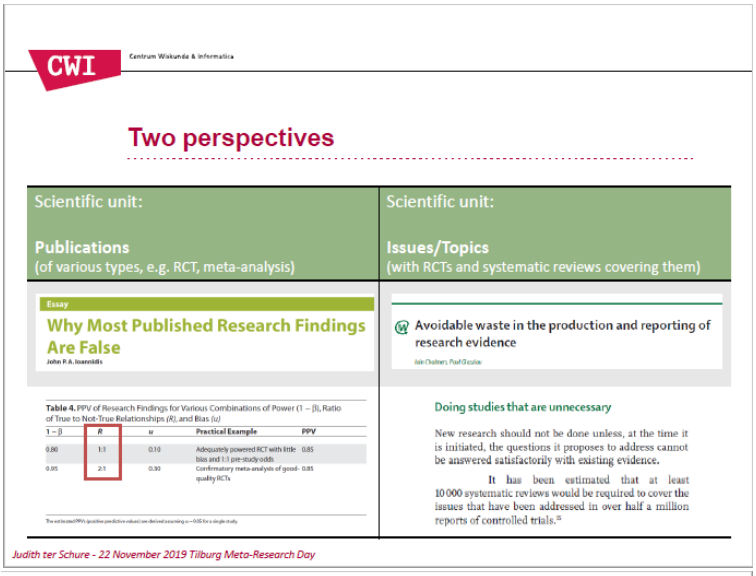
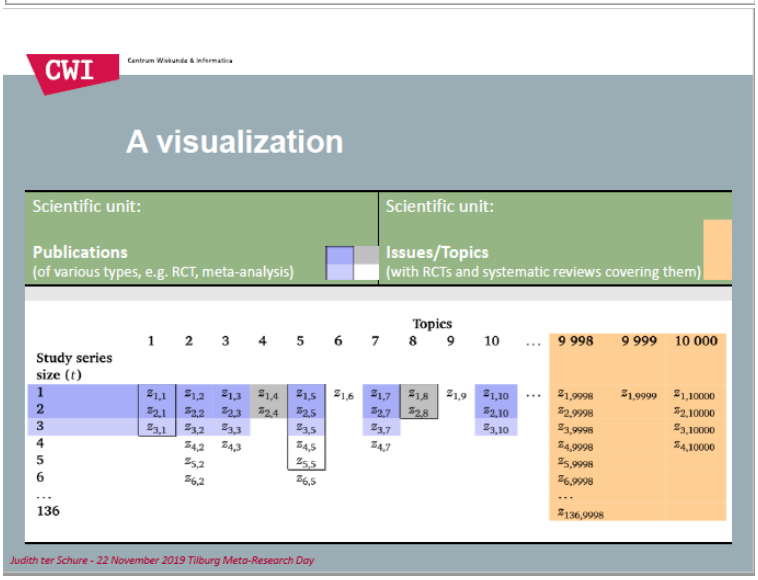
Take-away message:

It varies a lot per field whether scientists in their experimental design actually feel like they contribute to an accumulating series of studies. In some fields awareness exists that the results of an experiment will someday end up in a meta-analysis with existing experiments, while in others scientists aim to design experiments as 'refreshingly new' as possible. In a table that shows series of studies together in one column if they could be meta-analyzed, this latter approach shows scientists who mainly aim to initiate new columns. This pre-experimental perspective might be different from the meta-analysis perspective, in which a systematic search and inclusion criteria might still force those experiments together in one column, even though they weren't intended that way. This practice might erode trust in meta-analyses that try to synthesize effects from too different experiments.

The discussion was very hesitant towards enforcing rules (e.g. by funders or universities) on scientists in priority setting, such as whether a field needs more columns of 'refreshingly new' experiments, or needs replications of existing studies (extra rows) so a field can settle on a specific topic in one column with a meta-analysis.

In terms of statistical consequences, sequential processes might still be at play if scientists designing experiments know about the results of other experiments that might end up in the same meta-analysis. Full exchangeability in meta-analysis means that no-one would have decided differently on the feasibility or design of an experiment had the results of others been different. If that assumption cannot be met, we should consider studies as part of series in our statistical meta-analysis, even without forcing this approach in the design phase.





Note Judith beginning of session: The 'Issue/Topic' approach is very much inspired by Phase III clinical trials synthesized in the Cochrane database, in which an issue can be as specific as a treatment/patient/outcome-measure combination. Studies shown together in one column should share enough similarity for a meta-analyst to analyze them together. So conceptual replications might fall outside the scope of such an approach, and the appropriateness of this perspective might very much depend on the research field.

Why would we assign a different prior to a publication of an RCT (*R**= 1:1) than to a publication of a meta-analysis (*R** = 2:1)?

**R*: ratio of true to not-true relationships

Base *R* on available data

Of course the priors assigned in Table 4 of Why Most Published Research Findings Are False are illustrations of the underlying principle. The difference in *R* between single study publications of low bias RCTs and meta-analysis publications was based on findings in such publications in general (based on data): RCTs show positive results about half the time, while meta-analyses do so more often.

Correct available data for publication bias
Some research fields (e.g. priming in psychology) show very many positive results. So based on the available data, a new study in that field would be assigned a high prior, except that we sometimes should correct such appearance of the data for publication bias and assign a low prior instead.

Prior is not always constant
E.g. historically in tv advertising, the ratio of true to not-true relationships was high, while nowadays it is low. There is also a relation with the quality of the studies probing these relationships: recent studies are of better quality, maybe because few true effects are expected and the researchers do not want to be misled.
Effect sizes from studies can also seem to change in series of published studies, not due to changes in the true effect size, but due to changes in the selection of effects for publication (e.g. publication bias, outcome swithing)
Note Judith while writing this document: Bias in series due to selection can also occur not because scientists cheat in initial studies (filedrawering, outcome switching), but because scientists choose efficiently whether to expand initial studies; they do not select initial studies that look bad to replicate and thus expand into series, so we do not see initial underperforming studies as part of series (this is described in my Accumulation Bias paper).

Priors might be reflected in evidence cut-offs
In fields of physics where 7 sigma cutoffs are used to claim experimental findings, priors might be very different from fields that use alpha = 0.05. But the justification for an alpha threshold might also not depend on the priors of the researchers themselves (the physicists might have very high priors that the higgs particle exist) but on the prior of the outside world they want to confront with their evidence. So from that point of view it might make sense to not use very strict thresholds to provide evidence for relationships you expect the outside world assigns high prior (e.g. coffee is bad for you/good for you depending on who you consider 'outside world').

What processes drive the length of study series and occurrence of meta-analysis?

Hotness of the field
New studies might not be informed by previous results or science's need for new studies, but by scientists own enthusiasm about a set of seemingly similar relations (e.g. priming in psychology). Especially if these relations are researched without a solid underlying theory, these studies shouldn't be meta-analyzed together. So hotness of the field might not add to the depth of the table above ('a visualization'), but to the breadth of the table, by adding studies that actually research an issue so different from existing ones that it creates a new column.

Outside forces
Whether new studies are performed might depend on reward systems, and costs of experiments.

Priors
Available data might inform a prior for new studies (within the same column in the table above). But the decision to perform a new study might very well also depend on how risk-seeking a researcher is.
Note Judith during session: In Eindhoven, meta-researcher Anne Scheel intends to study whether research in the registered report format has a different prior from research outside that format. By studying evolution of priors during registered report phases (researcher/reviewers), she also hopes to find some data on differences in risk-seeking behavior among researchers. If these differences in priors exist, this could inform a publication type perspective.

What processes *should* drive the length of study series and occurrence of meta-analysis?

Informed by existing studies, but synthesize/conclude without them
Use existing studies to decide on the feasibility of a large prospective meta-analysis (many labs study/multi-center trial). But synthesize data only on the prospectively registered studies, which should be unbiased and have appropriate sample size.

What would this table look like if we start over with all new topics and would want to synthesize everything, not halfway throw away data?
Don't enforce topics on researchers. Let researchers pilot their own issues topics. But maybe even never force researchers to take into account the existing data to prioritize their topic of research in the social sciences: That recommendation of the Reduce Research Waste proposals might not be appropriate in the social sciences. Even if some topics grow into unreasonably large series (e.g. topic 9 998 with 136 studies in the visualization above), funders should not enforce researchers to set priorities differently.
Note Judith during session: Even in evaluation for phase III clinical trial grants, for which many funders have implemented the Reduce Research Waste recommendations for priority setting (e.g. ZonMW Goed Gebruik Geneesmiddelen), the availability of studies is hardly ever the reason not to award a grant. Researchers always seem to argue that their new trial is different than existing ones (subpopulation, variation of the treatment/outcome-measure) as the reason for performing it, and therefore existing evidence should not be directly used to evaluate the appropriateness of the new trial in terms of how it would change a meta-analysis. So in fact, researchers are often adding new topics, and expanding the table (visualization above) in breadth instead of in depth.

Pilot studies
Pilot studies should not be used to decide anything about the effect size. They only inform the feasibility of a study.
Note Judith while writing this document: This view is expressed more often. While discussing this point in more detail in the break after the session, I realized that I might use the term 'pilot study' in a nonstandard way. During this session (and in my paper about Accumulation Bias) I use it to denote the first study in a series of studies that might someday be meta-analyzed (or first in time, if a collection of studies is found in a systematic search). While the more standard use of the term is referring to a study that is more exploratory and is outside of a series of confirmatory studies to be meta-analyzed.

What is the better perspective: publications or topics?

Note Judith while writing this document: The answer might very much depend on the research field and whether new studies mainly add to the breadth of the table (a visualization) or its depth. But if a field publishes meta-analyses, even on very heterogeneous studies, then sequential processes might be at play. It might be very difficult to argue that of those studies that appear together in a meta-analysis (and therefore form a series, a column in the visualization table, at least according to someone's systematic search and inclusion criteria), these studies are completely exchangeable. The existence of results in some of the earlier studies might be known to the researchers involved in later studies, and therefore the existence and design of later studies might have been different had the results of earlier studies been different.

So why am I asking this question? Because it has statistical consequences

Note Judith during session: I imagine error control not to refer to imagined repetitions of your experiment in alternative worlds, but to existing experiments using the same method performed by others. In this view, the question arises: what are the other experiments/series or publications like mine for which the statistical method controls errors? The answer to this question depends on the perspective: publications or issues/topics.

CWI Centrum Wiskunde & Informatica

Thank you!

Statistical consequences

Scientific unit: **Publications** (of various types, e.g. RCT, meta-analysis)	Scientific unit: **Issues/Topics** (with RCTs and systematic reviews covering them)
Conventional meta-analysis (e.g. p-values for testing) • assumes fixed sample size and fixed study series size in meta-analysis • assumes only one analysis per topic • biased for any realistic retrospective meta-analysis	**Sequential meta-analysis** (e.g. alpha spending over meta-analysis updates) • runs over series of any size, including single studies! • assumes growing series, but with max size • assumes partly prespecified stopping rule
Bayesian analysis (e.g. posterior odds for testing) • conditions on everything • allows for updating, but any "most are (not) false" error control per publication type	**Martingale approach** (e.g. likelihood ratio universal bound, Safe Testing) • runs over series of any size, including single studies! • assumes growing series, without max size • works for any stopping rule: flexible

 Further Reading: Ter Schure & Grünwald (2019) Accumulation Bias in Meta-Analysis

CWI Centrum Wiskunde & Informatica

A visualization

Scientific unit: **Publications** (of various types, e.g. RCT, meta-analysis)	Scientific unit: **Issues/Topics** (with RCTs and systematic reviews covering them)

Study series size (t)	Topics 1	2	3	4	5	6	7	8	9	10	…	9 998	9 999	10 000
1	$z_{1,1}$	$z_{1,2}$	$z_{1,3}$	$z_{1,4}$	$z_{1,5}$	$z_{1,6}$	$z_{1,7}$	$z_{1,8}$	$z_{1,9}$	$z_{1,10}$	…	$z_{1,9998}$	$z_{1,9999}$	$z_{1,10000}$
2	$z_{2,1}$	$z_{2,2}$	$z_{2,3}$	$z_{2,4}$	$z_{2,5}$		$z_{2,7}$	$z_{2,8}$		$z_{2,10}$		$z_{2,9998}$		$z_{2,10000}$
3	$z_{3,1}$	$z_{3,2}$	$z_{3,3}$		$z_{3,5}$		$z_{3,7}$			$z_{3,10}$		$z_{3,9998}$		$z_{3,10000}$
4		$z_{4,2}$	$z_{4,3}$		$z_{4,5}$		$z_{4,7}$					$z_{4,9998}$		$z_{4,10000}$
5		$z_{5,2}$			$z_{5,5}$							$z_{5,9998}$		
6		$z_{6,2}$			$z_{6,5}$							$z_{6,9998}$		
…												…		
136												$z_{136,9998}$		