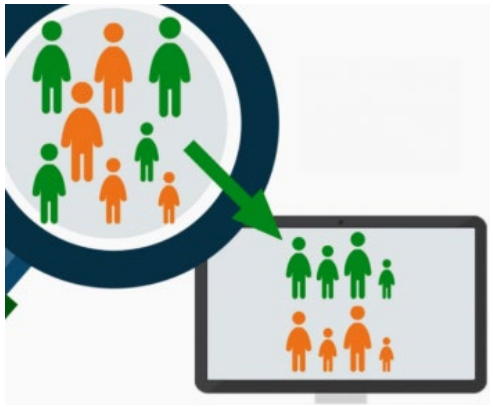
Comparative Effectiveness of DPP-4 Inhibitors Versus Sulfonylurea for the Treatment of Type 2 Diabetes in Routine Clinical Practice: A Retrospective Multicenter Real-World Study

Gian Paolo Fadini On behalf of the DARWIN-T2D Network





Unmeasur-ed/able variable bias



Measured

versus

Unmeasured

versus

Unmeasurable

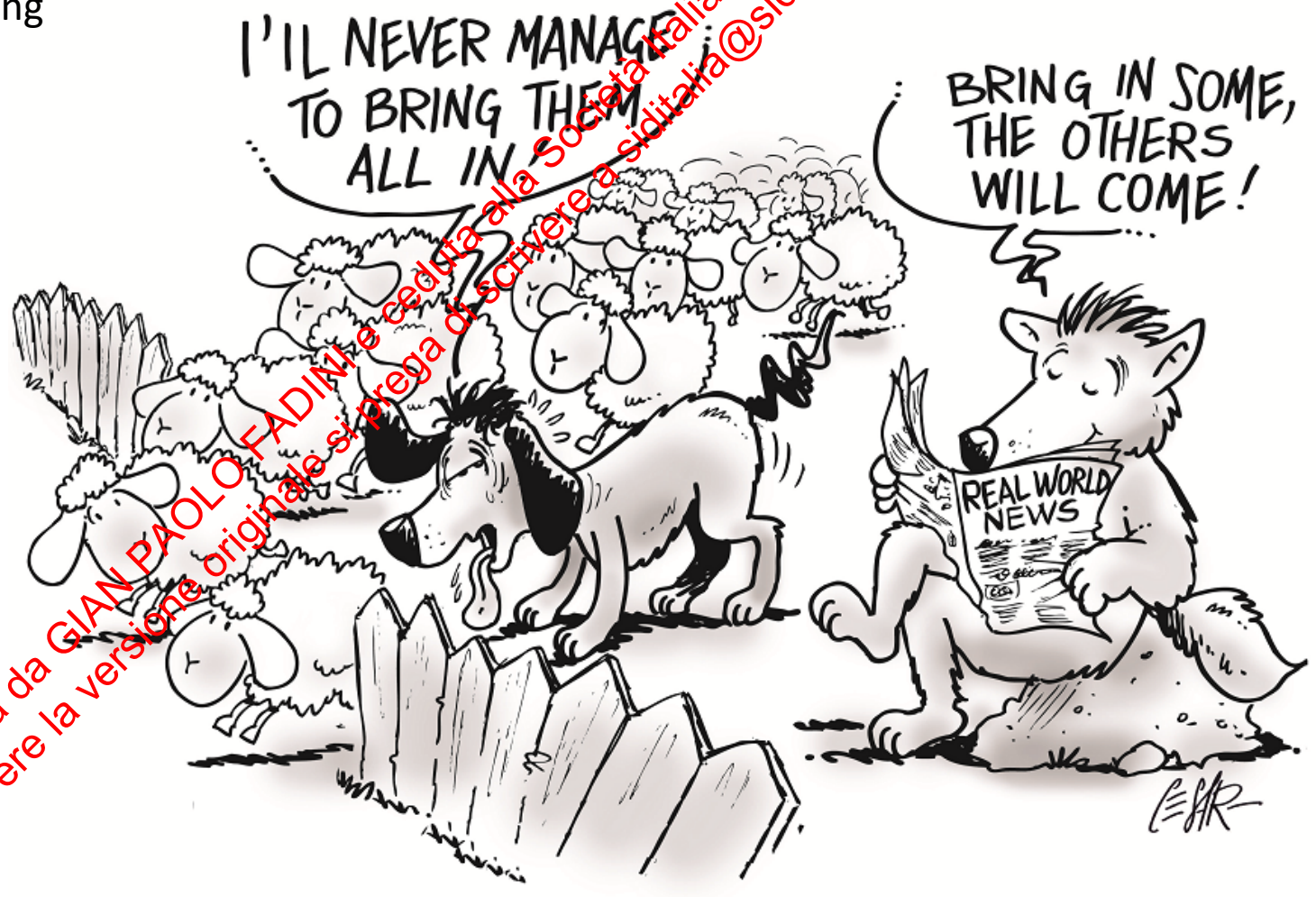
Variable	RCT	Real-world
Measured		
Age, sex	Y	Y
BMI, HbA1c, …	Y	Y / N
Unmeasured		
Socioeconomic status	Y	Y / N
Education	Y	Y / N
Pollution	Y	Y / N
Adherence	Y	Y / N
Unmeasureable		
Compliance	Y	N
Attitudes	Y	N

Unmeasur-ed/able variable bias

For how many variables one can adjust for, there will always be residual confounding driven by unmeasured and unmeasurable variables

Use proxy of variables that are difficult to measure

Define minimum sets of variables to be accounted for

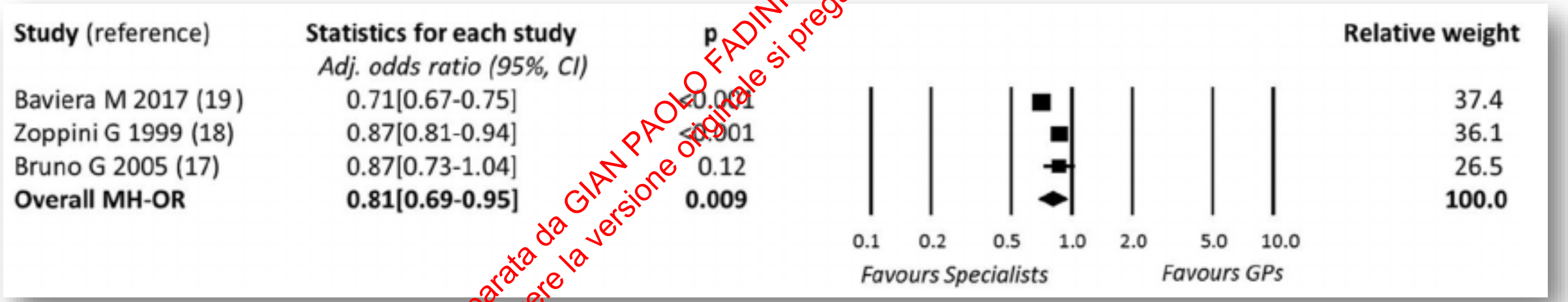


Controlling for the healthcare setting

META-ANALYSIS

Attending Diabetes Clinics is associated with a lower all-cause mortality. A meta-analysis of observational studies performed in Italy

E. Bonora [a,*], M. Monami [b], G. Bruno [c], G. Zoppini [a], E. Mannucci [b]



Reporting guidelines

STROBE

Strengthening the reporting of observational studies in epidemiology

https://www.strobe-statement.org/checklists/

STROBE Statement—Checklist of items that should be included in reports of *cohort studies*

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported
Objectives	3	State specific objectives, including any prespecified hypotheses
Methods		
Study design	4	Present key elements of study design early in the paper
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
Bias	9	Describe any efforts to address potential sources of bias
Study size	10	Explain how the study size was arrived at
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding
		(*b*) Describe any methods used to examine subgroups and interactions
		(*c*) Explain how missing data were addressed
		(*d*) If applicable, explain how loss to follow-up was addressed
		(*e*) Describe any sensitivity analyses