



Universiteit
Leiden
The Netherlands

ALL-IN meta-analysis

Schure, J.A. ter

Citation

Schure, J. A. ter. (2022, April 7). *ALL-IN meta-analysis*. Retrieved from <https://hdl.handle.net/1887/3281933>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3281933>

Note: To cite this publication please use the final published version (if applicable).

ALL-IN meta-analysis

Judith ter Schure

ISBN 978-90-619-6413-1

Cover design

Ilse Modder

Cover photo credits

Marjolein van Sommeren (conceptualization), Kilian Lafleur (digital edit),
Rinske ter Schure (photographer), Thomas de Jong (photographer),
Elsa ter Schure (photographer), Arnold ter Schure (encouragement)

ALL-IN meta-analysis

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op donderdag 7 april 2022
klokke 15:00 uur

door

Julia Anna (Judith) ter Schure

geboren te Meppel, Nederland
in 1992

Promotores:

Prof. dr. Peter D. Grünwald (Universiteit Leiden en
Centrum Wiskunde & Informatica, Amsterdam)

Dr. Daniel Lakens (Technische Universiteit Eindhoven)

Promotiecommissie:

Prof. dr. Frans A.J. de Haas

Prof. dr. Jelle J. Goeman

Dr. ir. Joanna in 't Hout (Radboud Universiteit)

Prof. dr. Glenn Shafer (Rutgers University)

Prof. dr. Alex J. Sutton (University of Leicester)

This work was funded by the Dutch Research Council (NWO) and carried out at
Centrum Wiskunde & Informatica (CWI), Amsterdam.



Centrum Wiskunde & Informatica



To Glenn Shafer, Stephen Senn, Peter Grünwald and Daniel Lakens

Subsets of you taught me the importance of fundamentals and history, the beauty of clinical trials, and – especially when a mathematical concept is necessary to make a point – the power of storytelling.

Origin of the material

The dissertation is based on the following earlier (pre-print) publications:

Chapter 1 is based on a paper that is under review at F1000 and available on ArXiv:

Judith ter Schure and Peter Grünwald. ALL-IN Meta-analysis: Breathing Life into Living Systematic Reviews. arXiv:2109.12141. 2021.

Chapter 2 is based on a paper that is available on ArXiv:

Judith ter Schure, Muriel F. Pérez-Ortiz, Alexander Ly and Peter Grünwald. The Safe Logrank Test: Error Control under Continuous Monitoring with Unlimited Horizon. arXiv:2011.06931. 2020.

Chapter 3 is based on a paper that is published at F1000 Research:

Judith ter Schure and Peter Grünwald. Accumulation Bias in Meta-analysis: The Need to Consider Time in Error Control [version 1; peer review: 2 approved]. *F1000Research*, 2019.

Chapter 4 is based on a blogpost at The Replication Network:

Judith ter Schure. Accumulation Bias: How to Handle It ALL-IN. *The Replication Network*. 2020.

Chapter 5 is based on a blogpost at The Replication Network:

Judith ter Schure and Peter Grünwald. Accumulation Bias: How to Handle It As a Bayesian. *The Replication Network*. 2022.

Chapter 6 is based on a paper published in STAtOR, the society magazine of the Netherlands Society for Statistics and Operations Research VVSOR:

Judith ter Schure, Peter Grünwald and Alexander Ly. Pandemic Preparedness in Data Sharing: Lessons Learned from Collaborating in a Live Meta-Analysis. *STAtOR*, 2021, 22.4: 47-52.

Contents

Preface	1
Introduction	5
1 ALL-IN meta-analysis	19
1.1 Statistics	27
1.2 Efficiency	34
1.3 Collaboration	37
1.4 Communication	42
1.5 Concluding remarks	44
Appendices	
1.A The inverse-conservative p -value	46
1.B R Code for calculations, simulations and plots	47
2 The Safe logrank test	49
2.1 Safe logrank tests	55
2.2 Comparing rejection regions	65
2.3 Comparing sample size	70
2.4 Variations and extensions	72
2.5 Discussion, Conclusion and Future Work	77
Appendices	
2.A Towards Continuous Time	79
2.B Expected Stopping Time, GROW and Wald's Identity	81
2.C Logrank test as a score test	82
2.D Details of sample size comparison simulations	86
3 Accumulation Bias	89
3.1 Accumulation Bias	91
3.2 A <i>Gold Rush</i> example: new studies after finding significant results	93
3.3 The Accumulation Bias Framework	99
3.4 <i>Time</i> in error control	107
3.5 Intermezzo: evidence for the existence of Accumulation Bias	111
3.6 Likelihood ratios' independence from meta-analysis time	112

3.7	The choice between error control conditioned and surviving over time . . .	116
3.8	Why likelihood ratios work: dependencies as strategy	117
3.9	Discussion	119

Appendices

3.A	Common/fixed-effect meta-analysis	121
3.B	Expectation <i>Gold Rush</i> conditional pilot Z-score	122
3.C	Expectation <i>Gold Rush</i> conditional meta-analysis Z-score	123
3.D	Mixture variance	124
3.E	Maximum time probability	124
3.F	Error control surviving over time in terms of a sum	125
3.G	Code availability	125

4 Accumulation Bias: How to handle it ALL-IN 127

4.1	Our example: extreme <i>Gold Rush</i> accumulation bias	128
4.2	The conditional sampling distribution under extreme <i>Gold Rush</i> accumula- tion bias	129
4.3	Accumulation bias can be efficient	131
4.4	The unconditional sampling distribution under extreme <i>Gold Rush</i> accu- mulation bias	132
4.5	ALL-IN meta-analysis	134
4.6	Accumulation bias from ALL-IN meta-analysis vs <i>Gold Rush</i>	136
4.7	Properties averaged over time	136
4.8	Multiple testing over time	138
4.9	Conclusion	141

Appendices

4.A	Extreme <i>Gold Rush</i> expressed in accumulation bias framework	142
4.B	Extreme <i>Gold Rush</i> conditional sampling distribution	143
4.C	The martingale underlying the table	145

5 Accumulation Bias: How to handle it as a Bayesian 147

5.1	Our example: extreme <i>Gold Rush</i> accumulation bias	148
5.2	Likelihood ratios	149
5.3	Two simple hypotheses	151
5.4	Bayesian error control under extreme <i>Gold Rush</i> accumulation bias	155
5.5	The prior odds are crucial	159
5.6	Beyond simple hypotheses	161
5.7	Pseudo-Bayesian error control	163
5.8	Conclusion	164

Appendices

5.A	Pseudo-Bayes posterior odds for exponential families and beyond	165
-----	---	-----

5.B Extension and Proof of Theorem 5.A.1	168
6 Data sharing in a live meta-analysis	173
6.1 Sharing live results while keeping researchers blinded	176
6.2 A central analysis	176
6.3 Data transfer agreements	178
6.4 Estimation	178
6.5 Conclusion	178
Discussion and future work	181
Bibliography	199
Samenvatting	203
Dankwoord	205
Curriculum Vitae	207

Preface

This Ph.D. research had its origin in a bar; a typical bar in Utrecht, in a historic wharf cellar at the central canal. On Wednesday, April 20th 2016, this bar served as the scenery for the Young Statisticians to host their night of beers and statistical discussion on the (ab)use of p -values in research: “*To p or not to p ?*” It was there that I heard Professor Peter Grünwald speak about how p -values are misunderstood and how much better we could do if we thought of statistics a bit more like gambling. I enjoyed every minute of it – also thanks to the great atmosphere that evening – and, fortunately, I still do.

Later that year I finished my Master’s *Statistical Science for the Life and Behavioural Science* while staying in contact with Peter. I was very lucky that the timing of my graduation matched with Peter’s procurement of funding for Ph.D. students. As a contender for a position, I had the advantage to have already made my job interview impression that day in that bar. Peter remembers it as quite unorthodox in mathematics for a student to simply walk up to him and state something along the lines of “This is so cool! Can I spend a Ph.D. studying this?”.

Now, almost four years of Ph.D. research¹ later, I am still not bored with p -value discussions. What is more, friends refer to my Ph.D. research as “the nemesis of the p -value”, and they have a point. What else could be the final blow to “science by p -value” than a paper (Ter Schure and Grünwald (2019), Chapter 3) that points out that in the cumulative science we idolize – “standing on the shoulders of giants” – the p -value is impossible to calculate correctly unless we do clinical trials and meta-analyses for random reasons?

¹Four full-time equivalent years: between May 1st, 2017 and February 1th, 2022 I spent 44 months working 80% of my working week (\approx 35 weeks full-time equivalent) on this Ph.D. research and 13 months working 100%, so 48 full-time months in total.



