Causal language use in diabetes guidelines

An evaluation of causality and alignment

Thesis research master health sciences

Keling Wang[1], supvr.: Jeremy A. Labrecque[1]

[1]Department of Epidemiology, Erasmus MC Rotterdam

27 November 2024

Causal language use in guidelines

- Clinical practice guidelines provide recommendations and standards of care.
- Recommendations about treatment options must involve **causal language**.
- Guidelines do not produce evidence, so this causation:
  a) *either* comes from (lower-level) original evidence,
  b) *or* comes from common knowledge and expertise.

Erasmus MC
University Medical Center Rotterdam

Causation flows

- **"Producing"** causation: a causal effect is *identified* from data with assumption sets.
- **Expressing** causation: an effect is reported in statistical terms and numbers in study reports.
- **Perceiving, interpreting, passing, re-expressing**: an effect size is read by researchers and used as evidence, passed down to more advanced users, and re-expressed by others.
- **Reaching** end-users and **practicing**: the causation is finally read and interpreted by decision makers, and a decision is being made based on it.

Erasmus MC
University Medical Center Rotterdam

Causation jumps

- Causation jumps can happen at any place in this causation flow. Some new things are added to the flow, with or without justified evidence.
- From non-causal to causal statement is a causation jump;
- Jumps are **not necessarily** harmful, and are required when moving from data to practice..., there are safe landing postures!

Causation jumps (cont'd)

Safe jumps:

- With assumptions that hold or are plausible, interpreting an association as a causal estimate. (*identification*).
- Adding justified existing domain-specific common causal knowledge. (*but better with the help of causal inference people*)

Dangerous jumps:

- Some new things are arbitrarily added to the flow without evidence, or an expression is incorrectly perceived and interpreted.

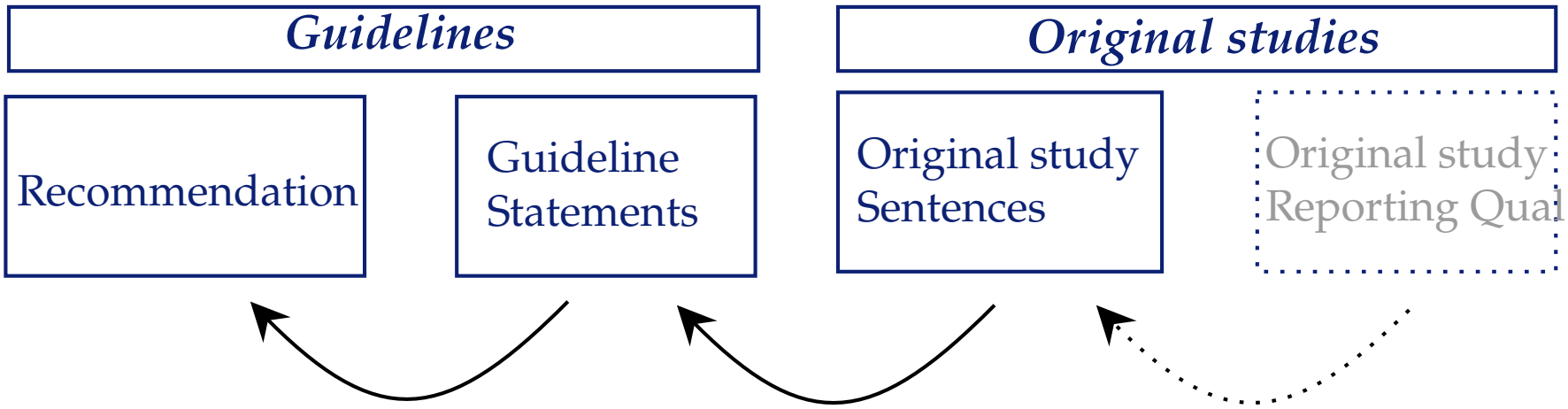
Erasmus MC
University Medical Center Rotterdam

Causation flow and jumps in guidelines

And when developing guidelines...

Causation flow and jumps in guidelines (cont'd)

A causation flow starts from **[1]** original evidence, goes through **[2]** supporting statements (guideline texts), and ends at **[3]** action recommendations.



Causation flow and jumps in guidelines (cont'd)

- Wherever not aligned, a **causation jump** will present.
- Almost all causation jumps in guidelines can be dangerous:
  - Guidelines do not conduct new studies and cannot access the patient data.
  - No causation can be produced.
  - Causal inference expertise is rarely involved to help a safe jump.

Use of causal language

- Causation is expressed, carried, and also perceived through words, *a.k.a.* causal language.
- Checking how causation is expressed through causal language is **the only possible way** to evaluate causation flows in guidelines.
- We mainly consider assessing the *expressed* strength of causation in causal language.

Aims

Investigate the form, the prevalence, and the pattern of

causation jumps

in clinical practice guideline causation flows,
via evaluating the use of causal language,
in the non-pharmacological part of clinical practice guidelines
about type 2 diabetes management.

Erasmus MC
University Medical Center Rotterdam

Methods overview

A1 Rate the **expressed strength of causation** in guideline statements, original study abstracts, and original study main-texts, and compared between them to check causation jumps (weighted by content-alignment).

A2 Rate the **expressed causation-dependence** in guideline recommendations, and check whether they can be effectively supported by causal statements.

Methods overview (cont'd)

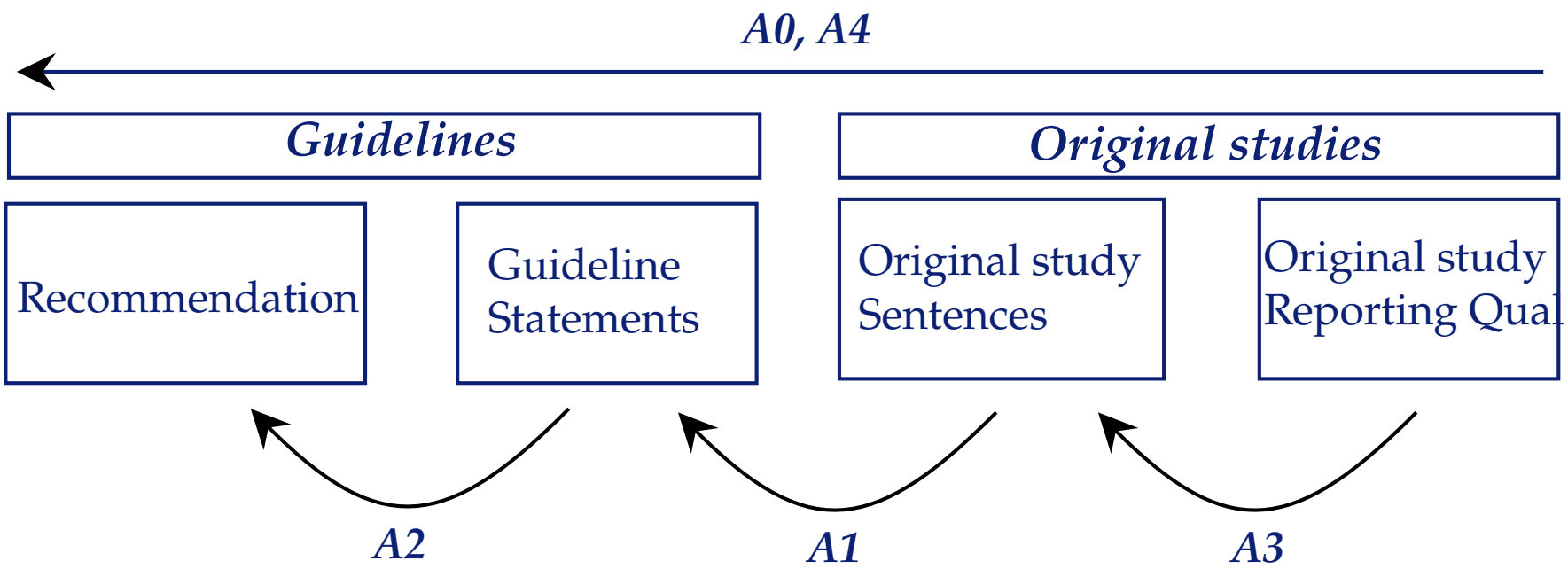
A3 Rate the reporting quality of necessary information for a causal interpretation in original studies. **Target Trial Emulation** component framework is used.

A4 Illustrate the causation flows and jumps, and explored the correlations among the scores.

Methods overview (cont'd)

Causation	Dependence	Content alignment	TTE reporting
3 – Strong	3 – Strong		
2 – Moderate	2 – Moderate		2 – Fully reported
1 – Weak	1 – Weak	1 – Exactly same	1 – Partially
		0.5 – Mentioned	
0 – Correlational	0 – Independent	0 – Not relevant	0 – Problematic
-1 – No relat.	NA – No rec.		

Haber et al. [1], Smit et al. [2], and Hernán et al. [3]



A0 Linking word frequency

A1 Causation strength and alignments

A2 Causation-dependence

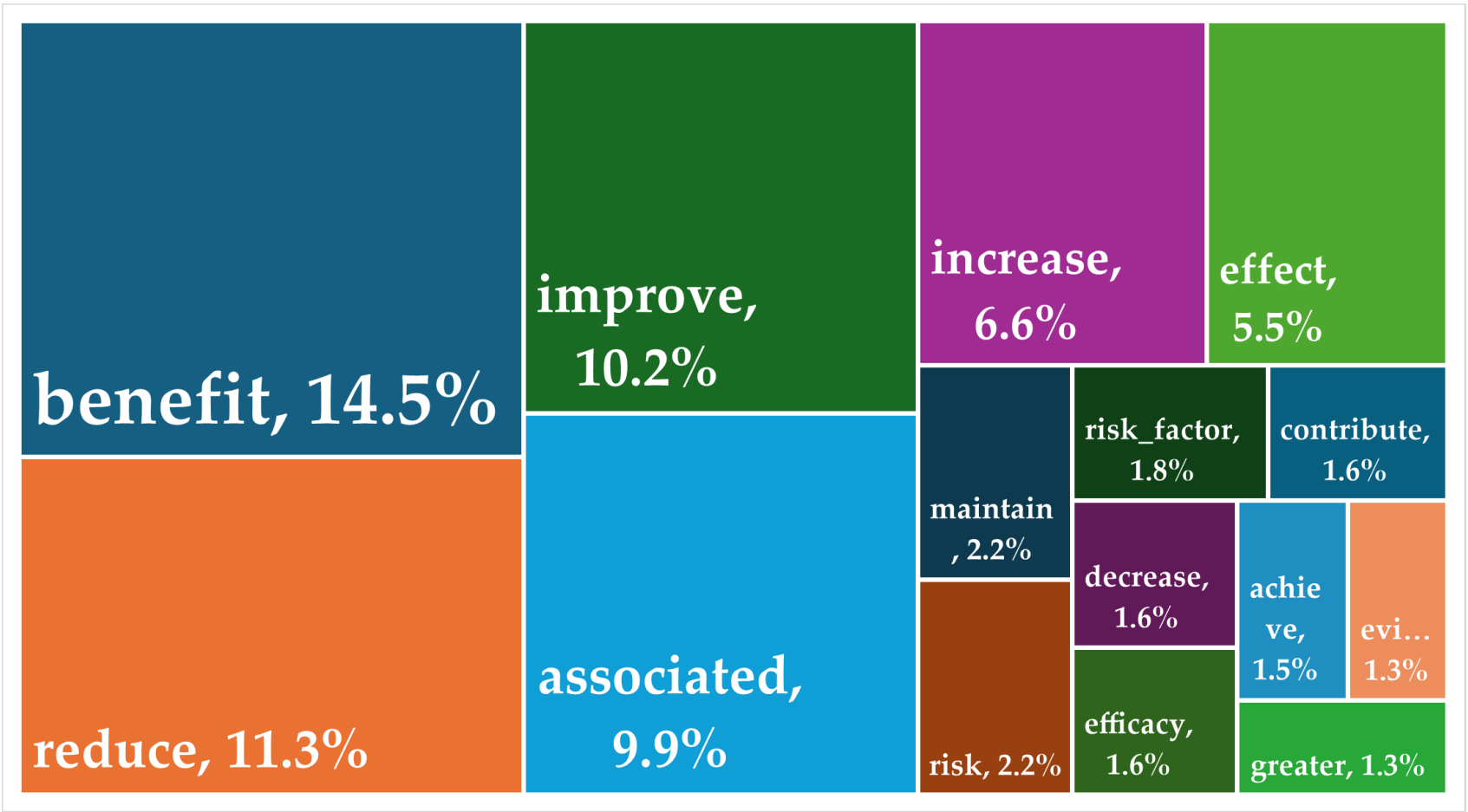
A3 TTE components

A4 Flows of causation in guidelines

Conclusion 0: Linking words

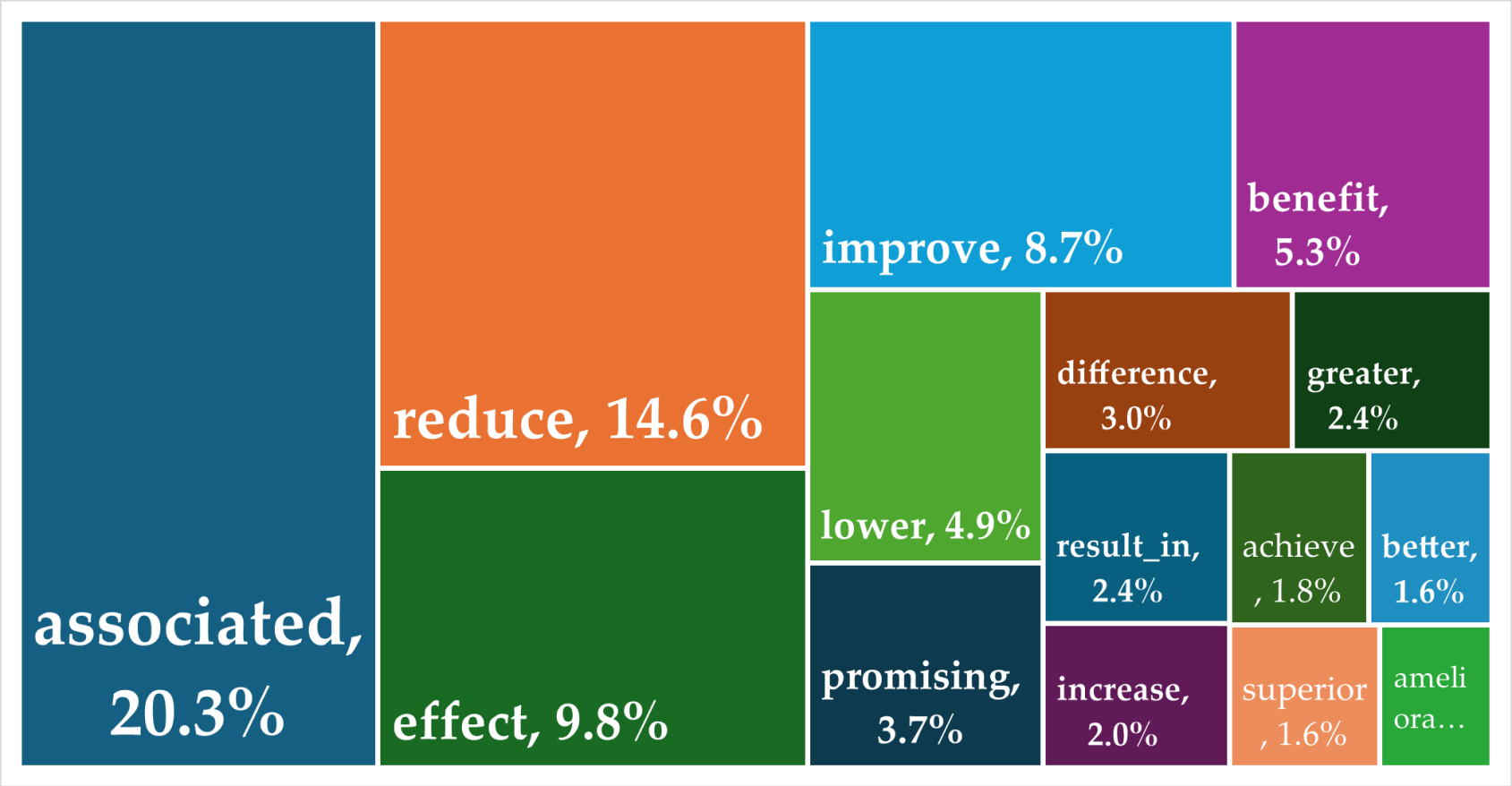
A0 $\Rightarrow$ Conclusion 0 "Associated" is used the most in original study reports to express causation. Guideline makers often avoided using this word and changed to stronger words.

Linking words in causal statements (guideline)



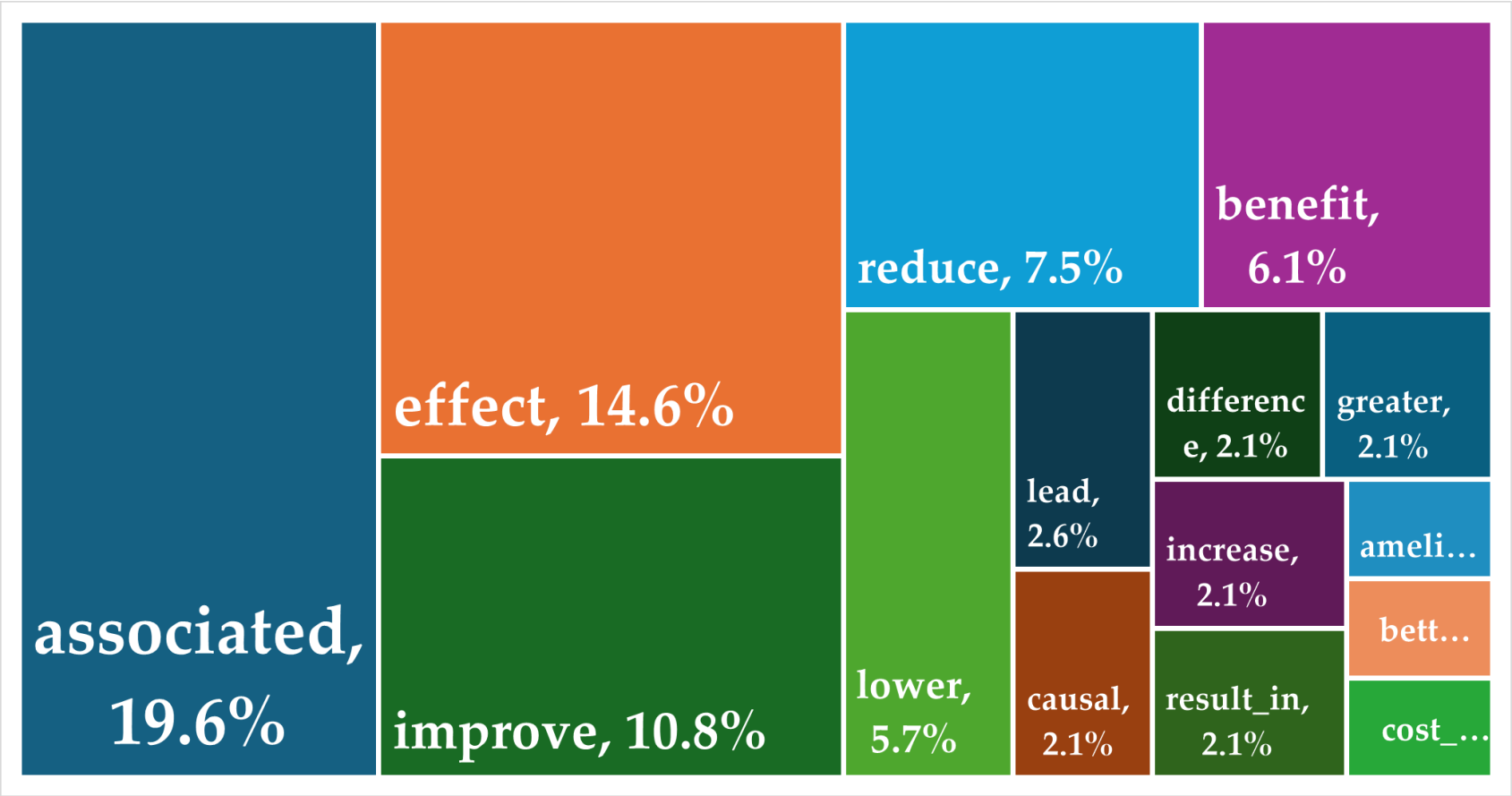
Erasmus MC
University Medical Center Rotterdam

Linking words in causal statements (OS abstract)



Erasmus MC
University Medical Center Rotterdam

Linking words in causal statements (OS main text)



Erasmus MC
University Medical Center Rotterdam

Conclusion 1: Causation strength

A1 ⇒ Conclusion 1 Expressed strength of causation was distributed bimodal and skewed. A single 4-item scale cannot capture all the features of causation strength in sentences.

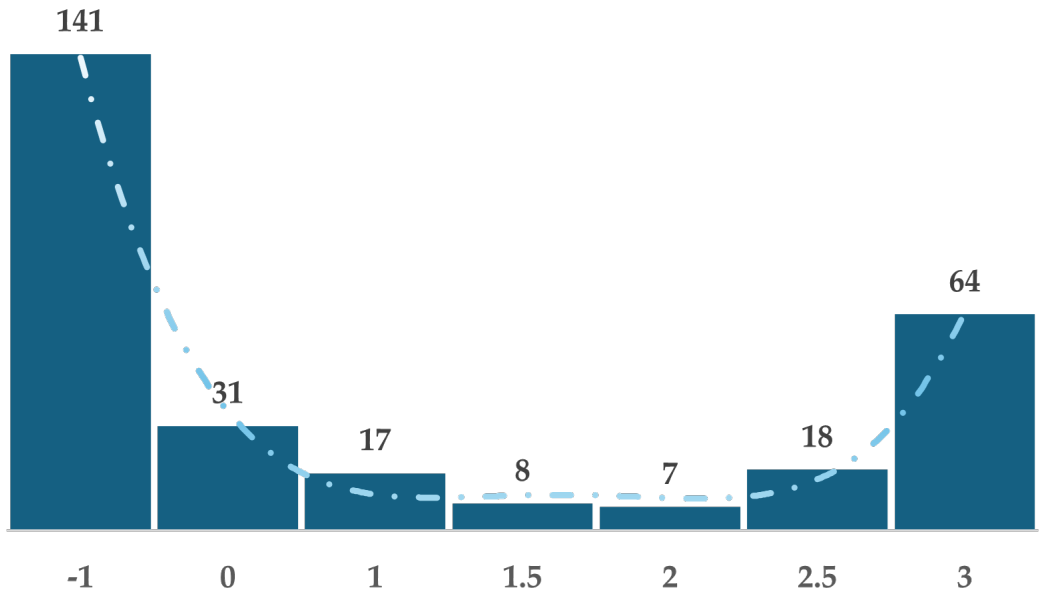
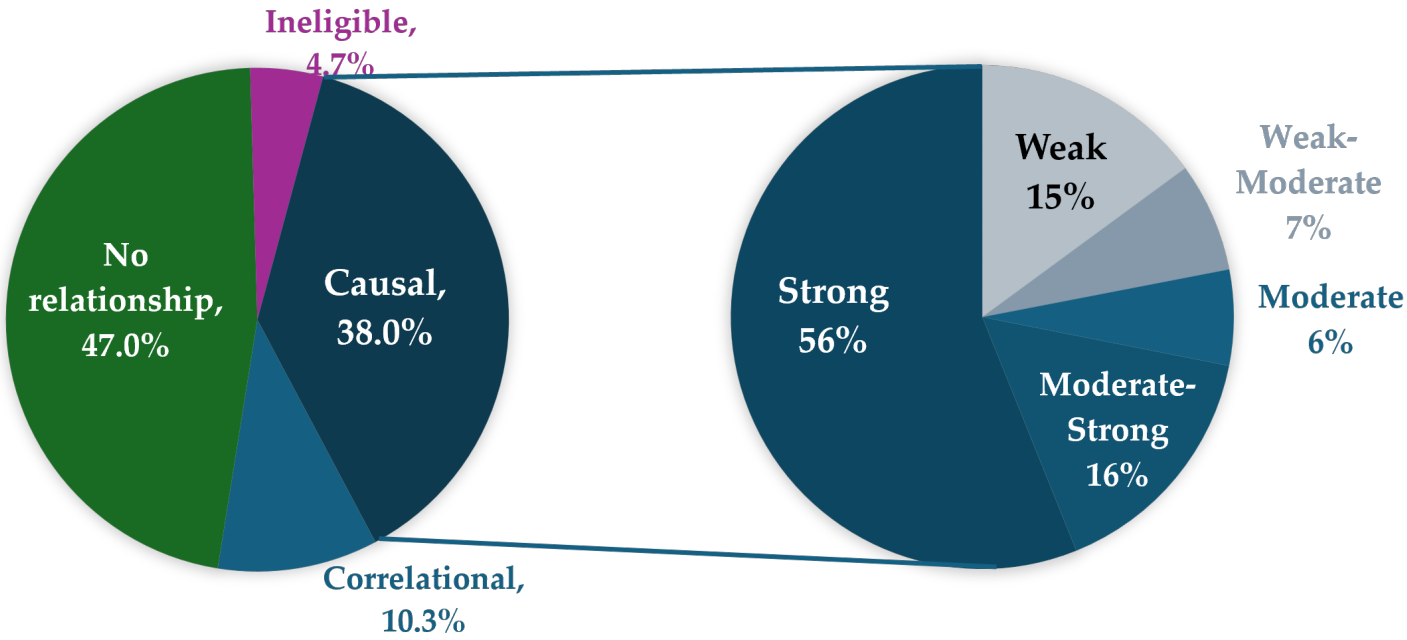
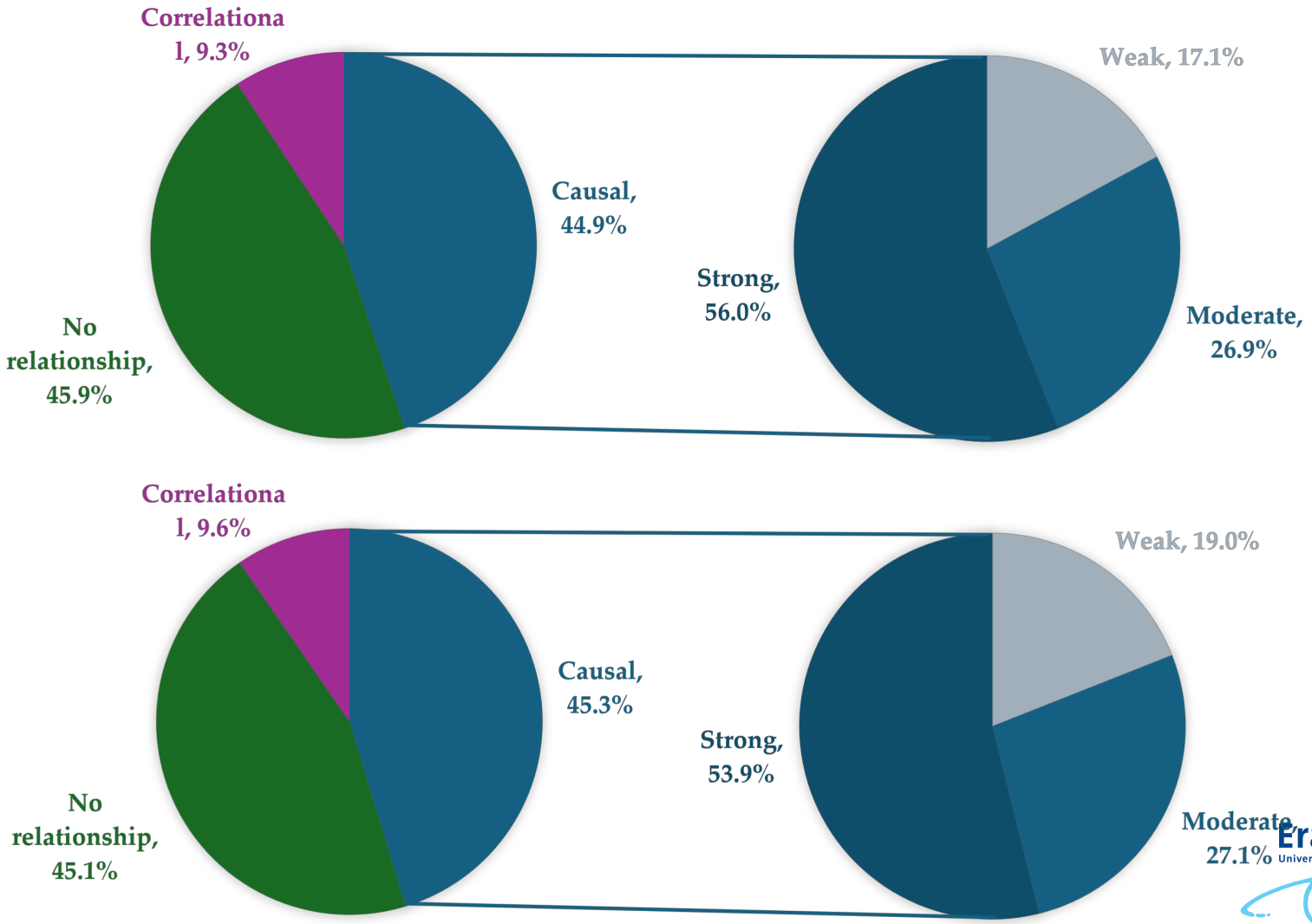


Figure 1: Distribution of causation strength in guidelines

Causation strength in 300 guideline statements (manually)



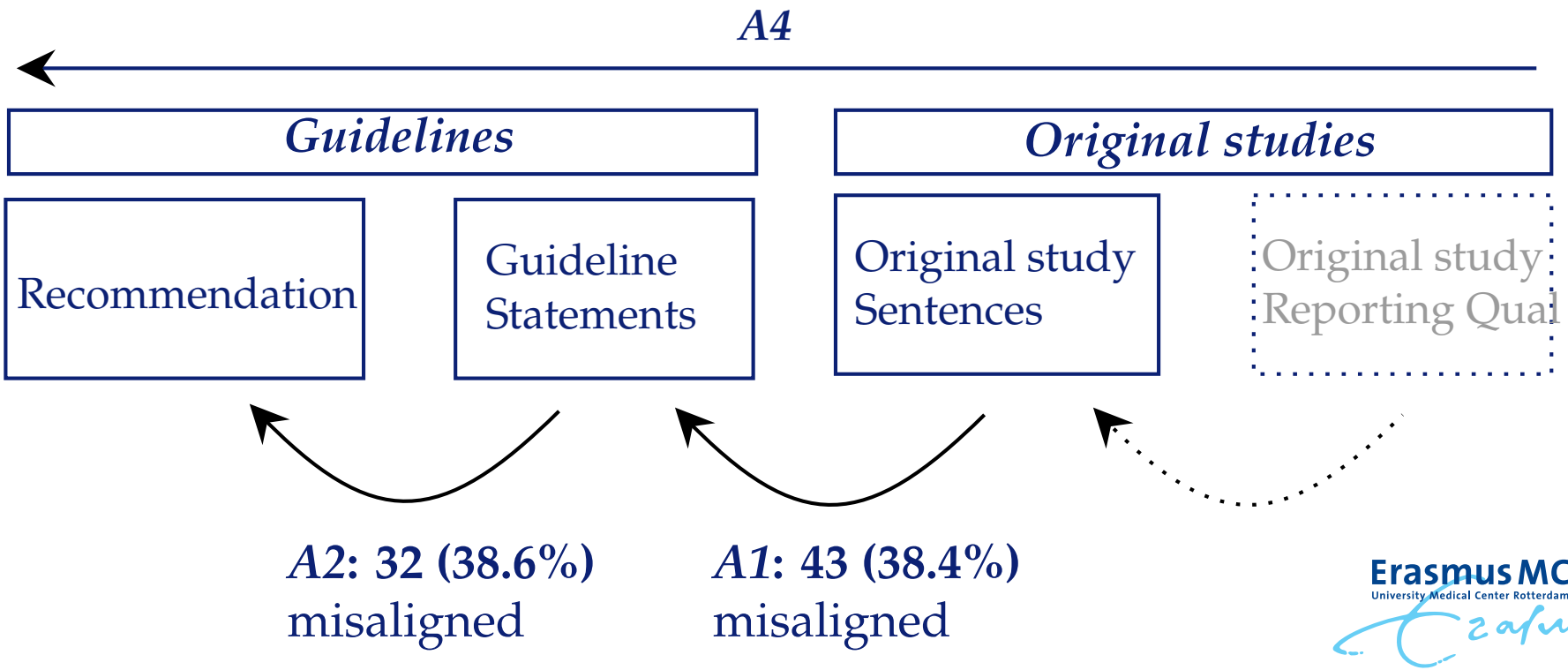
Causation strength in all statements (automated)



Erasmus MC University Medical Center Rotterdam

Conclusion 2: Causation jumps

A1+A2+A4 ⇒ Conclusion 2 Causation jumps were largely prevalent in guidelines. For both [recs – statements] and [statements – original studies], **more than 1/3** of the connections were misaligned.



Causation strength and alignments

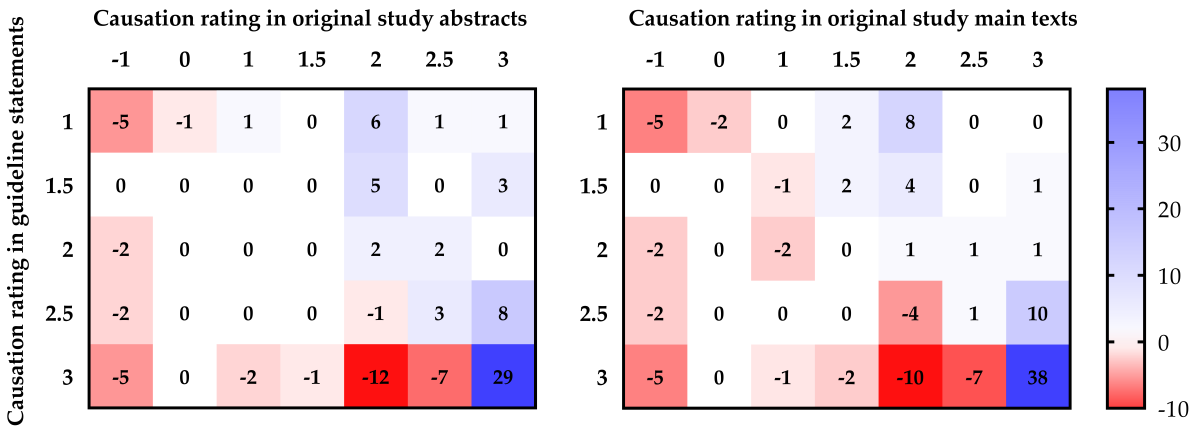


Figure 2: Alignment of causation strength ratings

38.6% (38 out of 99 *w.r.t.* abstracts) and **38.4%** (43 out of 112 *w.r.t.* main text) of the statements showed causation jumps.

Content-alignment weighted correlations ($\rho = 0.186; 0.344$) were almost half of the unweighted ones ($\rho = 0.332; 0.484$).

Erasmus MC
University Medical Center Rotterdam

Causation-dependence and alignments

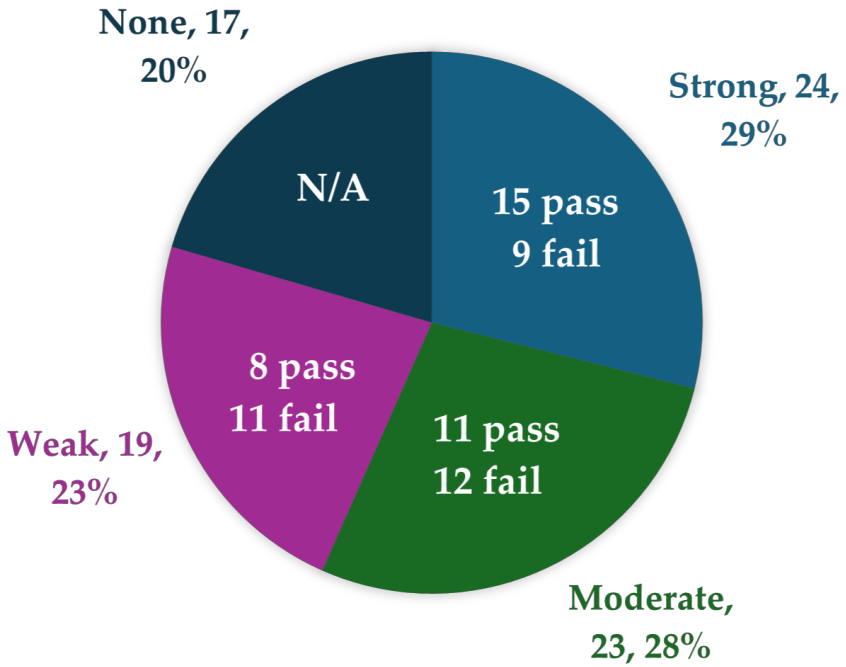


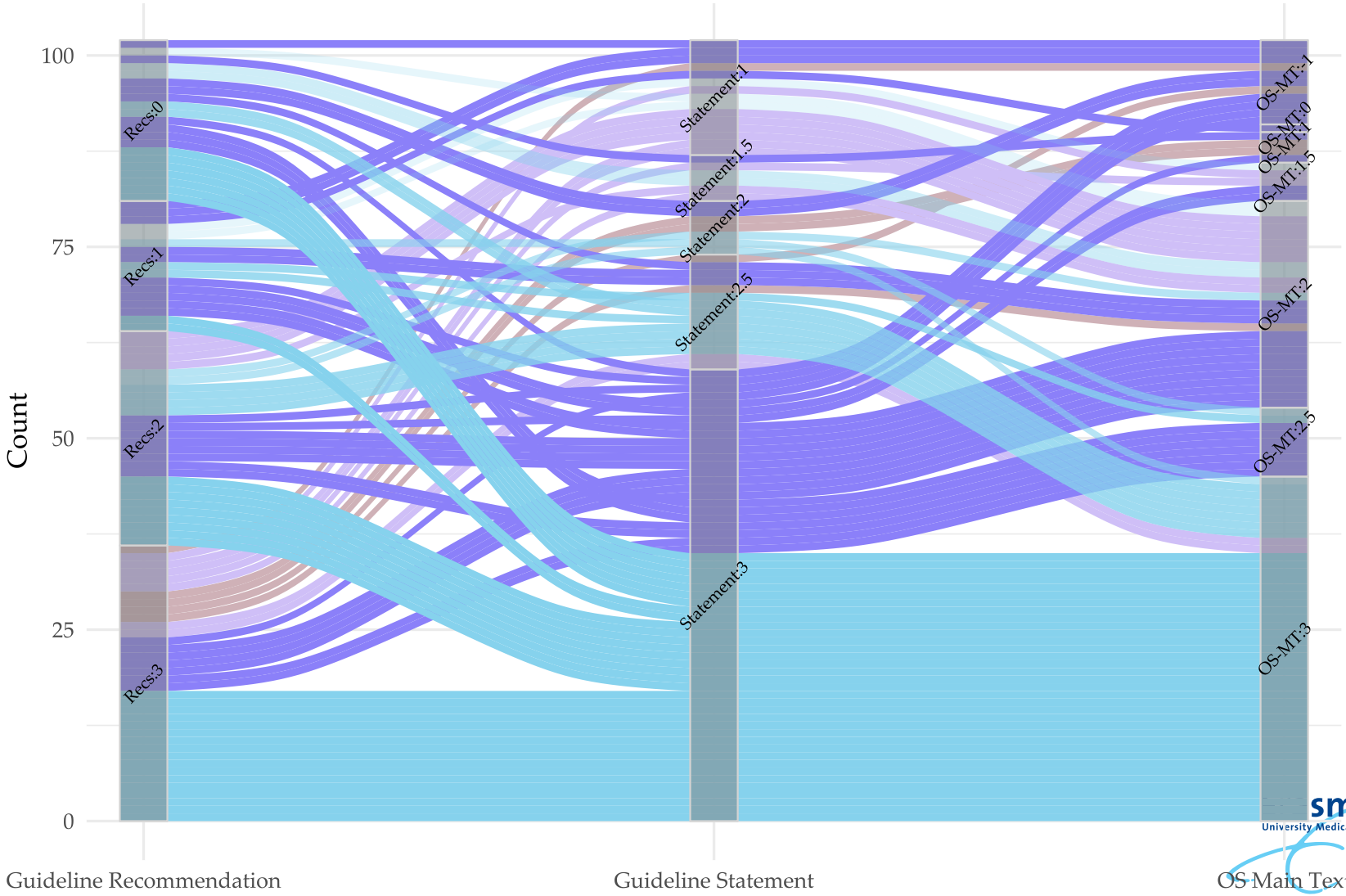
Figure 3: Causation-dependence of guideline recommendations

Of 83 evaluated guideline recommendations, **32 (38.6%) cannot be effectively supported** by their surrounding guideline statement:

There was **not any** guideline statement with its causation strength higher than causation-dependence of these guideline recommendations.

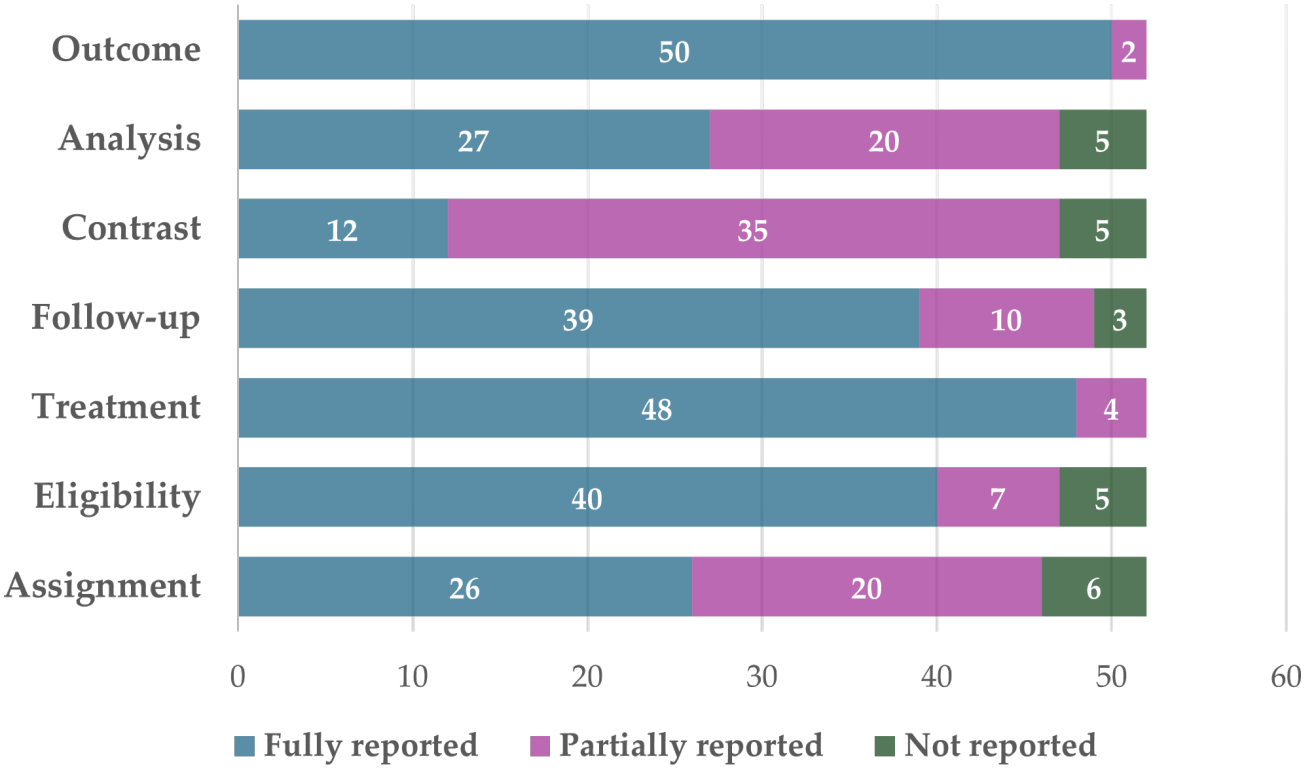
Erasmus MC
University Medical Center Rotterdam

Causation flows



Conclusion 3: TTE components

A3 ⇒ Conclusion 3 Causal contrasts and treatment assignment procedures were often ignored. RCTs reported TTE components clearer than non-RCTs, but were still not clear enough on all dimensions.



RCT vs. non-RCT reporting quality

RCT (m. 12.0/14, IQR 11.3 – 13.0) *vs.*
non-RCT studies (m. 10.0/14, IQR 9.00 – 11.0):
Wilcoxon test: locdiff = 2.00 [95%CI 1.00, 3.00], $P = 0.0002$

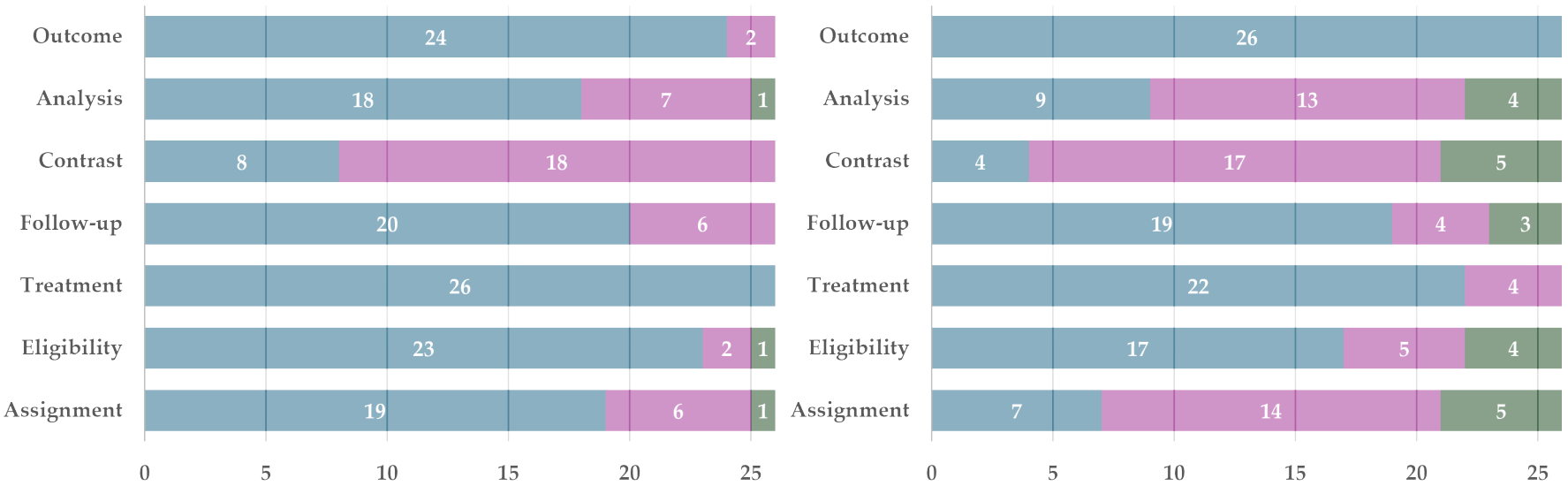
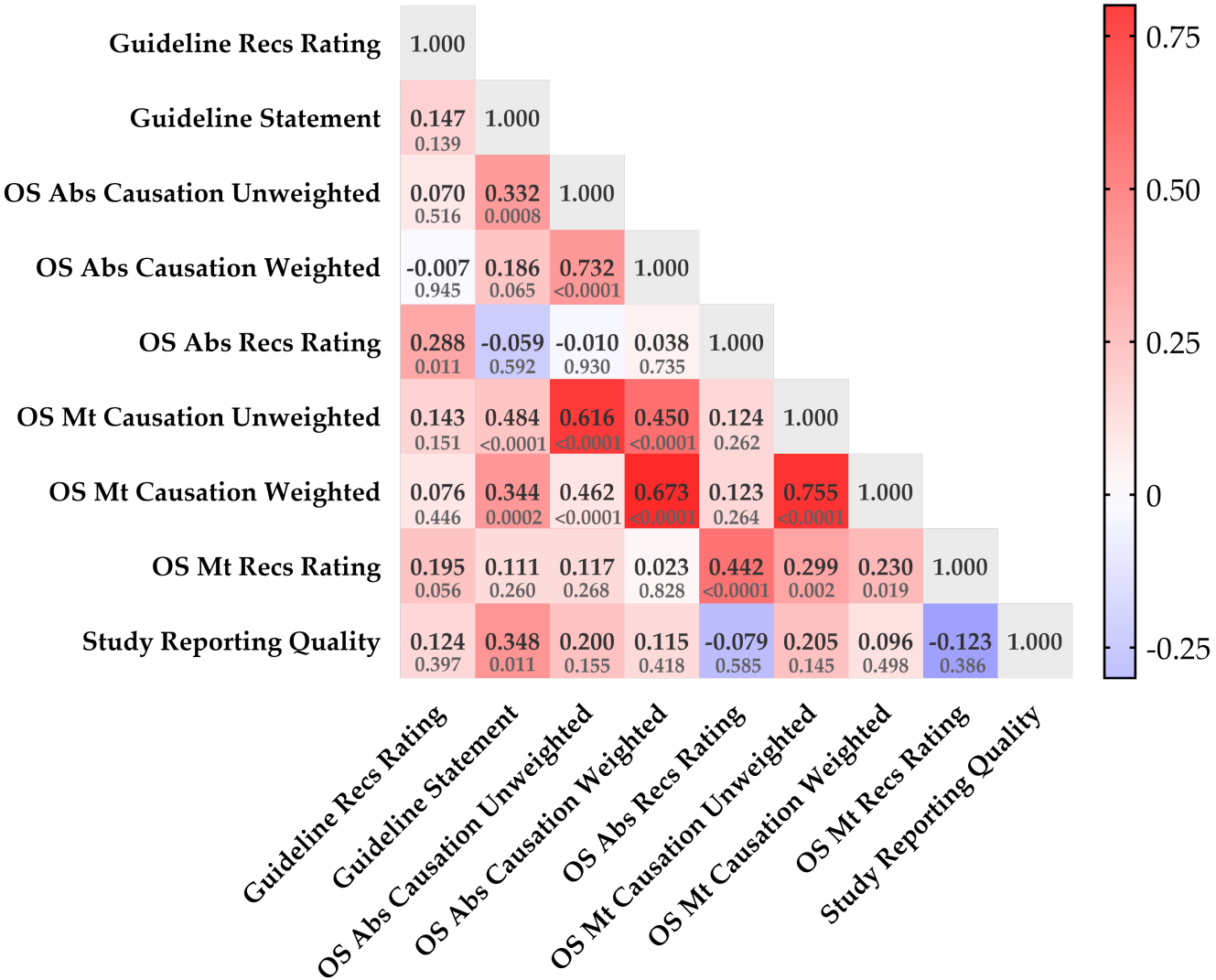


Figure 5: TTE component rating in RCTs and non-RCTs

Conclusion 4: Rephrasing guideline statements and recs

A1+A3+A4 ⇒ Conclusion 4 Guideline makers seemed to rephrase their expression of causation not only based on study conclusions, but also their reporting quality and study types. Conclusive sentences from original study main texts were more relevant to guidelines, but not abstracts.

Correlation among all ratings



Core conclusion

One third of the guideline statements were **causal**; among them, **causation jumps** were found in more than **one third** of the causation flows in the guidelines.

Implications

Guideline makers need to re-consider their wording choices, jump less, and jump more safely if inevitable. Researchers with causal inference expertise would help on this.

Erasmus MC
University Medical Center Rotterdam

Erasmus MC
University Medical Center Rotterdam

Progress of the thesis project and electives

- This is a novel study; my supervisor and I came up with the idea together. We had regular meetings across the whole lifespan of this project.
- I did most of the work in setting up, data collection, selection and analysis. My supervisor helped rate guideline statements and recommendations. Chang Wei (PhD candidate) helped all other ratings and solve divergences.
- Acknowledgment to prof.dr. Schram and dr. Wolters for their help and feedback.
- Most of the knowledge sets came from outside the NIHES electives. (Thanks to Miguel Hernán, AJE and Hugging Face)

Erasmus MC
University Medical Center Rotterdam

Other contributions

Published articles:

- Li L, Pan H, **Wang K***, et al. Factors influencing overdose and misuse of gender-affirming medication in Chinese transgender and gender diverse individuals: a qualitative study of experience and perspectives. ***International Journal of Transgender Health*** 2024. doi: 10.1080/26895269.2024.2316693.
    - Ranked #1 in transgender health. Co-correspondence.

Other contributions (cont'd)

Working projects:

- Formulating causal questions: causal estimands for everyone. (manuscript in preparation)
- Effect heterogeneity of screening for ovarian cancer prevention: a counterfactual prediction modelling study. (ongoing)
- Missing data mechanisms under potential outcome framework: a review (ongoing)
- Bounding the magnitude of effect heterogeneity using aggregated level data (ongoing)

Erasmus MC
University Medical Center Rotterdam

Questions and defense

Core conclusion

One third of the guideline statements were **causal**; among them, **causation jumps** were found in more than **one third** of the causation flows in the guidelines.

Implications

Guideline makers need to re-consider their wording choices in text, jump less, and jump more safely if inevitable. Causal inference expertise would help on this.

Erasmus MC
University Medical Center Rotterdam

Erasmus MC
University Medical Center Rotterdam

1. Haber NA, Wieten SE, Rohrer JM, et al. Causal and Associational Language in Observational Health Research: A Systematic Evaluation. American Journal of Epidemiology 2022;191:2084–97.
2. Smit JM, Krijthe JH, Kant WMR, et al. Causal inference using observational intensive care unit data: a scoping review and recommendations for future practice. npj Digital Medicine 2023;6:1–11.
3. Hernán MA, Wang W, and Leaf DE. Target Trial Emulation: A Framework for Causal Inference From Observational Data. JAMA 2022;328:2446–7.

Erasmus MC
University Medical Center Rotterdam

Limitations

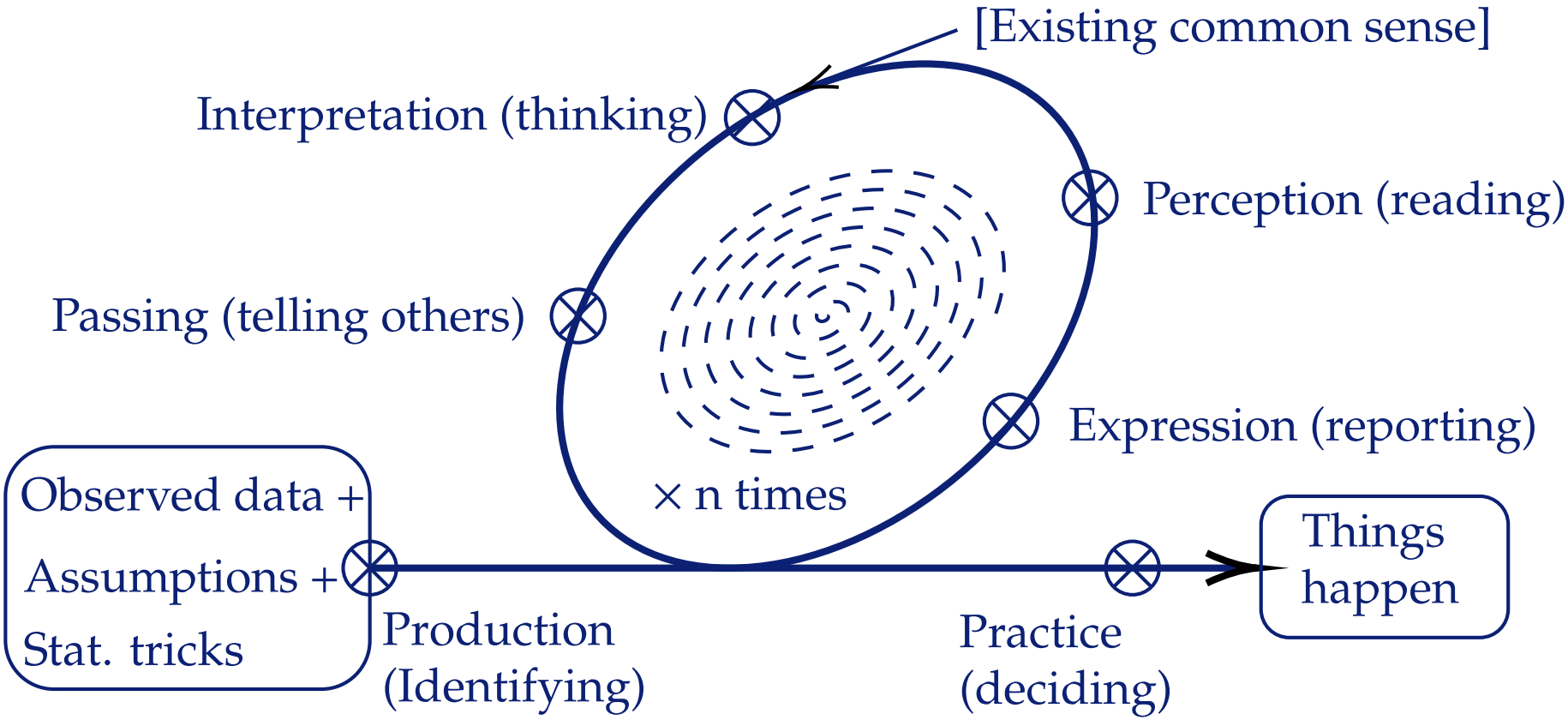
1. Only type 2 diabetes-related guidelines were evaluated.
2. The sample size (300 in total, 114 causal statements) was small.
3. Inter-rater divergences were frequent.
4. Only one study for one guideline statement was sampled and evaluated.

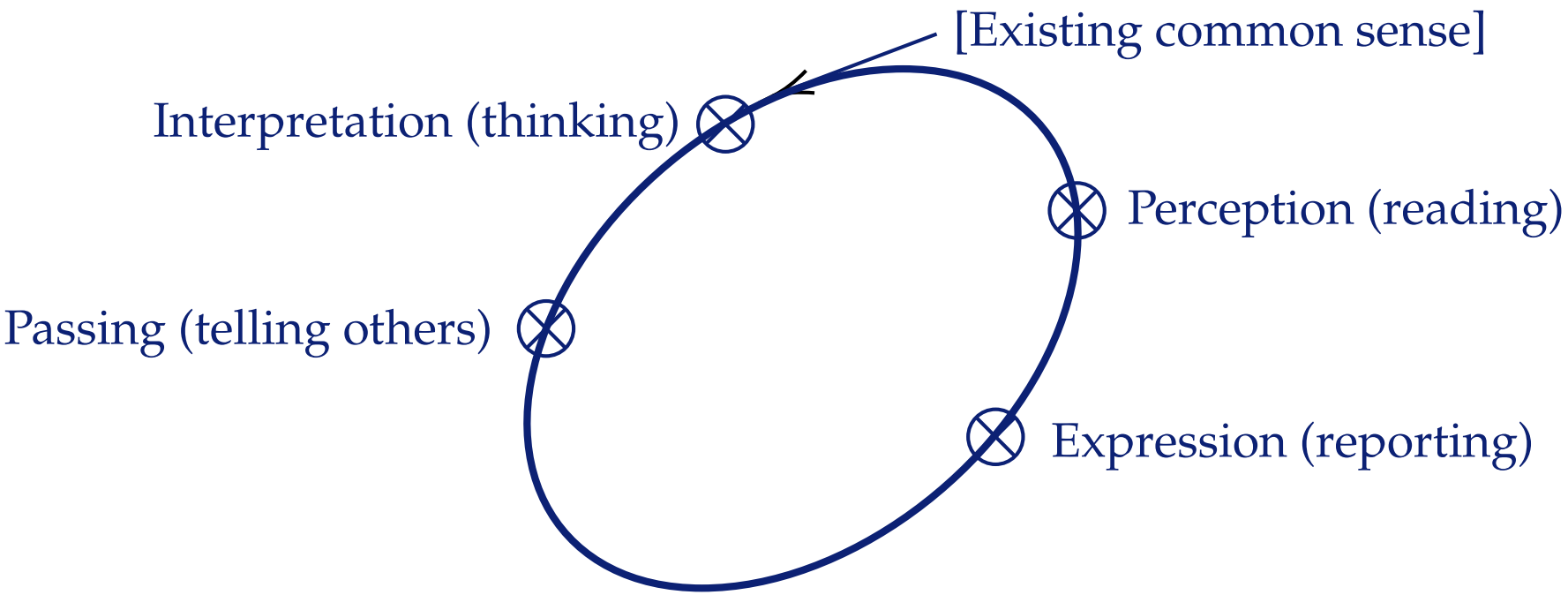
Erasmus MC
University Medical Center Rotterdam

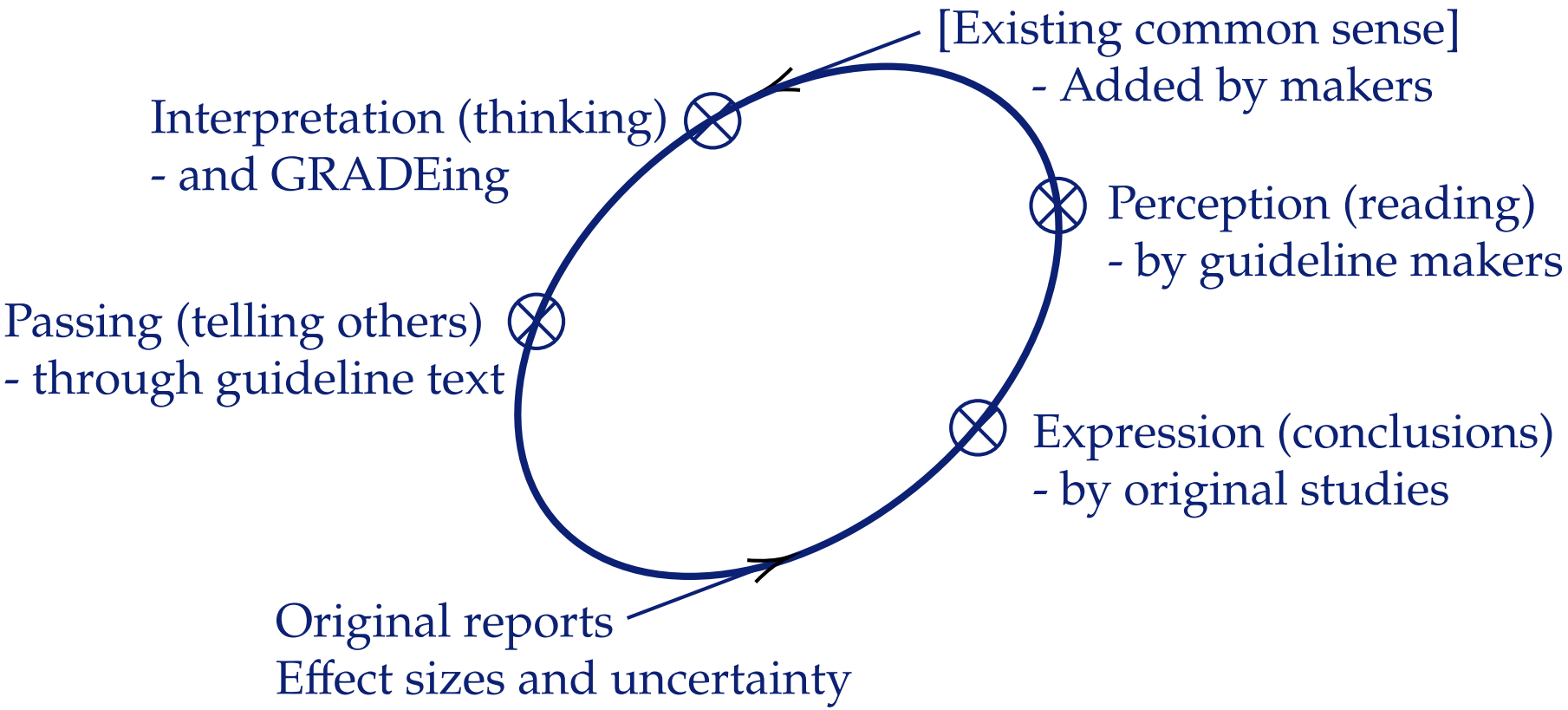
Strengths

1. The first to evaluate the causal language use in guidelines.
2. The first to propose the concepts “causation flow” and “causation jumps”.
3. Using TTE framework to evaluate trials.
4. NLP models assisting with automated ratings.

Erasmus MC
University Medical Center Rotterdam

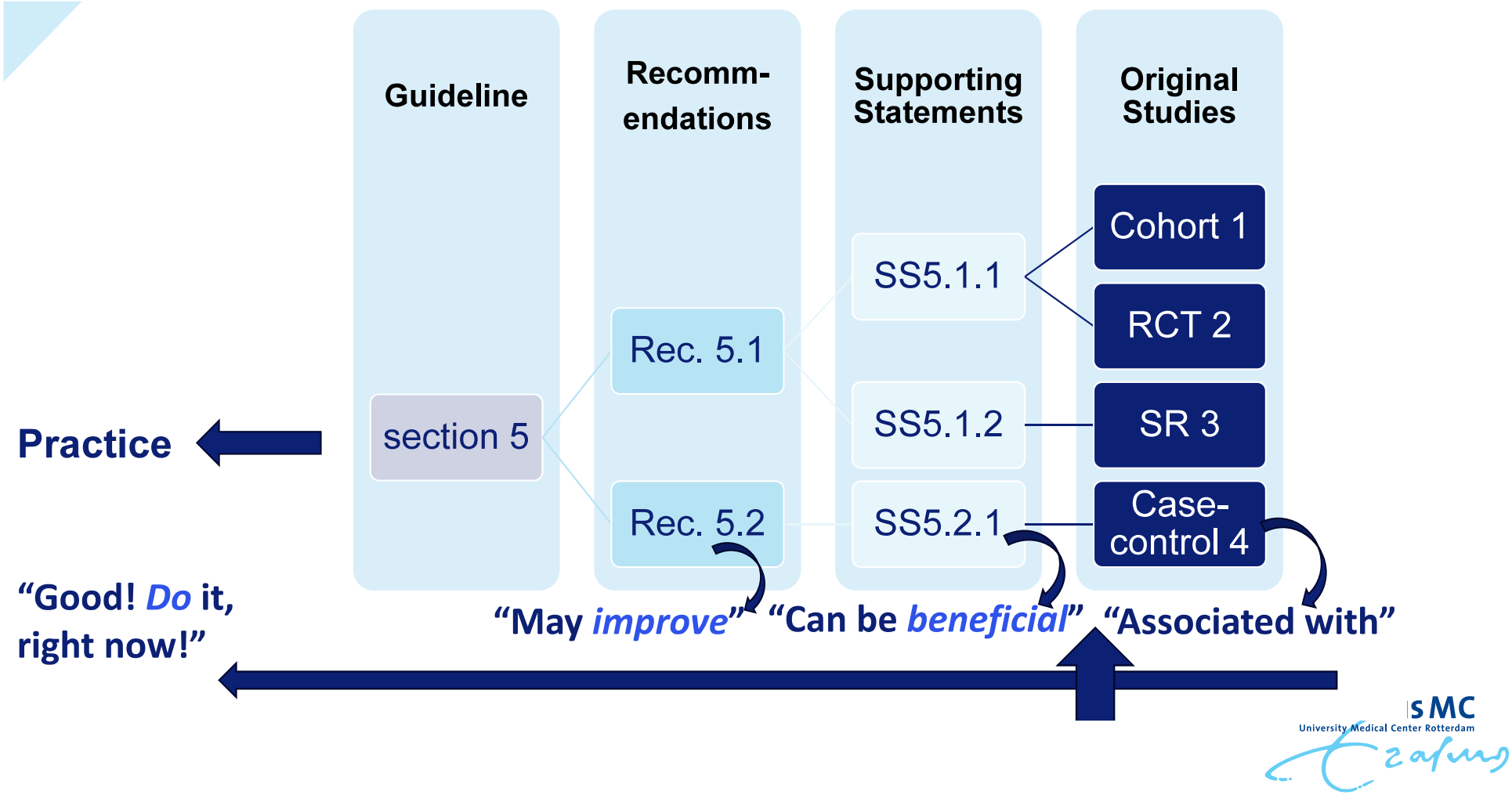

[Existing common sense]
Interpretation (thinking)
Perception (reading)
Passing (telling others)
Expression (reporting)
× n times
Observed data +
Assumptions +
Stat. tricks
Production
(Identifying)
Practice
(deciding)
Things
happen


[Existing common sense]
Interpretation (thinking)
Perception (reading)
Passing (telling others)
Expression (reporting)



Erasmus MC
University Medical Center Rotterdam

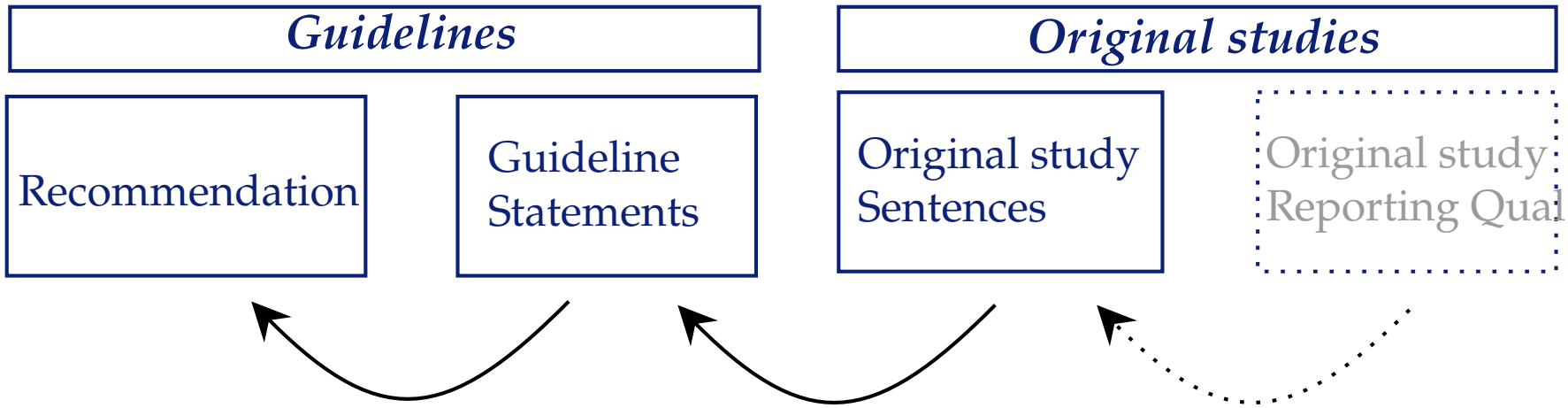
Causation flow and jumps in guidelines (cont'd)



Causation flow and jumps in guidelines (cont'd)

We... **do recommend** people drink 3 or more shots of expresso to prevent accidents.

After pooling results from 1000 65yr+... Morning coffee drinking **is associated with** lower road injury incidence ($RR = 0.77$ [0.36 – 0.98] per 3 shots espresso)



Erasmus MC
University Medical Center Rotterdam