

An investigation into the relationship between tinnitus distress and cognitive performance

Nathan Andrew Clarke

Thesis submitted to University of Nottingham for the degree of Doctor of Philosophy

Abstract

Chronic subjective tinnitus (herein referred to as tinnitus) is the conscious perception of sound in the absence of an external sound source. It is very common and a significant proportion of those who experience it find their tinnitus percept distressing. Cognitive task performance has been suggested to be poorer for people with subjective tinnitus, however, research on this topic to date has yielded inconsistent findings. This thesis contributes to understanding of the role of cognitive performance in subjective tinnitus through three studies.

The first study presents a comprehensive systematic review and meta-analyses of associations between cognitive performance and subjective tinnitus. The meta-analyses pooled data from a total of 1,863 participants, and found that subjective tinnitus was negatively associated with various domains of cognitive performance, but particularly for executive functions. It also highlighted potential confounding factors in primary research on this topic, such as the unconsidered influence of hearing loss.

The second study presents experimental data from an investigation into performance for core executive functions in a large sample of participants with distressing, non-distressing, and no tinnitus ($n=139$). Participant groups undertook a battery of cognitive tasks that tapped core executive functions, then between-group differences were analysed. Analyses initially suggested that executive performance was poorer for groups with both distressing tinnitus and non-distressing, but observed that *prima facie* evidence from the study was confounded, and precluded an independent association between tinnitus distress and poorer cognitive performance being inferred. Furthermore, confounding through methodological practises from previous research within the wider literature was highlighted.

Additionally, advanced analyses and modelling of this experimental data demonstrated independent associations between executive functions and distressing

tinnitus. However, the analyses also demonstrated that this is generally not the case for non-distressing tinnitus. Furthermore, non-distressing tinnitus was independently associated with improved performance for certain executive functions relative to the other groups, which suggested a causal role for executive functioning in distressing tinnitus.

Building on these insights, the third study proposed a conceptual model of tinnitus distress (the Executive Disruption Model or EDM), and an accompanying structural causal model (SCM). The EDM provides a novel conceptualisation in positing that the tinnitus percept is not the subject of focus for an attentional spotlight, but instead intrudes into conscious awareness because of inherent and/or disrupted executive abilities and causally contributes to distress.

Application of the SCM provided preliminary validation of the EDM in an independent dataset. Factor analysis was used to obtain latent measures of executive functioning that facilitated consistent estimates of a relationship between poorer executive functioning and greater tinnitus distress in both datasets.

The research presented in this thesis contributes new knowledge to the relationship between cognitive performance and distress related to subjective tinnitus through comprehensive evidence synthesis, primary research and model development, and subsequent application and validation. This offers consistency, both in terms of implicated cognitive domains and size of effect. The thesis presents novel theoretical and methodological insights into research on this topic with practical applications for cognitive research into tinnitus, related interventions for tinnitus distress, and potential clinical applications.

Publications

Publications arising from this thesis

Clarke, N. A., Akeroyd, M.A., Adams, B., Henshaw, H., Hoare, D.J. 2020. Associations between subjective tinnitus and cognitive performance: systematic review and meta-analyses. *Trends in Hearing*, 24:1-23.

Clarke, N. A., Akeroyd, M.A., Henshaw, H., Hoare, D.J. 2018. Association between subjective tinnitus and cognitive performance: protocol for systematic review and meta-analysis. *BMJ Open* 8, e023700.

Other publications during this thesis

Beck D. L, Bant S, Clarke N. A. 2020. Hearing loss and cognition: a discussion for audiologists and hearing healthcare professionals. *J Otolaryngol ENT Res*. 12(3):72-78.

Clarke, N. A., Hoare, D. J., Killan, E. 2019. Evidence for an association between disrupted sleep and hearing impairment: scoping review. *American Journal of Audiology*. 28:1015-1024.

Oral presentation

Clarke, N. A., Akeroyd, M., Henshaw, H., Hoare, D. An investigation into the relationship between tinnitus distress, concentration difficulties, and cognitive performance. British Society of Audiology Annual Conference, Virtual Conference. 12 October 2020.

Poster presentations

Clarke, N. A., Akeroyd, M., Henshaw, H., Hoare, D. Associations between tinnitus and cognitive performance: systematic review and meta-analyses. British Society of Audiology Annual Conference, Molineux Stadium, Wolverhampton. 5 June 2019.

Clarke, N. A., Akeroyd, M., Henshaw, H., Hoare, D. Associations between tinnitus and cognitive performance: systematic review and meta-analyses. CHSCOM 2019, Linköping, Sweden. 9 June 2019.

Declaration

I certify that this thesis is my own work, except where indicated by referencing. No part of this thesis has been submitting elsewhere for any other degree or qualification.

Nathan Clarke

Date

Acknowledgements

- My son, Elliott: born midway through this PhD and my proudest achievement.
- My wife, Olivia: for her constant support during challenging times.
- My in-laws, Horace and Elizabeth: for taking us in, looking after us, and making this thesis possible.
- My supervisory team: for their kindness, guidance, and encouragement.
- Colleagues at University of Nottingham Hearing Sciences: for their quickness to help and caring a great deal about what they do.
- The Medical Research Council Institute of Hearing Research (MRC IHR): for the opportunity, funding, and support.

List of Abbreviations

BEA	better-ear average
BF	Bayes factor
CBT	cognitive behavioural therapy
CFQ	Cognitive Failures Questionnaire
CHC	Cattell, Horn, Carroll
CHC-M	Cattell, Horn, Carroll, and Miyake
CRIq	Cognitive Reserve Index questionnaire
DAG	directed acyclic graph
DASS21	Depression, Anxiety, and Stress questionnaire
EDM	Executive Disruption Model
EF	executive functions
EFiT	Executive Functioning in People with Tinnitus study
ESS	effective posterior sample size
FSR	factor score regression
g	general intelligence
Gc	crystallised intelligence
Gf	fluid intelligence
Gl	learning/encoding efficiency
GLM	general linear model
Gl _r	long-term storage and retrieval
Gr	retrieval fluency
Gs	processing speed
Gsm	general short-term memory
Gv	visual processing
HLM	hierarchical linear model
HWM	high working memory
LWM	low working memory
MCI	mild cognitive impairment
MCMC	Markov Chain Monte Carlo
MMSE	Mini Mental State Examination
MoCA	Montreal Cognitive Assessment
MSA	measure of sampling adequacy
NHST	null hypothesis significance testing
N _p	unpractised words from unpractised categories
OVL	overlap statistic
PAF	principal axis factoring
PoS	probability of superiority statistic
PPC	posterior predictive check
PTA	pure tone audiometry
RIF	retrieval-induced forgetting
R _p -	unpractised words from practised categories

Rp+	practised words from practised categories
RRS	Rumination Response Scale
RT	response time
RVE	robust variance estimation
SALSA	Search, Appraisal, Synthesis, and Analysis Typology of Reviews
SCC	subjective cognitive complaints
SCM	structural causal model
SEM	structural equation model
SMD	standardised mean difference
STM	short-term memory
Supervisory Attentional System	SAS
TFI	Tinnitus Functional Index
THI	Tinnitus Handicap Inventory
TSCHQ	Tinnitus Sample Case History Questionnaire
U3	non-overlap statistic
WAIS	Weschler Adult Intelligence Scale
WASI	Weschler Abbreviated Scale of Intelligence
WoE	Weight of Evidence

Table of contents

1	Chapter 1 – Introduction to tinnitus and cognitive performance	1
1.1	Introduction	1
1.2	Overview of tinnitus	1
1.3	Pathophysiological models of subjective tinnitus	3
1.3.1	Peripheral and established models	3
1.3.2	Spontaneous neural activity	6
1.3.3	Contemporary models of tinnitus	6
1.3.4	Summary.....	8
1.4	Management and interventions for subjective tinnitus.....	8
1.5	Tinnitus distress, cognition, and cognitive performance.....	10
1.5.1	Models of tinnitus and cognitive performance	11
1.5.2	Review of cognitive performance and tinnitus reviews	16
1.6	Thesis aim and objectives	22
2	Chapter 2 – Associations between subjective tinnitus and cognitive performance: systematic review and meta-analyses	24
2.1	Introduction	24
2.2	Meta-analysis overview.....	24
2.2.1	Background.....	24
2.2.2	Meta-analytic process.....	25
2.2.3	Applying meta-analysis to associations between tinnitus and cognitive performance	27
2.3	Methods.....	30
2.3.1	Databases and search strategy	30
2.3.2	Inclusion criteria	31
2.3.3	Article screening and selection.....	32
2.3.4	Data extraction and data items.....	32
2.3.5	Weight of evidence (WoE) – record bias and quality	32
2.3.6	Categories of cognitive function	33
2.3.7	Statistical analyses	36
2.4	Results	40
2.4.1	Included study overview	40
2.4.2	Weight of Evidence (WoE) - record bias and quality appraisal	41
2.4.3	Model results.....	41
2.5	Discussion: is subjective tinnitus associated with poorer cognitive performance?	59

2.5.1	Which broad cognitive domains are associated with tinnitus?.....	59
2.5.2	Which narrow cognitive domains are associated with tinnitus?	62
2.5.3	Quality of available evidence	65
2.6	Conclusions.....	70
3	Chapter 3 – A comparison of executive functioning in distressing and non-distressing tinnitus groups (EFiT study).....	71
3.1	Introduction	71
3.2	What is Executive Functioning?.....	71
3.2.1	Controlled Versus Automatic Processes Models	72
3.2.2	Goal-Centred Processing Models	72
3.2.3	Multifactor Models.....	73
3.2.4	Core executive functions: Updating (of working memory), Inhibition, and Shifting	74
3.2.5	Fluid intelligence (or fluid ability)	75
3.3	EFiT Study (Executive Functioning in People with Tinnitus).....	76
3.3.2	Methods	82
3.4	Data analysis.....	93
3.4.1	Data checking	93
3.4.2	Processing response time data.....	93
3.4.3	Statistical inference.....	95
3.5	Results	99
3.5.1	Overview of study groups	99
3.5.2	Descriptive group differences in objective cognitive measures	105
3.5.3	Standardised Mean Differences (SMD)	105
3.5.4	Frequentist t-tests	110
3.5.5	Bayesian t-tests.....	112
3.5.6	Summary of results.....	114
3.6	Discussion: Methodological and Inferential Limitations Concerning Associations between Tinnitus Distress and Cognitive Performance	118
3.7	Conclusion	120
4	Chapter 4 – Advanced analyses and modelling to establish independent associations between cognitive performance and tinnitus	122
4.1	Introduction	122
4.2	Methods.....	122
4.2.1	General linear models (GLMs).....	124
4.3	Results	128
4.3.1	Frequentist GLMs	128
4.3.2	Bayesian GLMs	135
4.3.3	Summary of statistical inference	138

4.3.4	Retrieval-induced forgetting (RIF) sub-analysis.....	142
4.3.5	Hierarchical linear models for response time tasks	148
4.4	General discussion of advanced analyses and modelling.....	164
4.4.1	Objective measures.....	164
4.4.2	Self-report measures.....	169
4.4.3	Methods and analysis	171
4.4.4	Other variables.....	172
4.5	Consideration of observed effect sizes: small differences add up	174
4.6	Conclusions.....	179
5	Chapter 5 – Preliminary validation of a causal model of tinnitus distress.....	181
5.1	Introduction	181
5.2	From conceptual to causal cognitive models of tinnitus distress	181
5.3	Executive Disruption Model of Tinnitus Distress (EDM)	187
5.3.1	Causal assumptions	190
5.3.2	Percept loudness disruption.....	193
5.4	Structural Causal Modelling (SCM)	195
5.4.1	Directed acyclic graphs	197
5.4.2	Statistical “control” can be misleading	202
5.4.3	Minimum adjustment sets	203
5.5	Mapping the EDM to a DAG	204
5.5.1	Complete Variable DAG	204
5.5.2	Unobserved Variable DAG	207
5.6	Independent validation analysis.....	208
5.6.1	Overview of datasets.....	209
5.6.2	Principal axis factoring.....	210
5.6.3	Factor score regression (FSR).....	212
5.6.4	Results	213
5.7	Study strengths and limitations	214
5.8	Summary.....	215
6	Chapter 6 – General Discussion	217
6.1	Introduction	217
6.2	Discussion of key findings	217
6.2.1	Support for established models.....	222
6.2.2	Clinical relevance.....	228
6.2.3	Strengths and limitations	230
6.3	Executive functioning and the role of attention in tinnitus	230
6.3.1	Attention amplifies tinnitus severity	232
6.3.2	Attention originates and/or perpetuates the tinnitus percept	232

6.3.3	Attention distorts the tinnitus percept.....	232
6.3.4	The spotlight of attention	233
6.3.5	Causal relationship between constructs.....	234
6.4	Future work	236
6.4.1	EDM: implications and testable predictions.....	236
6.4.2	Wider applications of tinnitus and SCMs	238
6.4.3	Executive function interventions	240
6.5	Summary	244
	References	246
	Appendices	264
	Appendix A – WoE tool	264
	Appendix B – Retrieval-Practise Paradigm Category-Exemplars	266

List of Tables

Table 1—1. Overview of previous reviews investigating cognitive performance and tinnitus.	17
Table 2—1. Descriptive overview of the broad and narrow cognitive domains constituting the CHC-M taxonomy. Adapted from Webb et al. (2018). Executive function definition from Friedman and Miyake (2017).....	34
Table 2—2. Count of effect sizes in each broad cognitive factor.	41
Table 2—3. Overview of meta-analytic models, providing correlation, statistical significance, and number of effect sizes contained in its calculation.	41
Table 3—1. Overview of objective cognitive measures, outcome type, and cognitive construct measured.	85
Table 3—2. Clinical categorisation scoring system provided with the DASS21.	92
Table 3—3. Evidence categories for p values, for effect sizes, and for Bayes factors (adapted for Wetzels et al., 2011).	99
Table 3—4. Group counts, means and standard deviations of demographic and self-reported measures.	100
Table 3—5. Group means and standard deviations of objective cognitive measures. ..	105
Table 3—6. Standardised mean difference between and confidence intervals for objective cognitive measures between No Tinnitus, Non-Distressing Tinnitus, and Distressing Tinnitus groups.	109
Table 3—7. Objective cognitive measure standardised mean differences between No Tinnitus and Non-Distressing Tinnitus groups converted to correlations.....	110
Table 3—8. Objective cognitive measure standardised mean differences between No Tinnitus and Distressing Tinnitus groups converted to correlations.	110
Table 3—9. One-tailed, independent samples t-test results for No Tinnitus and Non-Distressing Tinnitus groups.....	111
Table 3—10. One-tailed, independent samples t-test results for No Tinnitus and Distressing Tinnitus Groups.	111
Table 3—11. One-tailed, independent samples t-test results for Non-Distressing Tinnitus and Distressing Tinnitus Groups	112
Table 3—12. Self-reported and objective cognitive measure Bayesian t-test results for No Tinnitus and Non-Distressing Tinnitus groups.....	112
Table 3—13. Self-reported and objective cognitive measure Bayesian t-test results for No Tinnitus and Distressing Tinnitus groups.	113

Table 3—14. Self-reported and objective cognitive measure Bayesian t-test results for Non-Distressing Tinnitus and Distressing Tinnitus groups	114
Table 4—1. Unstandardised general linear model results predicting objective cognitive measures with tinnitus group, age, hearing threshold, depression, anxiety, and stress as predictors.....	128
Table 4—2. Statistically significant predictors from general linear models predicting objective cognitive measures with tinnitus group, age, hearing threshold, depression, anxiety, and stress as predictors.	130
Table 4—3. Standardised general linear model results predicting objective cognitive measures with tinnitus group, age, hearing threshold, depression, anxiety, and stress as predictors.....	131
Table 4—4. Ranking of Distressing Tinnitus group partial beta coefficient size within general linear models of objective cognitive measures with tinnitus group, age, hearing threshold, depression, anxiety, and stress as predictors.	133
Table 4—5. Ranking of Non-Distressing Tinnitus group partial beta coefficient size within general linear models of objective cognitive measures with tinnitus group, age, hearing threshold, depression, anxiety, and stress as predictors.	134
Table 4—6. First and second ranked partial beta coefficients size within general linear models of objective cognitive measures with tinnitus group, age, hearing threshold, depression, anxiety, and stress as predictors.	134
Table 4—7. Bayesian general linear model predicting objective cognitive measures for Distressing Tinnitus group with tinnitus group, age, hearing threshold, depression, anxiety, and stress as predictors, standardised beta weights by rank with probability that beta is less than, or greater than zero.	135
Table 4—8. Bayesian general linear regression predicting objective cognitive measures for Non- Distressing Tinnitus group with tinnitus group, age, hearing threshold, depression, anxiety, and stress as predictors, standardised beta weights by rank with probability that beta is less than, or greater than zero	137
Table 4—9. Table detailing the ratio of beta weight ratios for each predictor in each GLM relative to the distressing tinnitus predictor.	139
Table 4—10. Mean proportions of correct recalls on Rp+, Rp-, Np, total, and RIF effect score for each group.....	143
Table 4—11. General linear models predicting Rp+, Rp-, and Np scores from tinnitus group category, age, hearing threshold, DASS_D, DASS_A, DASS_S, and RRS scores.	144
Table 4—12. Table showing partial beta weights and accompanying rank within each model for each retrieval-practise category.....	145

Table 4—13. Bayesian GLM with all predictors for RRS_scaled partial beta scores and none-zero probability estimates.	146
Table 4—14. Bayesian GLM with all predictors for tinnitus and distressing tinnitus for each retrieval-practise category with accompanying beta weights and none-zero probabilities.....	146
Table 4—15. Hierarchical linear model parameter estimates for Stroop task (incongruent) response time.....	153
Table 4—16. Hierarchical linear model parameter estimates for rule-switching task (switch trials) response time.....	159
Table 4—17. Effect size distribution summary statistics for Tinnitus and bothersome Tinnitus groups compared to the Control Group, showing standardised mean differences, overlap (OVL), non-overlap (U3), and probability of superiority (PoS) for objective cognitive measures.	175
Table 4—18. Quantile effect size distribution analysis from Clarke et al. (2020).	177
Table 5—1. Cognitive tasks and ascribed cognitive domains detailed in Mohamad (2015). WASI = Weschler Abbreviated Scale of Intelligence (2 nd ed.), WAIS-IV = Weschler Adult Intelligence Scale (4 th ed.), TEA = Test of Everyday Attention.....	210
Table 5—2. Frequentist factor score regression parameter estimates for Datasets 1 & 2.	213
Table 5—3. Bayesian factor score regression parameter estimates for Datasets 1 & 2	214

List of Figures

Figure 1—1. The Neurophysiological model of tinnitus (Jastreboff, 1990), showing a schematic representation of hypothesised relationships between classical auditory pathways including the limbic system, and sympathetic autonomic nervous system.	4
Figure 1—2. 'A model of the road to annoyance via different levels of cognitive functioning' from Andersson and McKenna (2006). The dashed line depicts a path through tinnitus and various "escape routes" towards no annoyance.	12
Figure 1—3. The 'Cognitive Behavioural Model' of tinnitus distress from McKenna et al. (2014).	13
Figure 1—4. Structural model depicting the proposed effects of tinnitus severity on cognitive performance adapted from Mohamad et al. (2016).	14
Figure 1—5. A 'Neurocognitive model' that links cognitive performance to chronic tinnitus from Trevis et al. (2016).	15
Figure 2—1. Hierarchical representation of the CHC-M composite model (adapted from Webb et al., 2018). g = general intelligence, Gf = fluid intelligence, Gc = crystallised intelligence, Gv = visual processing, Glr = long term storage and retrieval, Gsm = general short-term memory, Gs = processing speed, and EF = executive functions, STM = Short-term memory, LWM = Low working memory, HWM = High working memory. EF features a dotted outline to indicate its absence in the original CHC framework and subsequent addition by Webb et al., 2018 to create CHC-M.	33
Figure 2—2. PRISMA flowchart providing an overview of the systematic search and screening process.	40
Figure 2—3. Forest plot showing the correlation between tinnitus and response times in Executive Functioning tasks (EF-rt). Each study is represented by a point estimate bounded by a 95% CI, with the area of each square proportional the study's weight within the model. Summary effect size is displayed as a polygon at the bottom of the plot, with the width of the polygon representing the 95% CI.	42
Figure 2—4. Forest plot showing the correlation between tinnitus and error rates in Executive Functioning tasks (EF-error). Each study is represented by a point estimate bounded by a 95% CI, with the area of each square proportional the study's weight within the model. Summary effect size is displayed as a polygon at the bottom of the plot, with the width of the polygon representing the 95% CI.	43
Figure 2—5. Forest plot showing the correlation between tinnitus and correct responses in Executive Functioning tasks (EF-correct). Each study is represented by a point estimate bounded by a 95% CI, with the area of each square proportional the study's	

weight within the model. Summary effect size is displayed as a polygon at the bottom of the plot, with the width of the polygon representing the 95% CI.	44
Figure 2—6. Contour-enhanced funnel plot for EF-rt model. Individual study effect sizes are displayed on the x-axis, with their corresponding standard errors on the y-axis. The funnel lines are centred on the summary effect size, providing an indication of the spread. The unshaded region in the middle corresponds to p-values greater than .10, the light grey shaded region corresponds to p-values between .10 and .05, the dark grey region corresponds to p-values between .05 and .01, and the region outside of the funnel corresponds to p-values below .01.....	46
Figure 2—7. Forest plot showing the correlation between tinnitus and response times in Inhibition tasks (Inhibition-rt). Each study is represented by a point estimate bounded by a 95% CI, with the area of each square proportional the study's weight within the model. Summary effect size is displayed as a polygon at the bottom of the plot, with the width of the polygon representing the 95% CI.	47
Figure 2—8. Forest plot showing the correlation between tinnitus and response times in Shifting tasks (Shifting-rt). Each study is represented by a point estimate bounded by a 95% CI, with the area of each square proportional the study's weight within the model. Summary effect size is displayed as a polygon at the bottom of the plot, with the width of the polygon representing the 95% CI.....	48
Figure 2—9. Forest plot showing the correlation between tinnitus and response times in processing speed tasks (Gs-rt). Each study is represented by a point estimate bounded by a 95% CI, with the area of each square proportional the study's weight within the model. Summary effect size is displayed as a polygon at the bottom of the plot, with the width of the polygon representing the 95% CI.	50
Figure 2—10. Forest plot showing the correlation between tinnitus and correct responses in General Short-term Memory tasks (Gsm-correct). Each study is represented by a point estimate bounded by a 95% CI, with the area of each square proportional the study's weight within the model. Summary effect size is displayed as a polygon at the bottom of the plot, with the width of the polygon representing the 95% CI.	52
Figure 2—11. Forest plot showing the correlation between tinnitus and correct responses in General Learning and Retrieval tasks (Glr-correct). Each study is represented by a point estimate bounded by a 95% CI, with the area of each square proportional the study's weight within the model. Summary effect size is displayed as a polygon at the bottom of the plot, with the width of the polygon representing the 95% CI.	54
Figure 2—12. Forest plot showing the correlation between tinnitus and correct responses in General Learning tasks (Glr-GL-correct). Each study is represented by a point estimate bounded by a 95% CI, with the area of each square proportional the study's	

weight within the model. Summary effect size is displayed as a polygon at the bottom of the plot, with the width of the polygon representing the 95% CI.	55
Figure 2—13. Forest plot showing the correlation between tinnitus and correct responses in General Retrieval tasks (Glr-GR-correct). Each study is represented by a point estimate bounded by a 95% CI, with the area of each square proportional the study's weight within the model. Summary effect size is displayed as a polygon at the bottom of the plot, with the width of the polygon representing the 95% CI.	56
Figure 2—14. Forest plot showing the correlation between tinnitus and cognitive self-report measures (Self-report). Each study is represented by a point estimate bounded by a 95% CI, with the area of each square proportional the study's weight within the model. Summary effect size is displayed as a polygon at the bottom of the plot, with the width of the polygon representing the 95% CI.	57
Figure 2—15. . Forest plot showing the correlation between tinnitus and cognitive screening measures. Each study is represented by a point estimate bounded by a 95% CI, with the area of each square proportional the study's weight within the model. Summary effect size is displayed as a polygon at the bottom of the plot, with the width of the polygon representing the 95% CI.	58
Figure 3—1. A schematic representation of the unity and diversity of three core executive functions from Miyake and Friedman (2012).	73
Figure 3—2. A schematic diagram of interrelations between Updating, Inhibition, Switching, Fluid ability, and higher-level functions from Diamond (2013).	76
Figure 3—3. Schematic representation of the retrieval-practise paradigm from Anderson et al. (1994).	89
Figure 3—4. Simulated Stroop task ex-Gaussian density function of response time over 144 trials ($\mu = 600$, $\sigma = 80$, $\tau = 150$).	94
Figure 3—5. Spectrum of Bayes factor values that are typical within the biobehavioural sciences (Quintana and Williams, 2018).	96
Figure 3—6. a) States of evidence. b) What p-values provide. c) What Bayes factors provide. (Adapted from Dienes and Mclatchie, 2018)	97
Figure 3—7. Distribution of ages for No Tinnitus, Non-Distressing Tinnitus, and Distressing Tinnitus groups.	101
Figure 3—8. Proportion of DASS Anxiety score categories by group.	102
Figure 3—9. Proportion of DASS Depression score categories by group.	102
Figure 3—10. Proportion of DASS Stress score categories by group.	103
Figure 3—11. Proportion of moderate or higher DASS subscale score categories by group.	104
Figure 3—12. Plot displaying standardised mean differences (SMDs) for objective cognitive measures between No Tinnitus, Non-Distressing Tinnitus, and Distressing	

Tinnitus Groups. Positive SMDs represent superior performance to No Tinnitus group, while Negative SMDs show poorer performance. In this plot, the dashed vertical line represents the performance of the No Tinnitus group, while the purple and yellow squares represent the relative performance of the Distressing Tinnitus and Non-Distressing Tinnitus groups respectively. The signs of the SMDs have been transformed so that negative SMDs (i.e. squares to the left of dashed line) represent worse task performance than the No Tinnitus group, while positive SMDs (i.e. squares to the right of the dashed line) represent relatively better task performance.	107
Figure 3—13. Plot displaying standardised mean differences (SMDs) for objective cognitive measures between Non-Distressing Tinnitus, No Tinnitus, and Distressing Tinnitus Groups. Positive SMDs represent superior performance to Tinnitus group, while Negative SMDs show poorer performance than the No Tinnitus group. In this plot, the dashed vertical line represents the performance of the Non-Distressing Tinnitus Group, while the purple and yellow squares represent the relative performance of the No Tinnitus and Non-Distressing Tinnitus groups respectively. The signs of the SMDs have been transformed so that negative SMDs (i.e. squares to the left of dashed line) represent worse task performance than the Non-Distressing Tinnitus group, while positive SMDs (i.e. squares to the right of the dashed line) represent relatively better task performance.	108
Figure 4—1. Simple associative model tested by between-groups analysis presented in Chapter 3, showing an association between THI (tinnitus severity/distress) and EF (executive function).	122
Figure 4—2. A more complex associative model of tinnitus distress showing an association between THI (tinnitus severity/distress), EF (executive function), and a priori confounders.	123
Figure 4—3. Diagram illustrating the conceptual rationale of estimated multiple linear relationship in multiple regression analyses (dotted region represents the joint variation between X1 and Y).	124
Figure 4—4. Histogram of all incongruent responses in Stroop task.	150
Figure 4—5. Histogram of all switch responses in rule-switching task.	150
Figure 4—6. Posterior predictive checks for Stroop incongruent normal and shifted log-normal models.	153
Figure 4—7. HLM parameter estimates for each tinnitus groups predicting incongruent Stroop task response times (milliseconds). 1 = No Tinnitus, 2 = Non-Distressing Tinnitus, 3 = Distressing Tinnitus. THI = Tinnitus Handicap Inventory.	154
Figure 4—8. HLM parameter estimates for age predicting incongruent Stroop task response times (milliseconds) with accompany uncertainty boundaries.	155

Figure 4—9. HLM parameter estimates for participant hearing thresholds predicting incongruent Stroop task response times (milliseconds) with accompanying uncertainty boundaries.	155
Figure 4—10. HLM parameter estimates for DASS_A predicting incongruent Stroop task response times (milliseconds) with accompanying uncertainty boundaries. DASS_A = Depression, Anxiety, and Stress 21 (Anxiety).	156
Figure 4—11. HLM parameter estimates for DASS_D predicting incongruent Stroop task response times (milliseconds) with accompanying uncertainty boundaries. DASS_D = Depression, Anxiety, and Stress 21 (Depression).	156
Figure 4—12. HLM parameter estimates for DASS_S predicting incongruent Stroop task response times (milliseconds) with accompanying uncertainty boundaries. DASS_S = Depression, Anxiety, and Stress 21 (Stress).	157
Figure 4—13. Posterior predictive checks for rule-switch trials normal and shifted log-normal models.	158
Figure 4—14. HLM parameter estimates for each tinnitus groups predicting incongruent switch trials task response times (milliseconds). 1 = No Tinnitus, 2 = Non-Distressing Tinnitus, 3 = Distressing Tinnitus. THI = Tinnitus Handicap Inventory.	160
Figure 4—15. HLM parameter estimates for age predicting incongruent switch trials task response times (milliseconds) with accompanying uncertainty boundaries.	160
Figure 4—16. HLM parameter estimates for participant hearing thresholds predicting incongruent switch trials task response times (milliseconds) with accompanying uncertainty boundaries.	161
Figure 4—17. HLM parameter estimates for DASS_A predicting incongruent switch trials task response times (milliseconds). DASS_A = Depression, Anxiety, and Stress 21 (Anxiety) with accompanying uncertainty boundaries.	161
Figure 4—18. HLM parameter estimates for DASS_D predicting incongruent switch trials task response times (milliseconds) with accompanying uncertainty boundaries. DASS_D = Depression, Anxiety, and Stress 21 (Depression).	162
Figure 4—19. HLM parameter estimates for DASS_S predicting incongruent switch trials task response times (milliseconds) with accompanying uncertainty boundaries. DASS_S = Depression, Anxiety, and Stress 21 (Stress).	162
Figure 4—20. For the Matrix reasoning task, a Cohen's d of 0.69 implies that 75.5% of the No Tinnitus group will be above the mean of the Distressing Tinnitus group (Cohen's U3), 73% of the two groups will overlap, and there is a 68.7% chance that a person picked at random from the No Tinnitus group will have a higher score than a person picked at random from the Distressing Tinnitus group (probability of superiority)	176
Figure 5—1. Conceptual pathophysiological model of tinnitus generation and maintenance from Sedley et al. (2016).	182

Figure 5—2. Causal model of the effect of tinnitus on cognitive performance adapted from Mohamad et al. (2016).....	183
Figure 5—3. Illustration of how an 'inference engine' combines data with causal knowledge to answer queries of interest (Pearl and Mackenzie, 2018).	185
Figure 5—4. Illustration of Executive Disruption Model. A variable tinnitus percept is combined with variable latent EF and results in tinnitus distress.....	188
Figure 5—5. Illustration of compromised latent EF in Executive Disruption Model, depicting relatively more tinnitus distress for an unchanging percept following disrupted latent EF.	189
Figure 5—6. Causal model implicit in the 'loud sounds are distracting' conceptualisation.	191
Figure 5—7. General causal model proposed by Executive Disruption Model.	192
Figure 5—8. Illustration of variant Executive Disruption Model. A tinnitus percept of uniform loudness is combined with variable latent EF and results in variable tinnitus distress.	193
Figure 5—9. DAG where B is the exposure and C is the outcome. An ancestral relationship is also depicted, where B is a child of A and A is an ancestor of B.....	198
Figure 5—10. The elemental Chain DAG.	198
Figure 5—11. The elemental Fork DAG.	198
Figure 5—12. The elemental Collider DAG.	198
Figure 5—13. A DAG showing an open non-causal backdoor path from A to C through B.	200
Figure 5—14. A DAG showing a closed non-causal backdoor path from A to C through B.	200
Figure 5—15. A DAG with a closed backdoor path between A and C as B is a collider.	200
Figure 5—16. . A DAG with an open non-causal backdoor path between A and C because of conditioning on the collider B.	201
Figure 5—17. A DAG with an open non-causal backdoor path between A and C because of conditioning on D, which is a descendant of the collider B.....	201
Figure 5—18. Why the most newsworthy studies might be the least trustworthy. 200 research proposals are ranked by combined trustworthiness and newsworthiness. The top 10% are selected for funding. While there is no correlation before selection, the two criteria are strongly negatively correlated after selection (McElreath, 2016).....	203
Figure 5—19. Full DAG of Executive Disruption Model of tinnitus distress. Age = age, C_f = self-perception of cognitive failure, C_r = cognitive reserve, Dis = psychological distress, Ef = executive function, Gen = gender, H_h = hearing handicap, H_t = hearing thresholds, T_d = tinnitus duration, T_h = tinnitus handicap, T_l = tinnitus laterality, T_p = tinnitus presence, Rum = rumination.	204

Figure 5—20. Full DAG of Executive Disruption Model of tinnitus distress showing causal analysis path of interest in green. Age = age, C_f = self-perception of cognitive failure, C_r = cognitive reserve, Dis = psychological distress, Ef = executive function, Gen = gender, H_h = hearing handicap, H_t = hearing thresholds, T_d = tinnitus duration, T_h = tinnitus handicap, T_l = tinnitus laterality, T_p = tinnitus presence, Rum = rumination.	206
Figure 5—21. Full SCM of Executive Disruption Model of tinnitus distress showing causal analysis path of interest in green with added unmeasured confounds. Age = age, C_f = self-perception of cognitive failure, C_r = cognitive reserve, Dis = psychological distress, Ef = executive function, Gen = gender, H_h = hearing handicap, H_t = hearing thresholds, T_d = tinnitus duration, T_h = tinnitus handicap, T_l = tinnitus laterality, T_p = tinnitus presence, Rum = rumination, U_c = unmeasured confounds of tinnitus loudness and disrupted sleep.	207
Figure 5—22. Overlap DAG highlighting the variables that appear in both Dataset 1 and Dataset 2.	209
Figure 5—23. Principal axis factor from Dataset 1 (left) and Dataset 2 (right). Ovals depict latent executive factor, arrows show factor loading onto each task.....	211
Figure 6—1. Illustration of reciprocal link between tinnitus and cognitive problems from Tavanai and Mohammadkhani (2018).	235

1 Chapter 1 – Introduction to tinnitus and cognitive performance

1.1 Introduction

The purpose of this thesis is to investigate the relationship between tinnitus distress and cognitive performance. The first chapter will introduce the central concepts and relevant literature to be investigated within the thesis. I aim to achieve this by providing an overview of tinnitus, pathophysiological models of subjective tinnitus, and current management and intervention options. I will then introduce key terms relating to tinnitus distress and cognition, as well as models of tinnitus that pertain specifically to cognitive performance. Finally, to identify gaps in the current literature, I will undertake a critical review of literature reviews that have focussed on the relationship between tinnitus and cognitive performance. Having built a foundation for subsequent investigation, I will outline the key aims and objectives of the thesis.

1.2 Overview of tinnitus

Tinnitus is typically defined as the perception of sound in the absence of corresponding external stimuli. The perception has a variable phenomenology, with those experiencing tinnitus describing it as sounds ranging from simple tones to complex noise, and experiencing it as originating in the ears (peripherally) or “somewhere in the head” (centrally). Tinnitus is a very common experience, with most affected people reporting transient episodes during the course of their life (Henry et al., 2005).

There is no universally accepted classification system for tinnitus (Heller, 2003), however, there are several identifiable features that may indicate its underlying aetiology. Most broadly, tinnitus is parsed into objective and subjective forms. Objective tinnitus refers to a sound source that may be audible to both the individual experiencing it as well as an external observer. Causes of objective tinnitus include vascular

abnormalities, temporomandibular jaw disorders, or myoclonus of middle ear muscles. Tinnitus is also an indicative symptom of diseases such as Meniere's or Wilson's disease, and a potentially accompanying symptom of a neoplasm such as a vestibular schwannoma (Møller et al., 2010). However, in the majority of instances tinnitus is subjective (i.e. only being audible to the patient) and idiopathic in nature (Henry et al., 2005).

Epidemiological estimates of the number of individuals with tinnitus vary and have been fundamentally difficult to obtain due to a lack of objective means of identification, and a lack of standardisation amongst the definitions of tinnitus that are typically used by researchers (Baguley et al., 2012; Hunter et al., 2016). A large study in the United Kingdom estimated a prevalence of 10%, with estimates in the United States generally being 10-15% (Davis and El Refaie, 2000; Henry, 2016). Recent work within Europe has focused on the development of standardised questions for estimating tinnitus prevalence (Biswas et al., 2019), as well as the methodology for ascertaining epidemiological estimates of tinnitus, with Biswas and Hall (2020) suggesting that tinnitus affects 1 in 7 adults.

Tinnitus experienced for less than 3 months may be described as acute, but is typically described as a chronic condition if it is experienced for 6 months or more (Cima et al., 2019; Davis and El Refaie, 2000). Many individuals who experience chronic tinnitus appear to habituate to the percept, with recent evidence showing a small but statistically significant improvement in self-reported measures of tinnitus for patients enrolled into clinical trials for tinnitus, despite receiving no intervention (Phillips et al., 2018). However, there remain a significant number of those experiencing tinnitus who do not habituate to it and are sufficiently disturbed to seek medical attention. Cima (2018) notes that tinnitus is a generally a harmless auditory perception for the majority of people who experience tinnitus it; however, 3-6% of people experience distressing tinnitus.

Hallam et al. (1984) note a distinction between characteristics of the tinnitus percept per se and subsequent distress. Characteristics of the tinnitus percept have been the subject of psychoacoustic investigation and refer to analogous aspects of physical sound such as loudness, pitch, and timbre (Moore, 2012). Tinnitus distress is multifactorial and described by Hallam and colleagues as 'correlates of complaint', which includes depression, anxiety, disrupted sleep, and concentration difficulties. Tinnitus distress is also called bothersome tinnitus (or tinnitus severity) and is typically measured with questionnaires such as the Tinnitus Handicap Inventory or Tinnitus Functional Index (Cima, 2018). The remainder of this thesis will adopt a broad distinction of tinnitus percept as an aberrant internal signal, with tinnitus distress hereafter referring to the functional correlates of complaint related to the tinnitus percept.

1.3 Pathophysiological models of subjective tinnitus

1.3.1 Peripheral and established models

Early models of tinnitus were primarily described as peripheral models, which postulated the existence of aberrant activity originating from the auditory periphery, passing through the cochlear nerve and propagating to the auditory cortices (Baguley, 2002). Tinnitus was also considered to have arisen from processes such as spontaneous otoacoustic emissions, which may still be considered as a functional yet extremely rare subtype of tinnitus (Chung and Lee, 2016).

Despite recent progress in models of the pathophysiological mechanism(s) of tinnitus, the cause of subjective tinnitus remains enigmatic (Adjamian et al., 2009; Kaltenbach, 2011; Noreña, 2011; Rauschecker et al., 2010; Sedley et al., 2016; Shore et al., 2016). Current scientific consensus views tinnitus as a disorder of the central auditory system, as well as various associated non-classical pathways (Møller et al., 2010). The 'neuropsychological model' proposed by Jastreboff (1990) is a prominent model of tinnitus, which proposed that areas additional to classical auditory pathways

are responsible for the maintenance of tinnitus, including the limbic system, reticular formation, and sympathetic autonomic nervous system (Figure 1.1).

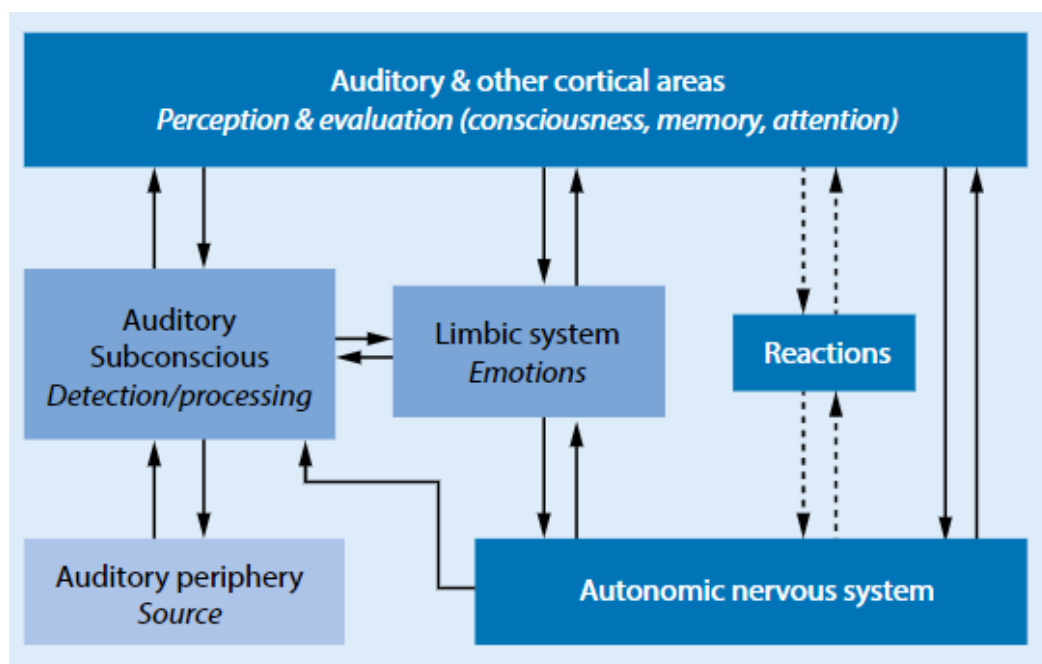


Figure 1—1. *The Neurophysiological model of tinnitus (Jastreboff, 1990), showing a schematic representation of hypothesised relationships between classical auditory pathways including the limbic system, and sympathetic autonomic nervous system.*

Investigations in animals have played an important role in the development of tinnitus theory by enabling invasive studies that are not possible in human participants. Much of this work has focused on demonstrating elevated spontaneous and sound-evoked activity in the auditory pathways of rats, ferrets, Guinea pigs, and hamsters. Eggermont and Roberts (2015) suggest that this work has a synergistic relationship with studies in humans, but note the fundamental difficulty arising from knowing if one of these animals in fact has tinnitus. The inference is typically made using either time-consuming associative learning paradigms or gap startle suppression paradigms; the latter is quicker to implement, but may be more susceptible to confounding through necessary insult to the peripheral auditory system that is required to presumably induce tinnitus in the subject. Nonetheless, there is a large body of animal evidence that has helped to develop theoretical accounts of tinnitus perception in humans.

Tinnitus is strongly associated with hearing impairment, which has led some authors to argue that hearing impairment is rarely absent when hearing assessment exceeds typical clinical sensitivities; therefore hearing impairment may be a necessary initial condition for the generation of tinnitus (Noreña, 2011; Rauschecker et al., 2010; Sedley et al., 2016).

A tinnitus percept appears to be inducible in individuals with normal hearing through acoustic deprivation in an anechoic chamber (Del Bo et al., 2008; Heller and Bergman, 1953). Heller and Bergman (1953) reported that 93.75% of normal hearing individuals described perceiving a sound while in an anechoic chamber. In a similar study 55 years later, Del Bo et al. (2008) reported that 92% of individuals also described hearing a sound in a room with a suggestive (but inactive) loudspeaker, with 83% reporting perceiving a sound in the absence of the speaker. Similar findings have been reported with less severe sound deprivation, where individuals without tinnitus wore an ear plug in a single ear for a week, approximating a mild-to-moderate high frequency hearing impairment (Schaette et al., 2012). In this study, 18 participants wore an ear plug for 7 days, with 14 individuals reporting experiencing tinnitus during that period - again, approximately 80%. Eleven participants reported the tinnitus to still be present while wearing the ear plug on the seventh day. With the exception of one person, the tinnitus had disappeared the day after removal of the ear plug, and no participant reported tinnitus a week after removal of the ear plug.

However, hearing impairment often occurs without tinnitus (Savastano, 2008), and the aforementioned anechoic and earplug studies that successfully induced tinnitus suggest that it is not hearing impairment per se that is required, but rather reduction in typical auditory input. Such observations creates difficulties for a model attempting to explain why tinnitus occurs in only some people with hearing impairment, and why it only becomes distressing in a minority of these individuals. This strongly suggests that further moderating and/or mediating variables are required for tinnitus distress to develop.

1.3.2 Spontaneous neural activity

The majority of pathophysiological tinnitus models postulate that increased spontaneous neural activity arises at cortical and subcortical areas (Noreña, 2011; Sedley, 2019).

Animal studies have reported correlates of tinnitus following exposure to noise trauma, including increased spontaneous neural activity from cochlear nucleus to auditory cortex (Kaltenbach et al., 1998; Noreña and Eggermont, 2003). Although electrophysiological and imaging studies have corroborated some of these findings, Sedley (2019) notes that such neural gain changes can be induced in response to damage to the auditory periphery irrespective of whether tinnitus is present.

Models that posit increased temporal synchronicity between populations of neurons have also been suggested to cause tinnitus through mechanisms such as ephaptic transmission (i.e. indirect communication between proximal nerve fibres or cells). This may cause temporally cohesive neural activity without an external stimulus (Møller et al., 2010, p. 78). Schaette and McAlpine (2011) have suggested that a renormalisation of neuronal response through homeostatic plasticity may cause tinnitus, after reporting a reduction in the amplitude of auditory brainstem responses for wave I (generated by auditory nerve activity), but an increase in wave V amplitude (generated auditory midbrain activity) for patients with tinnitus and a normal audiogram. Alterations in local firing rates have been suggested to induce 'maladaptive plasticity', which could alter tonotopic maps at various levels within the brain (Shore et al., 2016).

1.3.3 Contemporary models of tinnitus

Rauschecker (2010) has posited a model incorporating 'frontostriatal gating' that expands on increased subcortical activity arising from deafferentation. The model suggests that the frontostriatal system acts as a 'noise-reduction' mechanism. When dysfunctional, the normally unnoticeable tinnitus percept is allowed to reach conscious perception and is subject to limbic interactions. This model has facilitated comparisons between tinnitus and chronic pain (Rauschecker et al., 2015).

Global workspace models of tinnitus suggest that a 'minimum assembly of brain networks' must exist to bring tinnitus sensory cortex activity to consciousness.

Thalamocortical dysrhythmia has been suggested as a mechanism of achieving the required long-range brain subnetwork activity for such a model (De Ridder et al., 2014b, 2015).

Somatosensory models of tinnitus generation have also been put forward, which posit deafferentation and auditory-somatosensory interaction via the dorsal cochlear nucleus. These models have been proposed as an explanation of the observed ability of some individuals with tinnitus being able to modulate their percept via external manipulations of the head or neck (Levine, 1999; Shore et al., 2007).

Also proposed are 'filling-in' models, which are based on the concept that the 'Bayesian brain' serves as an 'inference machine', which actively seeks auditory input and makes inferences about what is likely to be heard next (De Ridder et al., 2014a). Reduced or absent input to auditory cortex (from deafferentation) therefore suggested to result in altered input from surrounding cortical areas that are still functioning normally, or from parahippocampal auditory memory (De Ridder et al., 2014a).

Sedley et al. (2016) have suggested a 'sensory precision' model of tinnitus that fits broadly within a Bayesian framework, and invokes concepts such as predictive coding and the free-energy principle (Friston, 2010). Sedley and colleagues also suggest that no single model of tinnitus to date can account for what they describe as "paradoxes" within the literature, which includes unpredictable improvement or exacerbation of the percept following auditory nerve section, occurrence of hearing loss without tinnitus, onset of tinnitus later than hearing loss, and a lack of spontaneous neural correlates in patients versus matched controls.

The apparent paradox may lie in attempts to capture both tinnitus generation, and tinnitus distress within a single model. It is noteworthy that attempts to theoretically explain such paradoxes within a single model rely heavily on cognitive concepts. For

example, attention has been implicated in tinnitus formation, maintenance, and severity (Roberts et al., 2013; Sedley et al., 2016), while other models also carve out a role for memory through implication of auditory memory in parahippocampal structures and formation of 'pervasive memory traces (De Ridder et al., 2014a; Sedley, 2019).

1.3.4 Summary

This section has presented an overview of pathophysiological tinnitus models and their development. In doing so, several noteworthy points arise: first, causal postulates of initial peripheral models are still generally assumed to be true; i.e. although, to date, tinnitus-specific hyperactivity (e.g. increased spontaneous firing or synchronicity) has been problematic to demonstrate consistently in humans, aberrant hyperactivity is typically presumed to exist *somewhere* along the peripheral to cortical auditory pathway. This theoretical line of reasoning has subsequently been reified through research in animal models. Second, many models of tinnitus have attempted to account for tinnitus generation and chronification (i.e. the chronic appearance of a percept in consciousness) and subsequent tinnitus distress. Finally, various cognitive themes are central components in these models (e.g. attention), making it increasingly important to understand the nature of their role in the pathophysiology of tinnitus.

1.4 Management and interventions for subjective tinnitus

There is currently no demonstrable cure for tinnitus, however, many interventions are available for therapeutic use. These typically target comorbidities, offering no predictable impact on the percept per se. Eggermont (2016) notes that the aetiology of subjective tinnitus appears to be less important for effective intervention than its psychological, neurological, and audiological comorbidities. Given the importance of comorbidities, a multidisciplinary approach is often required in the management of tinnitus (i.e. the pathway through assessment to intervention), which may include specialties such as general practitioners, ENT specialists, audiologists, and clinical psychologists. Current

interventions for tinnitus may be broadly split into audiological, psychological, and medical categories.

Medical interventions are usually pharmacological in nature, and typically target either comorbidities or presumed underlying aetiologies that may be causing the percept. Historically, numerous and wide-ranging pharmacological treatments have been proposed for treating tinnitus, including drugs that affect the circulatory system (e.g. vasodilators), glutamate antagonists, and even botulinum toxin (Baguley et al., 2012, pp. 149–159). However, Cochrane systematic reviews concerning specific pharmacological interventions have reported no evidence for their efficacy in directly treating tinnitus (Baldo et al., 2012; Hoekstra et al., 2011; Wegner et al., 2018).

Psychological therapies for tinnitus typically entail modifying a person's reaction to their tinnitus percept. Cognitive behavioural therapy (CBT) is the most evidenced psychological approach to treating tinnitus. It aims to facilitate habituation to the percept through "remodelling" and "challenging" unhelpful thoughts and consequent behaviours. There is reasonable evidence that CBT has a positive effect in the management of tinnitus (Fuller et al., 2020).

Audiological interventions may include the use of hearing aids or sound generators. Hearing aids amplify sounds to restore speech intelligibility, while sound generators attempt to 'mask' the tinnitus percept. To date, Cochrane reviews have suggested that there is no evidence to refute or support the usage of either as an intervention for tinnitus (Sereda et al., 2018).

A recent study found that access to available interventions was variable, and was essentially dependant on local service practise and availability (McFerran et al., 2018). However, the National Institute of Clinical Excellence has since published guidance on tinnitus assessment and management in the UK (NICE, 2020), which may lead to increased standardisation of service delivery, access to therapeutic management, and treatment options. This goal is part of a wider research agenda for such standardisation

and improved access both in the UK and across Europe. For example, Cima et al. (2019) recently published multidisciplinary European guidelines for diagnostics, assessment, and treatment of tinnitus, while other researchers are working towards broadly accessible interventions such as internet-delivered CBT (Beukes et al., 2018, 2017).

A further potential intervention for chronic subjective tinnitus is perceptual and cognitive training (Hoare et al., 2010). Some studies have suggested that generalised cognitive training for people with subjective tinnitus could provide therapeutic benefit (Kallogjeri et al., 2017; Krings et al., 2015). However, it is necessary to improve understanding of cognitive domains associated with tinnitus to achieve efficacy with cognitive-related interventions. Interventions based on an enhanced understanding of the cognitive domains associated with tinnitus would, therefore, bolster current management strategies.

1.5 Tinnitus distress, cognition, and cognitive performance

Mohamed et al. (2016, p. 2) note that the term concentration is typically a “lay person’s label for his/her personal experience”; however, its associated difficulties are typically discussed using the term ‘cognition’ within the tinnitus literature. This is a broad term and can be used inconsistently due to its varied historical applications and use within common discourse, as well as the fields of philosophy and psychological science (Brandimonte et al., 2006). Cognition is not a monolithic entity and is often conceptualised as comprising of many interrelated systems that operate in unison to give rise to cognitive function (Craik, 2007). Indeed, Baguley et al. (2012) propose that cognition can be broadly discussed with regards to the contents of thoughts or the efficiency of thought processes. The use of the term within the tinnitus literature is typically associated with CBT and has subsequently been more closely linked with contents than efficiency.

The efficiency of cognitive systems can be probed with tasks that are thought to index specific elements of theoretical constructs such as speed of information processing (Salthouse, 1996). Use of such cognitive tasks is well established within the psychological literature, and has been combined with conceptual models and brain imaging studies that provide the current theoretical underpinnings of separable, but overlapping systems, such as working memory, long-term memory, and executive functions (Baddeley, 2012; Diamond, 2013; Miyake et al., 2000; Petersen and Posner, 2012). The efficiency of operation of such systems is arguably best captured by the term 'cognitive performance'.

Cognitive, attentive, and concentration difficulties have been found to be an important concept to people with distressing tinnitus (Colagrosso et al., 2018; Mohamad et al., 2016). For example, a recent synthesis by Hall et al. (2018) aimed to provide a comprehensive and authoritative collection of reported problems in everyday life for adults with tinnitus and their significant others. The review performed a comprehensive search of peer-reviewed and grey literature databases, and identified 84 records that detailed complaints generated from the perspective of both patients with tinnitus and their significant others. The authors reported good convergence between complaints collected using open- and closed-format question, finding that 'concentration difficulties', 'ability to ignore tinnitus', and 'cognitive difficulties' were frequently reported under the category of 'Functional difficulties due to tinnitus'.

1.5.1 Models of tinnitus and cognitive performance

While many established models of tinnitus distress feature cognitive components (Cima, 2018; Cima et al., 2019), the following section introduces models that pertain specifically to the relationship between tinnitus and cognitive performance.

1.5.1.1 Andersson and McKenna (2006) - Road to annoyance model

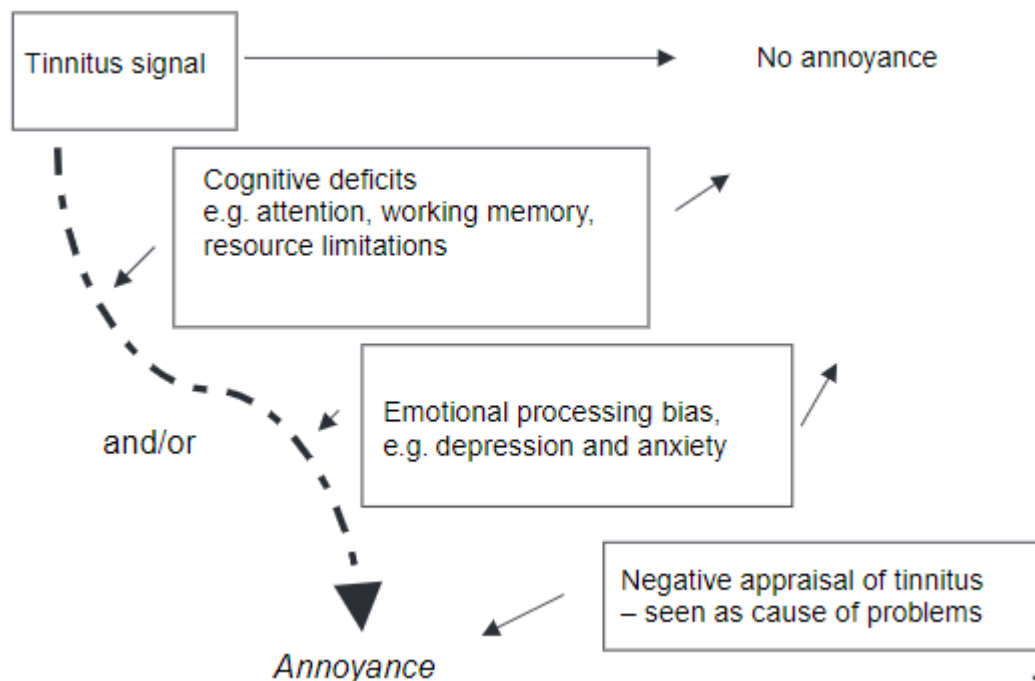


Figure 1—2. 'A model of the road to annoyance via different levels of cognitive functioning' from Andersson and McKenna (2006). The dashed line depicts a path through tinnitus and various "escape routes" towards no annoyance.

Andersson and McKenna (2006) proposed a model that was based on the limited behavioural evidence available at the time and clinical observation/intuition (Figure 1.2). The model essentially describes a "road to annoyance via different levels of cognitive functioning". It proposes cognitive involvement over various levels of cognitive functioning, and suggests corroborating behavioural evidence for each level. The model suggests that cognitive deficits in information processing (resulting from resource limitations due to tinnitus interference) affect attention and working memory. It also suggests that cognitive bias influences emotional processing, linking selective attention to anxiety, and memory to depression. Andersson and McKenna propose subgroups of tinnitus distress could exist based on these biases. Conscious appraisal of tinnitus is also implicated; this is not meta-cognition per se, but more thoughts and beliefs related to tinnitus, and therefore relates to the contents of thought rather than efficiency of process.

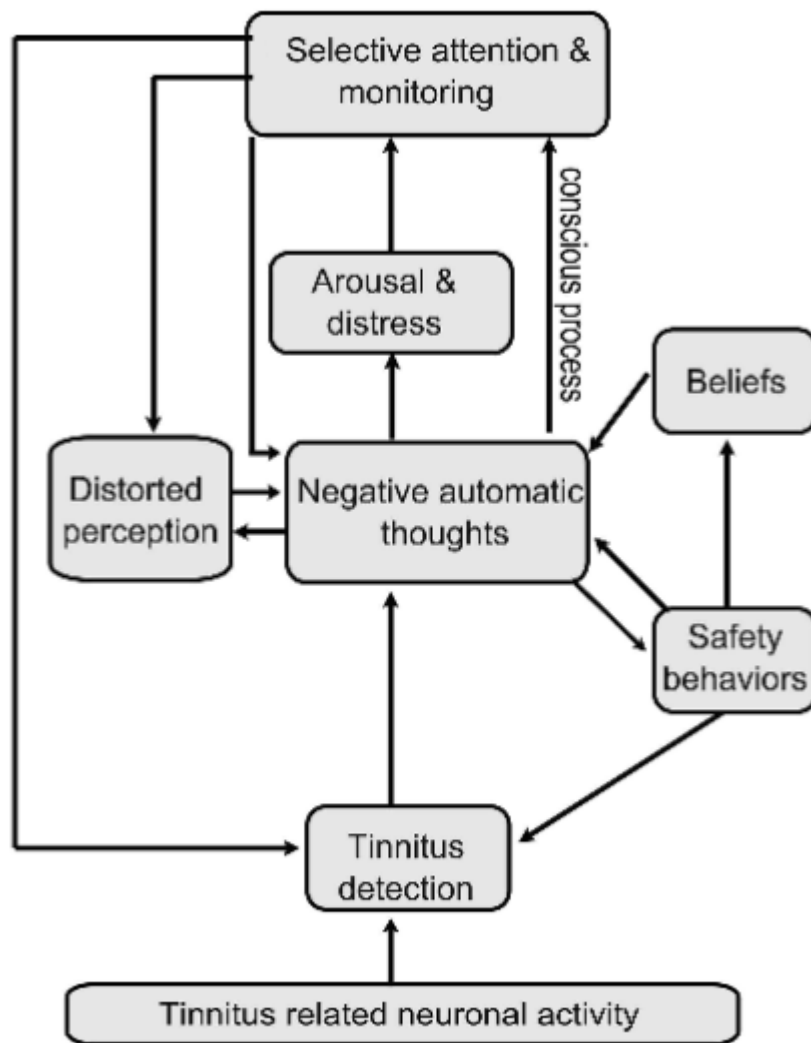


Figure 1—3. The 'Cognitive Behavioural Model' of tinnitus distress from McKenna et al. (2014).

McKenna et al. (2014) proposed a broad cognitive model and aimed for it to be scientifically testable (i.e. that generated hypotheses that can be falsified). The model (Figure 1.3) links tinnitus to cognitive performance by suggesting that selective attention deficits may exist in people with tinnitus. It also posits a role for sustained attention and monitoring; although the link to sustained attention is less detailed within the model, it is suggested to arise through hypervigilance. Both processes are hypothesised to lead to distorted tinnitus perception (notably its loudness), leading to “overly negative” appraisal of the percept. In addition to commenting on the importance of testable hypotheses, McKenna and colleagues also highlighted the importance of considering causality within

tinnitus models, and subsequently suggested structural equation modelling as a means of testing causal claims.

1.5.1.3 Mohamad et al. (2016) - Structural Model

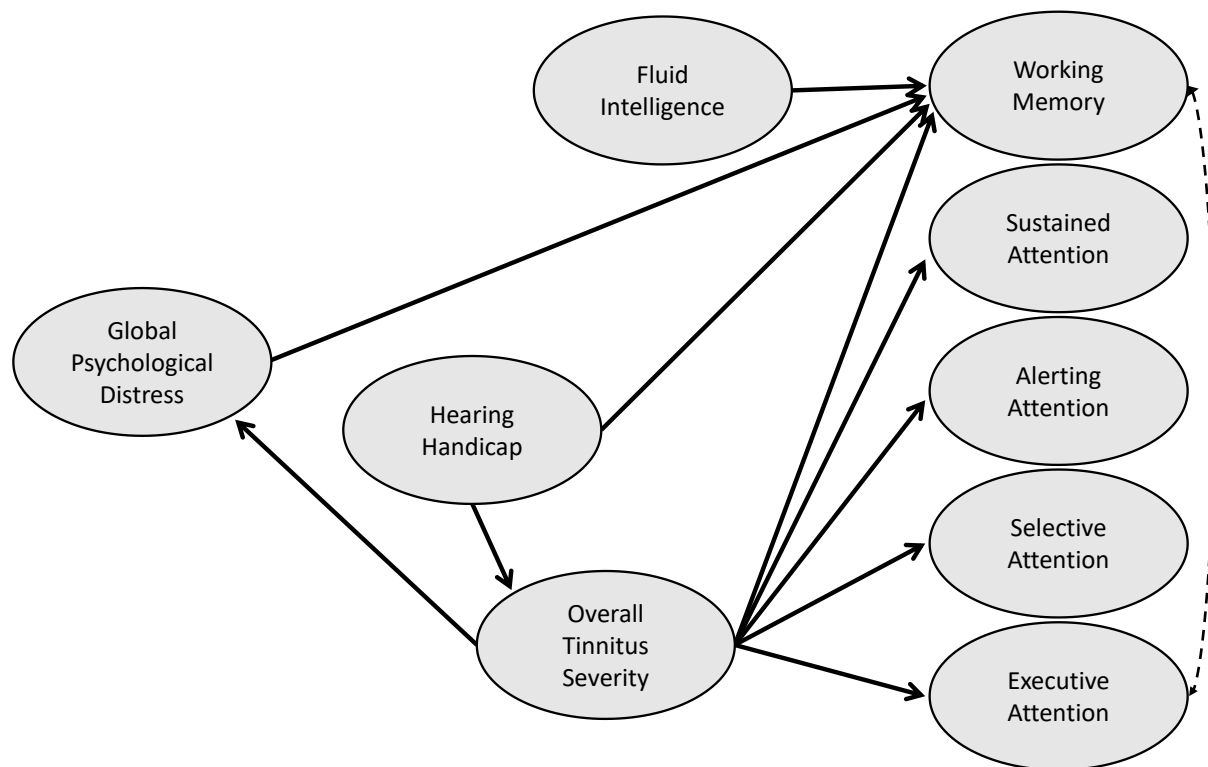


Figure 1—4. Structural model depicting the proposed effects of tinnitus severity on cognitive performance adapted from Mohamad et al. (2016).

Following a review of the consequences of tinnitus and tinnitus severity on cognition, Mohamad et al. (2016) proposed a “best informed guess” concerning various cognitive components and their interrelationships; these included fluid ability, working memory, sustained attention, alerting attention, selective attention, and executive attention (Figure 1.4). The model suggests that tinnitus severity impacts various elements of cognitive performance (whether this is severity of the percept or related distress is not specified). The model also demonstrates the increasing relevance of causal hypotheses and highlights that structural equation modelling and related techniques can be used to test causal claims relating to this research topic (Handscomb et al., 2019; McKenna et al., 2014; Mohamad et al., 2016).

1.5.1.4 Trevis et al. (2016) - Neurocognitive model

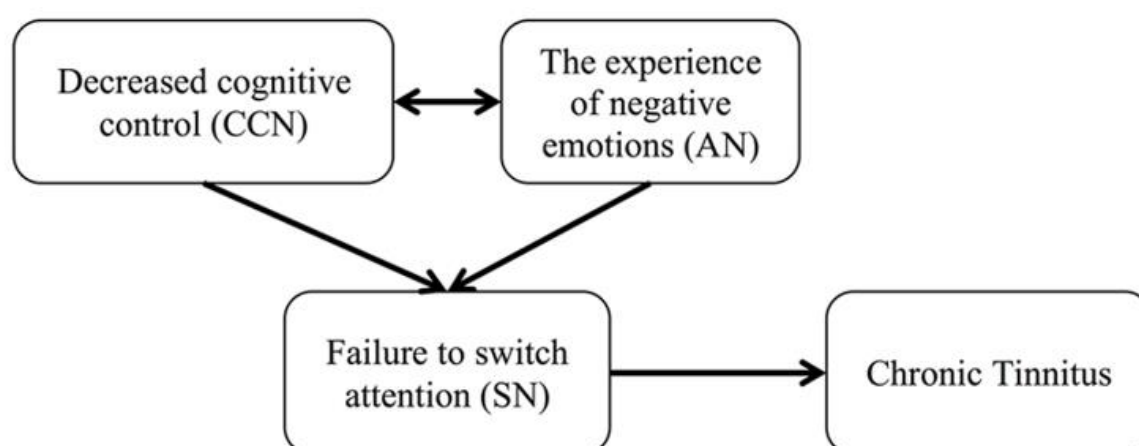


Figure 1—5. A 'Neurocognitive model' that links cognitive performance to chronic tinnitus from Trevis et al. (2016).

Using insights from functional magnetic resonance imaging and behavioural tasks, Trevis et al. (2016, 2017b) identified various cognitive networks and proposed a neurocognitive model of tinnitus, which posits over- and under-activation in various neural networks is responsible for chronic tinnitus. This model (Figure 1.5) specifically suggests dysfunction through a hyperactive autobiographical memory network (leading to increased rumination) and hypoactive cognitive control network (resulting in memory and attention deficits). Trevis and colleagues suggest that when viewed this way, chronic tinnitus may reflect a fundamental failure of the cognitive control network to inhibit the tinnitus percept if it is emotionally salient to the individual, as well as reduce proficiency at “switching attention away from the tinnitus sound compared to people with healthy hearing”.

1.5.1.5 Summary

Multiple studies have reported deficits in cognitive performance in individuals with tinnitus across a number of cognitive domains including attention, processing speed, and working memory performance (Das et al., 2012; Heeren et al., 2014; Rossiter et al., 2006). Consequently, various models posit a necessary role for cognitive performance in the generation, and subsequent chronification of the percept, and/or its ability to cause distress. Similarly, current theoretical models of tinnitus pathophysiology also suggest

that cognitive concepts such as attention may be a crucial component of tinnitus (Sedley et al., 2016); however, to date findings from primary research have presented inconsistent evidence relating to impairment of cognitive performance in people with tinnitus. Understanding the proposed link between tinnitus and cognitive performance is, therefore, important and there is a need for effective synthesis to consolidate current knowledge on this topic.

1.5.2 Review of cognitive performance and tinnitus reviews

Several reviews have addressed the association between tinnitus and cognitive performance (Andersson and McKenna, 2006; Mohamad et al., 2016; Tegg-Quinn et al., 2016; Trevis et al., 2017a). While systematic reviews are desirable, there are many methods of evidence synthesis and a wide range of literature reviewing methods; the overlap between these methods can make it difficult to distinguish exact criteria for a systematic review, and in many cases this can be left in the eye of beholder (Munn et al., 2018; Schick-Makaroff et al., 2016). Furthermore, reviewing methodology and associated nomenclature is developing with a quickening pace that presents additional barriers to understanding for those unfamiliar with the field. Not all research synthesis is equal, and it is becoming increasingly important to understand related capabilities and limitations inherent in synthesis methods.

Grant and Booth (2009) have suggested a pragmatic framework that focuses on the key areas of review development: search, appraisal, synthesis, and analysis (SALSA). The SALSA framework has identified 14 types of review that are characterised by variations in these key areas of the reviewing process, and provides a useful lens for focusing discussion concerning previous syntheses of tinnitus and cognitive performance.

Table 1—1. Overview of previous reviews investigating cognitive performance and tinnitus.

Study	Databases searched	Example search strategy provided	No. of records retrieved during database search	No. of studies included in review	SALSA typology	Synthesis of association between cognition and tinnitus
Trevis et al. (2018)	PsycINFO, MedLine	(1) tinnitus AND psych* (all fields), (2) tinnitus AND mood (all fields), (3) tinnitus AND depress* (all fields), (4) tinnitus AND anx* OR stress (all fields)	725	35	Systematic /Critical Review	Meta-analysis of 'psychological functioning' in tinnitus participants. Narrative review of association between 'cognitive functioning' (n = 16); i.e. behavioural cognitive task performance and chronic tinnitus. awareness and severity
Mohamad et al. (2016)	PubMed	(((((tinnitus[Title]) AND cogniti*[Title])) OR ((tinnitus[Title]) AND attention[Title])) OR ((tinnitus[Title]) AND memory[Title]))	65	9	Mapping Review/ Systematic Map	Narrative review of 'behavioural research' addressing the impact of tinnitus and its severity of various aspects of cognitive performance in domains of working memory and attention.
Tegg-Quinn et al. (2016)	PubMed, MedLine, CINAHL, SCOPUS, EMBASE	(tinnitus) and (cognition OR memory OR attention OR concentration OR cognitive function OR mental activity) NOT (infant OR child OR adolescent OR paediatric OR animal OR balance OR hyperacusis OR implant OR pharmaceutical OR drugs)	2236	18	Systematic Review	Narrative review of behavioural cognitive tasks, electrophysiological correlates of cognition and self-reported cognitive function measures.

Andersson and McKenna (2006)	MedLine, Psychological Abstracts	Not reported	Not reported	Not reported	Critical Review	Narrative review of 'cognitive deficits' (i.e. behavioural cognitive tasks), 'cognitive bias and 'conscious appraisal of tinnitus'.
------------------------------	----------------------------------	--------------	--------------	--------------	-----------------	---

When applying the SALSA framework to the reviews that have previously synthesised measures of tinnitus and cognitive performance, it is evident that no review has systematically synthesised all of the available literature and quantitatively analysed the results (Table 1). Andersson and McKenna (2006) described their review as a 'selective systematic review', with the stated aim of reviewing the literature at the interface between tinnitus and cognitive function. The review was organised with reference to different levels of information processing. Grant and Booth (2009) suggest that typological characteristics that define this style of review include narrow database searches, lack of an explicit study quality appraisal, and conceptual organisational synthesis. Andersson and McKenna's review would, therefore, be deemed a critical review according to the SALSA typological standards. However, the authors note that there were relatively few available studies on this topic at the time of publication.

Mohamad et al. (2016) performed a review of the behavioural evidence concerning the consequences of tinnitus and its severity on cognition. This review included details about the scope of literature searches (providing time and scope constraints), provided a formal quality assessment of the likelihood of bias included studies, featured graphical and tabular synthesis, and identified further research developments and requisite methods. According to the SALSA synthesis framework, this synthesis should be considered a 'mapping review' or 'systematic map'.

Tegg-Quinn et al. (2016) published a systematic review of the impact of tinnitus upon cognition in adults. The review fits requirements for the majority of typological characteristics required for a systematic review. Analysis was undertaken predominantly through a clinical perspective with its stated aims being 'to facilitate accurate

informational counselling' and 'assist effective rehabilitation'. These aims are arguably too broad to be effective as a means of facilitating a systematic review, which requires a focused and concise question. Furthermore, this review extends beyond cognitive performance measured via behavioural tasks to include other associated measures of cognitive performance; it extracted data from 18 records, including 24 behavioural tests, 9 electrophysiological tests, and an oculomotor test. The broad stated aims and inclusion of various measurement types presented challenges to synthesis that was noted by the authors: "The significant heterogeneity of study designs, cognitive tests, subject characteristics, and management of covariables in the 18 studies reviewed makes it difficult to compare and contrast the studies directly" (Tegg-Quinn et al., 2016, p. 539).

Finally, Trevis et al. (2017a) undertook a systematic review and meta-analysis of 'psychological functioning' in tinnitus. While this fulfilled most typological characteristic requirements, their focus was on the broad construct of psychological functioning (conflating emotional wellbeing and cognitive functioning), with the quantitative synthesis being limited to 'emotional wellbeing' in participants with tinnitus (i.e. anxiety and depression). Evidence for an association between tinnitus and its severity and cognitive performance through behavioural measures was subject to a narrative critique under the heading of 'cognitive functioning'. The authors suggested a 'limited number of studies using similar tasks or assessing common cognitive domains' as reason for a meta-analysis on the topic being inappropriate. A potential issue with this assertion is the limited scope of the searches performed, which included only MedLine and PsycINFO. According to the SALSA framework, a typological requisite of both systematic review and meta-analysis is 'exhaustive, comprehensive searching'. Trevis et al. (2017a) arrived at their conclusion of an association between tinnitus and cognitive performance through informal vote counting or narrative synthesis; their review of this specific aspect of the literature can only reasonably be considered a critical review, being neither comprehensive in search scope or providing quantitative synthesis.

Taken together, previous reviews investigating the relationship between tinnitus and cognitive performance have all reported mixed evidence for an association, and/or challenges in synthesising the available literature. Suboptimal typology of applied synthesis methods may account for mixed results and analytic difficulties encountered. Furthermore, conclusions of mixed evidence may be due to all reviews on this topic to date having used an informal (i.e. unspecified) method of 'traditional vote counting' to form conclusions (i.e. tallying up the statistically significant versus non-significant conclusions). Evidence from the field of research synthesis demonstrates that this method lacks power (Bushman and Wang, 2009) and, at worst, can lead to the wrong conclusion (Ellis, 2010; Friedman, 2001). Previous reviews have also not typically included a quantitative synthesis of the available literature concerning cognitive performance (i.e. meta-analysis). To date, only Trevis et al. (2017a) used meta-analysis within their review; however, only emotional wellbeing variables (i.e. anxiety and depression) were subject to quantitative synthesis.

Typology aside, reviews to date have not been comprehensive in the scope of available evidence concerning associations between tinnitus and cognitive performance. This is evidenced by the variability in the number of studies included in previous reviews, and the limited degree of overlap in included studies. For example, although three reviews were published within a year of each other Mohamad et al. (2016) reviewed 9 studies and Trevis et al. (2017a) reviewed 35 studies (with 16 of these concerning cognitive performance), yet only 3 studies were included in both reviews. It is likely that differences in search strategies (i.e. the search terms that were used and the electronic databases that these were performed in) were primarily responsible for the lack of overlap. Furthermore, tinnitus is a heterogeneous condition, making it inherently difficult to define and specify in the context of a systematic review. Different inclusion criteria and working definitions of tinnitus are likely to influence the number and type of records included within a review. For example, Trevis et al. (2017a) defined 'chronic tinnitus' as experience of tinnitus for at least 1 month. An alternative approach would be to not

attempt to temporally specify a population, but rather to investigate this variable through further quantitative analysis if feasible. Therefore, previous work has essentially investigated the association between tinnitus and cognitive performance with different datasets (Table 1.1).

A further important area of consideration is subjective cognitive complaints (SCC). To date, only Mohamad et al. (2016, p. 5) have considered SCC, noting that “the more successful we become at examining part of the cognitive system in isolation, the less our data are likely to tell us about cognition in everyday life. If self-reported difficulties in mental concentration are indeed attributable to a negative effect of tinnitus on cognitive performance, then it is important for such a link to be supported by empirical evidence”. However, the authors concluded that there was not enough evidence to form a conclusion about the relationship between SCC and severity of tinnitus.

An improved understanding of cognitive performance in participants with tinnitus has the potential to provide corroborating functional evidence for brain structures implicated in tinnitus generation and maintenance. This is particularly important given the role that attention plays in many prominent theories of tinnitus generation and its subsequent maintenance (Roberts et al., 2013; Sedley et al., 2016). Furthermore, there is considerable research interest in training cognitive abilities (Diamond & Ling, 2016), with training of associated cognitive domains representing a promising yet unexplored means of therapeutic management, both for amelioration of concentration difficulties, and for the impact of such training on the tinnitus percept. It is important to understand which cognitive domains are associated with tinnitus before an effective investigation of potential therapies can begin.

Furthermore, with the exception of executive attention, previous reviews have summarised the relationship between tinnitus and cognitive performance in various domains, such as working memory, as having mixed results (Mohamad et al., 2016;

Tegg-Quinn et al., 2016), potentially leaving researchers and clinicians uncertain as to whether tinnitus is associated with poorer cognitive performance.

In summary, although several studies have reviewed the link between tinnitus and cognition, there has been little overlap between included studies and types of research synthesis applied to existing evidence. The findings of previous reviews therefore highlight the difficulties in drawing firm conclusions about the current evidence solely through narrative synthesis, which has noted shortcomings such as author bias, opacity of methods and quality assessment, and questionable comprehensiveness. No review to date has specifically reviewed the association between subjective tinnitus and cognitive performance through meta-analysis.

1.6 Thesis aim and objectives

The aim of this thesis is to investigate the relationship between tinnitus distress and cognitive performance to enhance theoretical understanding of the cognitive associations with tinnitus, and potentially inform development of cognitive therapeutic interventions.

The specific objectives are:

- i. to undertake a comprehensive, systematic review and meta-analysis of the literature concerning tinnitus and cognitive performance, and to quantitatively synthesise the available data to identify specific cognitive domains associated with tinnitus (Chapter 2)
- ii. to subsequently investigate the relationship between tinnitus distress and associated domains of cognition (Chapter 3)
- iii. to establish whether tinnitus distress is independently associated with cognitive performance (Chapter 4)
- iv. to propose and validate a model of cognitive performance and tinnitus distress in an independent dataset (Chapter 5)

- v. to examine conceptual and methodological issues relating to tinnitus distress and cognitive performance, and identify areas for novel research into interventions (Chapter 6)

2 Chapter 2 – Associations between subjective tinnitus and cognitive performance: systematic review and meta-analyses

2.1 Introduction

The purpose of this chapter is to present a comprehensive systematic review and meta-analyses of associations between subjective tinnitus and cognitive performance, which addresses limitations of previous reviews on this topic that were identified in the previous chapter. The analysis presented here was published in the peer-reviewed journal *Trends in Hearing* (Clarke et al., 2020). The first part of the chapter aims to provide an overview of meta-analysis and established meta-analytic processes. Methodological issues related to meta-analysis, and how these issues were addressed will also be discussed, followed by presentation of the analysis.

2.2 Meta-analysis overview

2.2.1 Background

A common technique for reviewing medical literature is the “classic” narrative review (Grant and Booth, 2009), where various studies that are relevant to the review question are selected at the discretion of the author(s), and pertinent aspects are subsequently critically appraised and summarised. While this style of evidence synthesis remains common, it has faced criticism as a primary means of scientific research synthesis due to various limitations including a lack of systematicity, potential author bias, and a lack of transparency (Quintana, 2015). Effective research synthesis of complex topics requires methodological rigour that is at least equal to that of primary research (Chalmers et al., 2002).

Meta-analysis is the quantitative, statistical analysis of effect sizes generated within primary studies. When combined with systematic review, it is widely considered

the gold standard for research evidence synthesis and sits at the top of the NHS England hierarchical pyramid for evidence standards (BNSSG Research and Evidence Team, 2018; Shamseer et al., 2015). Because meta-analysis increases sample size through aggregation, it has been described as allowing researchers to “arrive at conclusions that are more accurate and more credible than can be presented in any one primary study or in a non-quantitative, narrative review (Rosenthal and DiMatteo, 2001).

The concept of meta-analysis faced criticism during its inception, however, modern critiques tend to focus on its misapplication rather than conceptual validity (Stegenga, 2011), with Rosenthal and DiMatteo (2001) noting that all criticisms that can be applied to its usage are also applicable to traditional narrative techniques. Subsequently, there has been widespread adoption of meta-analysis, and numerous authors have suggested that it is a crucial method for summarising knowledge of effects in various disciplines (Cumming, 2017; Ellis, 2010).

2.2.2 Meta-analytic process

Ellis (2010, p. 97) suggests that there are six steps in conducting a meta-analysis. They are used here as a basis for providing an overview of the meta-analytic process:

1. Collect studies – despite electronic databases having made the process of collecting studies more efficient, relying on data solely from the peer-reviewed literature may give a distorted picture due to publication bias. Databases containing grey literature such as working papers, conference abstracts, or PhD theses may also be consulted and included within the review. The study collection process should be adequately documented.
2. Code studies – relevant information to encode from studies should be decided on *a priori* and extracted by multiple members of a review team to form a database of information that can be used for the meta-analysis. Despite potentially being considered a mundane subject, knowing exactly what information is required is a crucial step within the meta-analytic process. Piloting data collection forms can

help to refine the coding process and increase cohesion across members of a reviewing team.

3. Calculate a mean effect size – Rosenthal and DiMatteo (2001) describe effect sizes as “Chief Coins of Meta-Analytic Realm”. Deciding on which effect size should be reported is not a trivial issue, particularly as standardised mean differences and correlations may be converted between one another (Lipsey and Wilson, 2000; McGrath and Meyer, 2006). Choosing and calculating an effect size statistic should therefore be performed with reference to the aims of the analysis and to facilitate subsequent interpretation.
4. Compute the statistical significance of the effect size statistic – in traditional analyses, computing the statistical significance of the mean is a choice between fixed- and random-effects models, with the fundamental difference between the two lying in the assumptions regarding the population from which studies are drawn (Kock and Gemünden, 2009). A fixed-effects model assumes that the population effect size is identical for all studies, while random-effects models assume that studies are drawn from a larger population of studies, with the underlying effect size varying from study to study.
5. Examine the variability in the distribution of effect size estimates - a powerful, yet perhaps underappreciated aspect of meta-analysis lies in its capacity to identify and explain inconsistencies, and facilitate discovery of moderators and mediators in bodies of research (Rosenthal and DiMatteo, 2001). This can be achieved through scrutiny of confidence intervals, homogeneity/heterogeneity statistics, and meta-regression analyses.
6. Interpret the results - interpretation of the results is a challenging process, and popularised guidelines exist for interpretation of the magnitude of standardised effect sizes (such as Cohen’s guidelines for ‘small’, ‘medium’, and ‘large’ effects). However, domain-specific knowledge is ultimately required for effective interpretation and summary effect sizes should therefore be considered in this

context (Cohen et al., 2003; Correll et al., 2020; Funder and Ozer, 2019; McGrath and Meyer, 2006)

While “how to” guidance for performing meta-analysis generally concurs with the outlined process (Lipsey and Wilson, 2000), other sources suggests that transparent, reproducible meta-analyses should begin at an earlier stage through protocol registration on a database such as PROSPERO, and end with standardised reporting methods, such as Preferred Reporting Items for Systematic Review and Meta-analyses (Moher et al., 2009; Quintana, 2015).

2.2.3 Applying meta-analysis to associations between tinnitus and cognitive performance

Effectively applying meta-analysis to investigate associations between tinnitus and cognitive performance required consideration of established criticisms when designing the systematic review; therefore, noted criticisms and solutions to mitigate them were considered in the design of the meta-analysis, and were detailed in an analysis protocol prior to commencing the reviewing process. The complete protocol was published a peer-reviewed journal (Clarke et al., 2018).

Criticisms of meta-analysis are perhaps best thought of as criticisms of research synthesis more generally (Rosenthal and DiMatteo, 2001). Nonetheless, debate concerning the conceptual validity and quality of meta-analyses has refined criticism of evidence synthesis and yielded established methodological issues that should be considered as part of an effective review. These can be summarised as the problems of “apples and oranges”, “garbage in – garbage out”, and the “file draw problem” (Kock and Gemünden, 2009).

Early criticism of meta-analysis noted that studies typically vary in their operationalisation and measurement of variables, and so meta-analysis attempts to aggregate study results that are incommensurable (Franke, 2001). This criticism is generally referred to as comparing apples and oranges. Analyses can therefore restrict admitted studies and increase validity or increase generalisability through consideration of heterogeneity and moderator variables to investigate important difference between

primary studies (Franke, 2001; Kock and Gemünden, 2009; Lipsey and Wilson, 2000, p. 9). Categorising cognitive tasks for meta-analysis presented a major challenge, and due to their interrelatedness, has been compared to 'slicing smoke' (Mather and Jaffe, 2016). The apparent flexibility in ascribing a task to various domains of cognition may allow the researchers' own bias to enter the process. To limit this, it was decided to use an established cognitive taxonomy that contained an appendix of cognitive tasks that had already been categorised (Webb et al., 2018). This method had various advantages including building on previous research to expedite the review process through categorisation of the tasks, subsequently reducing author bias, and offering established theoretical links to the cognitive psychological literature.

While restriction of studies may be applied to the problem of garbage in – garbage out, the generality of the analysis would then be reduced, as well as introducing author bias into the analysis (Kock and Gemünden, 2009; Lipsey and Wilson, 2000, p. 9). An alternative approach is to include all studies regardless of their perceived methodological quality. A rating system for methodological quality can then subsequently be applied and included in the analysis. Sensitivity analyses can then demonstrate whether the substantive findings of the meta-analysis were altered by studies that had certain methodological quality or characteristics. It was decided that generality and comprehensiveness of this systematic review should be afforded due consideration. Therefore no predetermined study quality exclusion criteria were set. Instead, a study quality grading tool was created for the review (Gough, 2007; Weed, 2005) and used to rate aspects of studies that were crucial to answering the review question including its methodological quality, methodological relevance, and topic relevance.

The file drawer problem refers to the effect of selection bias that is inherent in the established publication selection process. This arises because studies that report a result of achieving statistical significance are more likely to be published than those that do not obtain a statistically significant result (Kock and Gemünden, 2009; Simonsohn et al.,

2014). To limit the impact caused by the 'file drawer problem', various methodological and practical solutions were implemented in the design of the meta-analysis. This included creating a broad and encompassing search strategy through consideration of included search terms and their application to specific electronic databases. Efforts were also made to identify grey literature, and study authors were contacted to request information that was not included within study manuscripts. Finally, assessment of the likelihood of publication bias affecting the meta-analysis was performed through various graphical and statistical methods such as Egger's regression, funnel plots, and Baujat plots (Baujat et al., 2002; Quintana, 2015; Sterne et al., 2011).

In addition to established criticisms of meta-analysis, reproducibility of psychological research has become a prominent issue in recent years, and was therefore considered in the design of this review (Open Science Collaboration, 2015). This is a controversial and nuanced topic and a detailed discussion is outside the scope of this section. However, critics have argued that reproducibility of research is propagated through questionable research practises such as "hypothesising after the results are known", and an excess of potential analytic methods/strategies or "researcher degrees of freedom", giving rise to issues such as "p-hacking", where analytic practises are biased towards those that will yield statistically significant results (Quintana, 2015; Simonsohn et al., 2015; Wicherts et al., 2016). Various efforts were made to attempt increase confidence in this analysis in light of such issues. These included pre-registration of the meta-analysis on PROSPERO, providing a published protocol in a peer-reviewed journal, and categorising cognitive tasks using an existing taxonomy to limit researcher degrees of freedom and perceived bias. Furthermore to provide transparency, the search history, data, and code for analyses are available on the Open Science Framework (https://osf.io/gk5zh/?view_only=d2fa693eced945ea9e12c5dcce2d3957).

Having presented an overview of meta-analysis and described the key considerations and decisions in its application to this research, the rest of the chapter presents the results of the systematic review and meta-analyses. This is the first review to apply a

cognitive taxonomy to connect theorising regarding tinnitus and cognitive performance to current theories within cognitive psychology. The aim was to undertake a comprehensive systematic review and meta-analyses, to answer the question: 'is subjective tinnitus associated with poorer cognitive performance?' To answer this question, an established cognitive framework (Webb et al., 2018) was used to analyse:

- Which broad domains of cognition are associated with tinnitus
- Which narrow domains of cognition are associated with tinnitus
- How strong the associations are between tinnitus and performance in various cognitive domains
- What the quality of the available evidence is
- Which moderating variables confound any identified associations between cognitive performance and tinnitus

2.3 Methods

The review is reported according to Preferred Reporting Items for Systematic Review and Meta-analysis (Moher et al., 2009) and published in a peer-reviewed journal (Clarke et al., 2020). The review protocol was pre-registered on the PROSPERO database (CRD42018085528), and published in a peer-reviewed journal (Clarke et al., 2018).

2.3.1 Databases and search strategy

Electronic searches for peer-reviewed journal articles were performed in MEDLINE (via Ovid SP), EMBASE (via Ovid SP), PsycINFO (via Ovid SP), ASSIA (via ProQuest), Cumulative Index to Nursing and Allied Health Literature or CINAHL (via EBSCO Host), Scopus, PubMed, and Web of Science (Science and Social Science Citation Index). Grey literature including PhD theses and conference proceedings were eligible for inclusion. However, all studies identified in grey literature were subsequently matched to peer-reviewed publication. We therefore reported the peer-reviewed version of the study record only.

Search terms entered into the databases were identified using free text, controlled vocabularies (e.g. Medical subject headings – MeSH, and CINAHL Headings), literature review, opinion of authors, and scrutiny of pilot search results. A record of search activity and the relevant search strings for each database is available at Open Science Framework (https://osf.io/gk5zh/?view_only=d2fa693eced945ea9e12c5dcce2d3957). Hand searches of key journals (identified via electronic database searches and author discretion) were also performed. Searches were conducted in February 2018, and updated in January 2019.

2.3.2 Inclusion criteria

Published or in-press studies, written in English or with an available English translation, were eligible for inclusion. No date restrictions were applied to searches. Review inclusion criteria were specified according to Participant, Intervention/Interest, Comparator, Outcome, and Setting (PICOS):

Participants

Studies including adults (≥ 18 -years-old) with subjective tinnitus. Studies that included both children (< 18 -years-old) and adults were excluded, unless the adult data were reported separately.

Intervention/Interest

Studies including participants with self-reported tinnitus.

Comparator

Studies reporting at least one established measure of cognitive performance (behavioural or self-report).

Outcome measures

Studies reporting a measure of association between tinnitus and cognitive performance or studies containing potential requisite descriptive data to calculate an association.

Study design

Cross-sectional, longitudinal, experimental, quasi-experimental, and observational studies were included.

2.3.3 Article screening and selection

Article screening and record management was performed using Covidence systematic review software (<https://www.covidence.org>). Titles and abstracts of records were independently screened by at least two members of the review team and discrepancies discussed. Records required consensus by both reviewers to be taken forward for full text review. A third member of the review team adjudicated in instances where consensus was not obtained between initial reviewers. Articles were then taken forward to full text screening, including articles where there was too little information to make an initial decision via title and abstract screening. Included records were those that met the inclusion criteria following full text review.

2.3.4 Data extraction and data items

Included records were subjected to data extraction using a data extraction form developed by the review team. Each record was subject to data extraction independently by two members of the review team. Extracted data items were then compared across reviewers for accuracy and consistency. In instances where extracted data items differed for a record, discussion and comparison to clarify extracted data was undertaken between the reviewers.

2.3.5 Weight of evidence (WoE) – record bias and quality

Records were rated according to their relevant risk of bias and study quality using a Weight of Evidence (WoE) framework (Gough, 2007; Weed, 2005). A WoE tool was devised by the review team (Appendix A) and each record was rated independently by two reviewers. In instances where WoE ratings between reviewers did not agree, discussion was undertaken to clarify rating rationale until consensus was achieved. WoE ratings between members of the review team displayed excellent agreement with 2/3 of

ratings achieving complete agreement on initial comparison. Complete agreement for all quality ratings was attained following initial clarification and discussion of ratings, with no need for third author adjudication.

2.3.6 Categories of cognitive function

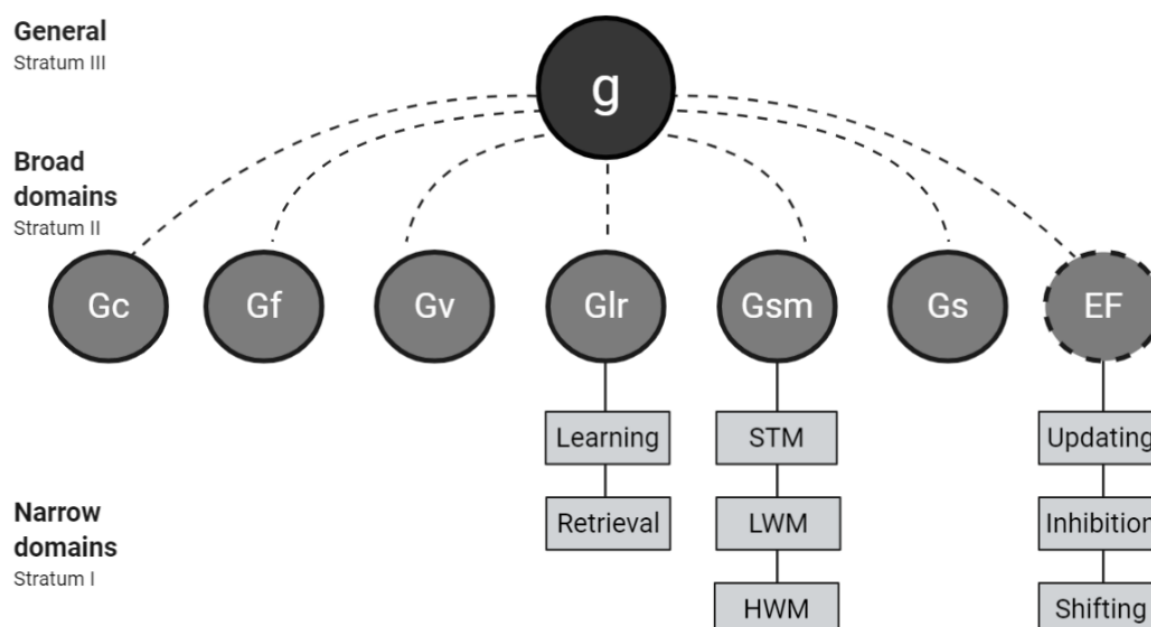


Figure 2—1. Hierarchical representation of the CHC-M composite model (adapted from Webb et al., 2018). *g* = general intelligence, *Gf* = fluid intelligence, *Gc* = crystallised intelligence, *Gv* = visual processing, *Glr* = long term storage and retrieval, *Gsm* = general short-term memory, *Gs* = processing speed, and *EF* = executive functions, *STM* = Short-term memory, *LWM* = Low working memory, *HWM* = High working memory. *EF* features a dotted outline to indicate its absence in the original CHC framework and subsequent addition by Webb et al., 2018 to create CHC-M.

The CHC-M cognitive taxonomy (Table 2.1) is based on a framework presented in Webb et al. (2018). CHC-M integrates both the Cattell-Horn-Carroll (CHC; Jewsbury et al., 2016) and Unity-Diversity (Friedman and Miyake, 2017) conceptualisations of executive function. CHC theory originates from psychometric research, providing an extensively studied theoretical framework for conceptualising human cognitive abilities (Carroll, 2010). The taxonomy presents a “three-stratum” taxonomical hierarchy of cognitive abilities (Figure 2.1). Each unique factor is identified as being distinct based on theoretically different processes and statistically divergent correlations (Webb et al., 2018). Atop the hierarchy in Stratum I is Charles Spearman’s general intelligence (*g*). Stratum II contains broad cognitive factors, including fluid intelligence (*Gf*), crystallised

intelligence (Gc), visual processing (Gv), long-term storage and retrieval (Glr), general short-term memory (Gsm), and processing speed (Gs). Stratum III contains narrow cognitive factors that constitute Stratum II factors, for example, abstract reasoning and verbal reasoning constitute Gf. Although the original CHC model does not feature EF, the Unity-Diversity conceptualisation of executive function was incorporated by Webb et al. (2018). In this conceptualisation, EF is comprised of the narrow factors of Updating, Inhibition, and Shifting. Unity-Diversity is considered synergistic with CHC as it uses the same psychometric criteria applied in the original CHC model (i.e. theoretically different processes and statistically divergent correlations). The construct of executive functioning is of growing clinical interest, with related deficits being studied in various clinical populations such as depression, schizophrenia, and attention deficit hyperactivity disorder (Webb et al., 2018).

Table 2—1. Descriptive overview of the broad and narrow cognitive domains constituting the CHC-M taxonomy. Adapted from Webb et al. (2018). Executive function definition from Friedman and Miyake (2017).

Broad cognitive domain		Narrow cognitive domain	
<i>Fluid reasoning (Gf)</i>	Deliberate but flexible control of attention to solve novel ‘on the spot’ problems that cannot be performed by relying exclusively on previously learned habits, schemas, and scripts.	<i>Abstract reasoning</i>	Ability to recognise (induct) and apply logical rules that govern sequence changes in abstract stimuli. Induction (I) and Sequential reasoning (RG) are defining features of Gf tasks.
		<i>Verbal reasoning</i>	Ability to understand and evaluate the logic of various verbal arguments; Reasoning with verbal material and knowledge acquired previously.
<i>Long-Term Memory and Retrieval (Glr)</i>	The ability to store, consolidate, and retrieve information over periods of time measured in minutes, hours, days, and years.	<i>Learning/Encoding Efficiency (Gl)</i>	Ability for efficient learning of new information that is held over longer periods than what is typical of short-term memory (Gsm) tasks.
		<i>Retrieval Fluency (Gr)</i>	The ability to rapidly access and recall information previously stored in long-term memory.

General Short-Term Memory (Gsm)	The ability to encode, maintain, and manipulate information in one's immediate awareness and includes memory span and working-memory. The working memory literature distinguishes between single-task coordination (e.g., manipulation of one-stream of stimuli in memory for retrieval) and multi-task coordination (e.g., of two or more streams of information).	High Working Memory (HWM)	Multi-task coordination: Manipulation/processing of multiple streams of information for a coordinated response. Can entail inhibition of one-or more streams. Entails both processing and storage of multiple sources of information.
		Low Working Memory (LWM)	Entails a manipulation that is more cognitively complex than STM processes, but less complex than updating and High WM. Entails both processing and storage of a single (rather than multiple) source of information (e.g., recall a list in some specified order other than how it was presented (e.g., in reverse order).
		Short-term Memory (STM)	The ability to encode information, maintain it in immediate awareness (e.g., in primary memory), and immediately reproduce the information in the same sequence which it was presented.
Executive Function (EF)*	High-level cognitive processes, often associated with the frontal lobes, that control lower level processes in the service of goal-directed behaviour	Updating	The active process of monitoring incoming information and 'updating' items held in working memory by replacing irrelevant information with task-relevant information.
		Shifting	Shifting between mental sets or operations by disengaging from an irrelevant mental set and actively engaging in a set relevant to the current task.
		Inhibition	Active and deliberate overriding of a dominant or automatic response in order to complete a task.
Processing speed (Gs)	The ability to perform simple repetitive cognitive tasks quickly and fluently.	Perceptual speed	The speed at which visual stimuli can be compared for similarity or difference

Visual processing (Gv)	The ability to make use of simulated mental imagery (often in conjunction with currently perceived images) to solve problems.	Sensory perception	The efficiency of primary senses to process and provide information, picked up through primary senses (typically vision), necessary for task completion.
		Visual perception	The ability to perceive complex patterns and mentally simulate how they might look when transformed (e.g., rotated, changed in size, partially obscured). It is central to Gv (like induction is central to Gf)

Supplementary material from Webb et al. (2018) provides a categorisation of cognitive tasks subsequently used to categorise cognitive tasks in this analysis. When a cognitive task did not feature in the Webb et al. (2018) categorisation, two members of the review team independently categorised the task according the cognitive functions defined in the CHC-M taxonomy. Independent categorisations were compared and consensus was achieved through discussion for instances where there was disagreement between categorisations. Within each cognitive construct, outcome effect sizes were further partitioned into task performance assessing response times and accuracy rates.

In the published protocol (Clarke et al., 2018), a distinction was made between objective and subjective cognition (i.e. tasks that aim to measure a specific cognitive process versus self-report of perceived cognitive performance). However, an unanticipated categorisation issue arose in the form of clinical tests specifically designed to screen for mild cognitive impairment (MCI), such as the Montreal Cognitive Assessment or MoCA (Nasreddine et al., 2005). As MCI screening tests are designed to assess overall cognitive ability and not performance in specific cognitive domains, we categorised such tasks as 'MCI screening tests'. A separate meta-analysis was conducted for this category.

2.3.7 Statistical analyses

2.3.7.1 Calculation of effect sizes

The effect size of interest was the product-moment correlation (r). For records where measures of association between cognitive tasks and tinnitus severity were not reported,

an estimate of the effect size was calculated from available descriptive statistics. The variety of study designs included in the review meant that available data necessary to calculate estimates of effect sizes differed across records. Estimates of the correlation coefficient between tinnitus and cognitive performance were derived from different descriptive statistics, dependant on what data were available. The correlation coefficient was chosen for its flexibility as an effect size metric (McGrath and Meyer, 2006), allowing for a less selective data synthesis strategy. Ellis (2010, p. 99) recommends reporting correlation over a standardised mean difference (SMD) because of inherent advantages, noting that conversion of a correlation into an SMD usually entails a loss of information. To calculate effect sizes for inclusion in this review, the following schema was developed:

- When a record reported a single Pearson correlation per measure of cognitive performance the correlation was used, and the data item was coded as "reported".
- When a record reported multiple tinnitus measurements and multiple Pearson correlations per measure of cognitive performance, then the correlation between the primary tinnitus questionnaire and measure of cognitive performance was used, and the data item was coded as "reported".
- When a record did not report a Pearson correlation between tinnitus questionnaires and cognitive performance, then the data item was coded as 'not reported'.

A Pearson correlation was calculated from available descriptive and/or test statistics using an effect-size calculator

(<http://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-Home.php>). If a record did not report sufficient descriptive statistics to derive a Pearson correlation, then a standardised mean difference was calculated (Lipsey and Wilson, 2000, p. 198). SMDs were converted to point-biserial correlations using the formula (Polanin and Snilstveit, 2016):

$$r = \frac{d}{\sqrt{d^2 + a}}$$

where r is the correlation coefficient, d is the SMD, and a is given by

$$a = \frac{(n_1 + n_2)^2}{n_1 * n_2}$$

where n is the size of each group.

For some records, relevant statistics could be estimated from figures. Estimation was performed using WebPlotDigitizer (<https://automeris.io/WebPlotDigitizer/>). Estimated values were then used to calculate an effect size using one of the aforementioned steps. An overview audit of the method of derivation of each record effect size is provided as supplementary material, as well as details of records that were suitable for inclusion but did not contain enough information to calculate an effect size (https://osf.io/gk5zh/?view_only=d2fa693eced945ea9e12c5dcce2d3957). In those instances, study authors were contacted with a request to provide a correlation from their original dataset, descriptive statistics, or points of clarification that would facilitate calculation of a correlation.

2.3.7.2 *Meta-analyses*

Meta-analyses were performed in R (R Core Team, 2018). Meta-analyses were performed using the Metafor package (Viechtbauer, 2010). Random-effects models were fitted to data from broad cognitive domains (i.e. EF, Gs, Gsm, Glr, and Gv), and narrow cognitive domains where available data permitted. Due to the heterogeneous nature of studies included in this analysis (i.e. studies produced from different laboratories, using different methods, and different populations), a random-effects model was chosen and specified prior to analysis (Clarke et al., 2018). As some models featured multiple effect sizes from a single study, the within-subjects statistical dependency of these effect sizes was accounted for using robust variance estimation (RVE; Hunter and Schmidt, 2004; Moeyaert et al., 2017; Quintana, 2015).

Model heterogeneity was evaluated using Q and I^2 statistics. The Q-statistic tests the null hypothesis that all studies are evaluating the same effect. The I^2 statistic ranges between 0-100% interpreted as to low (25%), moderate (50%), and high (75%) heterogeneity (Higgins et al., 2003; Quintana, 2015). Identification of potential outlying cases were performed in accordance with guidelines provided in Viechtbauer and Cheung (2010). Outlier diagnostics of the various meta-analytic models included: Externally standardised residuals, difference of fits (DFFITS), Cook's distances, Covariance ratios, difference of betas values (DFBETAS), Estimates of R^2 when each study is removed in turn, Test statistics for (residual) heterogeneity when each study is removed in turn, Diagonal elements of the hat matrix, and Weights (in percent) given to the observed outcomes during the model fitting. Baujat plots were used to visualise overly influential data points within models (Baujat et al., 2002; Quintana, 2015). A 'leave-one-out' analysis was also performed on each model, which repeatedly fits the model, excluding a single study on each iteration, to verify that model results are not being driven by a single overly-influential study.

Moderator analyses were performed through meta-regression models. Moderator variables were specified *a priori* and coded during the data extraction stage and included: mean age of tinnitus participant sample, yes/no measurement of anxiety and/or depression, and hearing thresholds. Publication bias was assessed visually using contour-enhanced funnel plots, with statistical significance being tested using Rank correlation and Egger's regression tests (Quintana, 2015). If a record could not be included within the meta-analyses, it was considered as part of the narrative synthesis.

2.4 Results

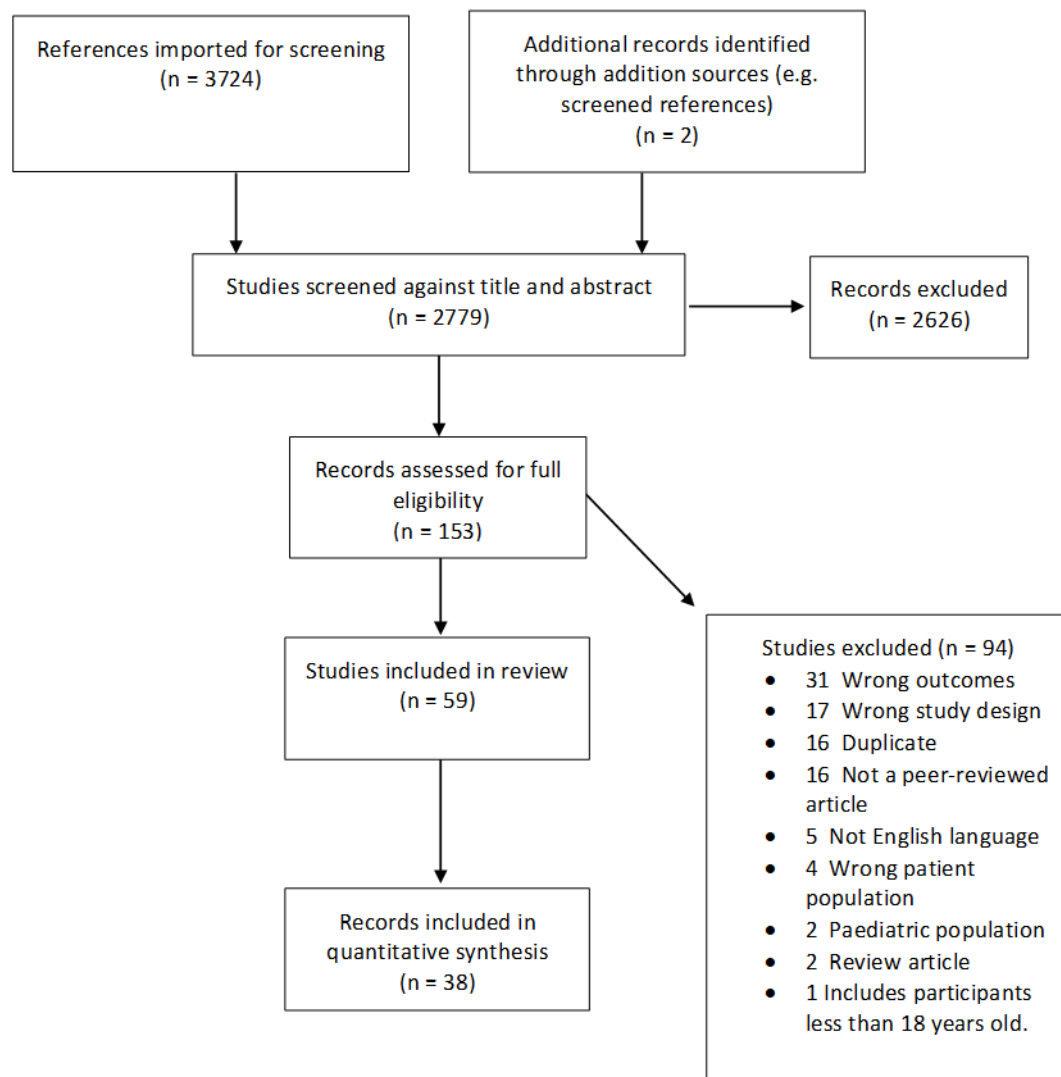


Figure 2—2. PRISMA flowchart providing an overview of the systematic search and screening process.

2.4.1 Included study overview

Thirty-eight records (1,863 participants) reported correlations or featured calculable correlations, and were therefore included in the analyses. 93 correlations were calculated from included records. Table 2.2 details the count of effect sizes across broad cognitive factors.

Table 2—2. Count of effect sizes in each broad cognitive factor.

Broad cognitive factors	Effect size total
EF	35
Gs	19
Glr	16
Cognitive screening	9
Gsm	8
Self-report	4
Gv	2

2.4.2 Weight of Evidence (WoE) - record bias and quality appraisal

The overall quality of studies varied widely. Assessment of the quality of each record in the context of its ability to answer the main review question was performed with WoE appraisal (Appendix A). WoE scores were subsequently used as a variable within moderator analyses of the broad cognitive domain models.

2.4.3 Model results

An overview of the model results is provided in Table 2.3. The following sections detail analysis of each cognitive domain and the models that were applied.

Table 2—3. Overview of meta-analytic models, providing correlation, statistical significance, and number of effect sizes contained in its calculation.

Model	r	p-value	Count
EF-rt	0.35	< 0.001	24
EF-errors	0.3	< 0.001	5
EF-correct	0.1	n.s.	6
EF-Inhibition-rt	0.32	< 0.001	14
EF-Inhibition-errors	0.24	n.s.	2
EF-Shifting-rt	0.48	< 0.001	8
EF-Shifting-errors	0.37	0.003	3
EF-Updating-rt	-0.14	n.s.	2
Gs-rt	0.34	< 0.001	13
Gs-errors	0.27	0.002	3
Gs-correct	-0.2	n.s.	3
Gsm-correct	-0.21	0.024	7
Gsm-STM-correct	-0.11	n.s.	3
Gsm-LWM-correct	-0.31	n.s.	2

Gsm-HWM-correct	-0.28	n.s.	2
Glr-correct	-0.16	0.003	16
Glr-GL-correct	-0.14	0.007	10
Glr-GR-correct	-0.19	0.024	6
Gv-rt	0.3	n.s.	2
Self-report	0.51	0.04	4
Cognitive screening	-0.47	0.002	9

2.4.3.1 Executive function (EF)

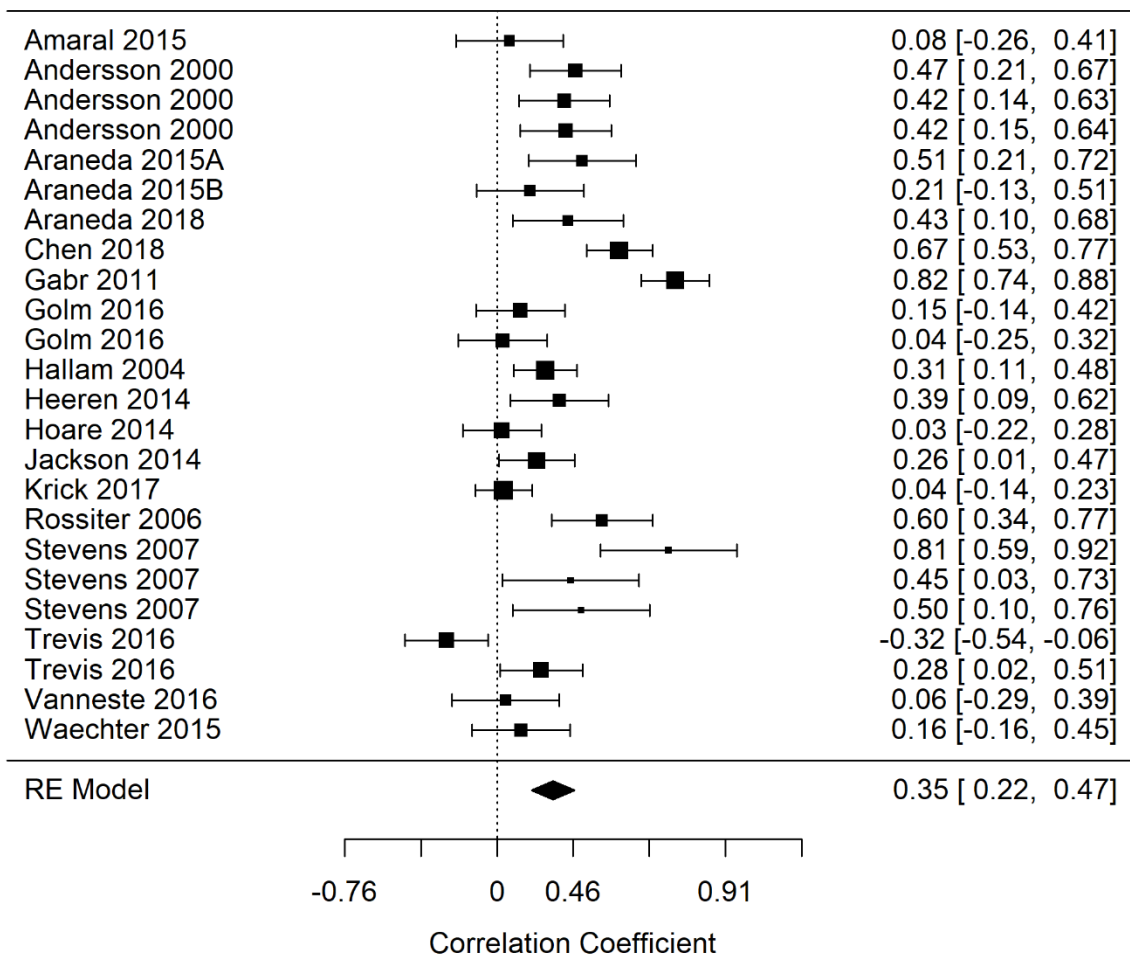


Figure 2—3. Forest plot showing the correlation between tinnitus and response times in Executive Functioning tasks (EF-rt). Each study is represented by a point estimate bounded by a 95% CI, with the area of each square proportional the study's weight within the model. Summary effect size is displayed as a polygon at the bottom of the plot, with the width of the polygon representing the 95% CI.

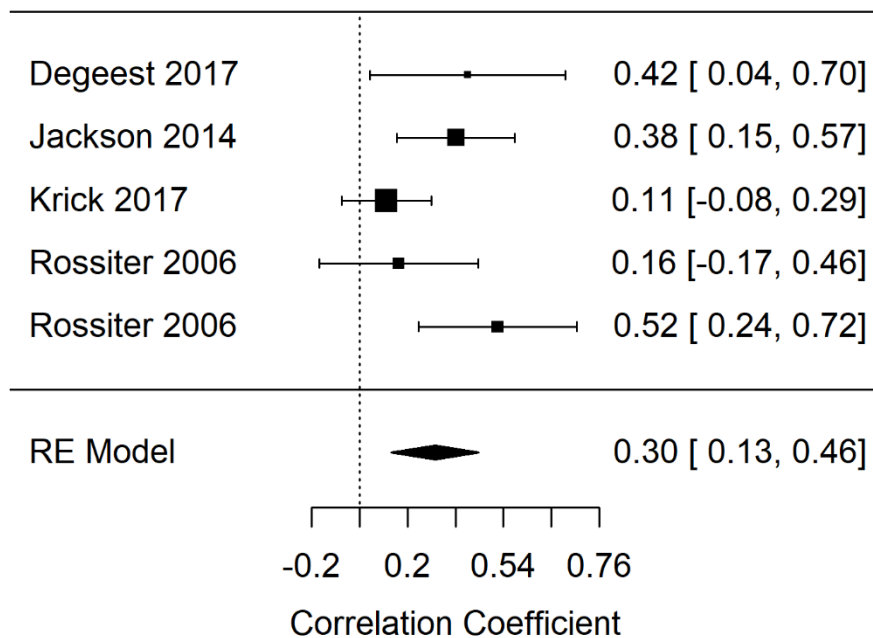


Figure 2—4. Forest plot showing the correlation between tinnitus and error rates in Executive Functioning tasks (EF-error). Each study is represented by a point estimate bounded by a 95% CI, with the area of each square proportional the study's weight within the model. Summary effect size is displayed as a polygon at the bottom of the plot, with the width of the polygon representing the 95% CI.

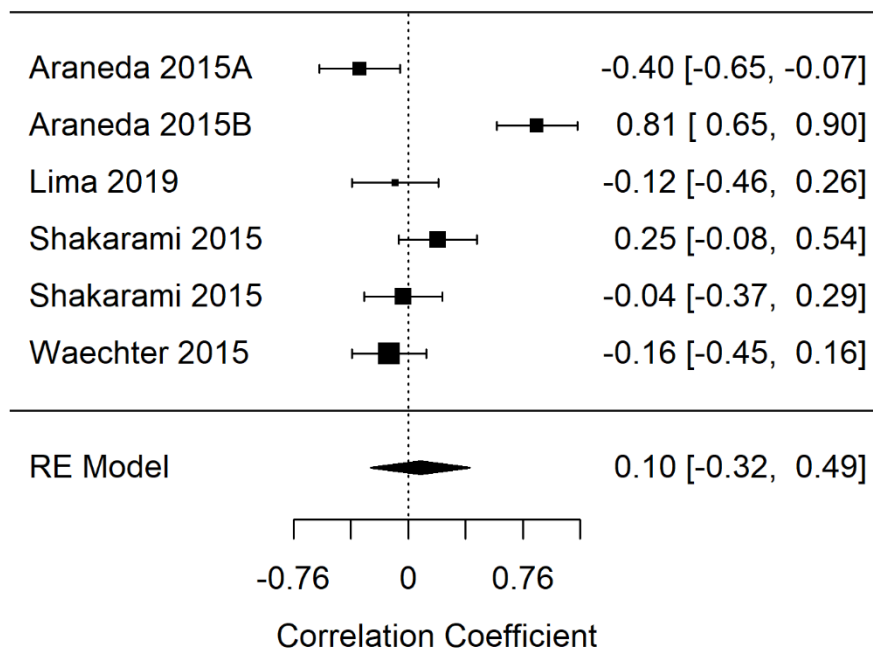


Figure 2—5. Forest plot showing the correlation between tinnitus and correct responses in Executive Functioning tasks (EF-correct). Each study is represented by a point estimate bounded by a 95% CI, with the area of each square proportional the study's weight within the model. Summary effect size is displayed as a polygon at the bottom of the plot, with the width of the polygon representing the 95% CI.

Random-effects meta-analyses for EF (broad domain) response time (EF-rt), errors (EF-errors), and correct responses (EF-correct) were performed. Figures 2.3, 2.4, and 2.5 show model summary correlations. Statistically significant summary correlations were identified in models for EF-rt [$r = 0.35$; 95% CI (0.22, 0.47), $p < 0.001$] and EF-errors [$r = 0.3$; 95% CI (0.13, 0.46), $p < 0.001$]. These results demonstrate longer response times and increased error rates in executive functioning tasks as tinnitus severity increased.

In a leave-one-out analysis, EF-rt remained statistically significant in every iteration, indicating that the result was not being driven by the effect of a single study. However, EF-errors was not statistically significant following the removal of Jackson et

al. (2014), indicating that this study was solely responsible for driving the statistical significance of the original model. Analysis of EF-rt using RVE supported the statistically significant finding [0.37; 95% CI (0.19, 0.54), $p < 0.001$].

Both models featured significant Q statistics, indicating that they do not share a common effect size. EF-rt [$Q(df = 23) = 143$, $p < 0.001$], EF-errors [$Q(df = 3) = 7.53$, $p = 0.06$]. High and moderate heterogeneity was evident for both response time and error models. EF-rt [$I^2 = 81.98$] and EF-error [$I^2 = 59.35$].

EF-correct was not statistically significant. Furthermore, as there were only six studies with the necessary outcome measures reported, there were too few data points to further investigate narrow executive function domains with correct response measures.

Figure 2.6 shows a contour-enhanced funnel plot of EF-rt that suggests an absence publication bias. Egger's regression and rank correlation tests were not significant, proving further support for a lack of publication bias within executive function studies. Similarly, there was evidence for an absence of publication bias in all broad domain models, which are subsequently omitted for brevity.

Various moderator variables were tested in the EF-rt model to explain observed heterogeneity. These included presence/absence of hearing measurement, presence/absence of depression measurement, presence/absence of anxiety measurement, risk of bias quality score, and mean hearing threshold of tinnitus participants included in the record. No moderator variable was statistically significant within the broad EF domain. The number of broad EF studies available meant

examination of executive function narrow factors was also possible.

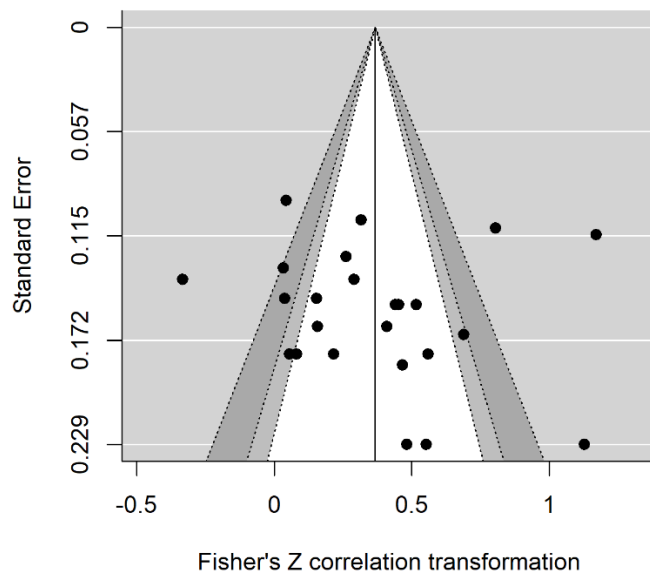


Figure 2—6. Contour-enhanced funnel plot for EF-rt model. Individual study effect sizes are displayed on the x-axis, with their corresponding standard errors on the y-axis. The funnel lines are centred on the summary effect size, providing an indication of the spread. The unshaded region in the middle corresponds to p -values greater than .10, the light grey shaded region corresponds to p -values between .10 and .05, the dark grey region corresponds to p -values between .05 and .01, and the region outside of the funnel corresponds to p -values below .01.

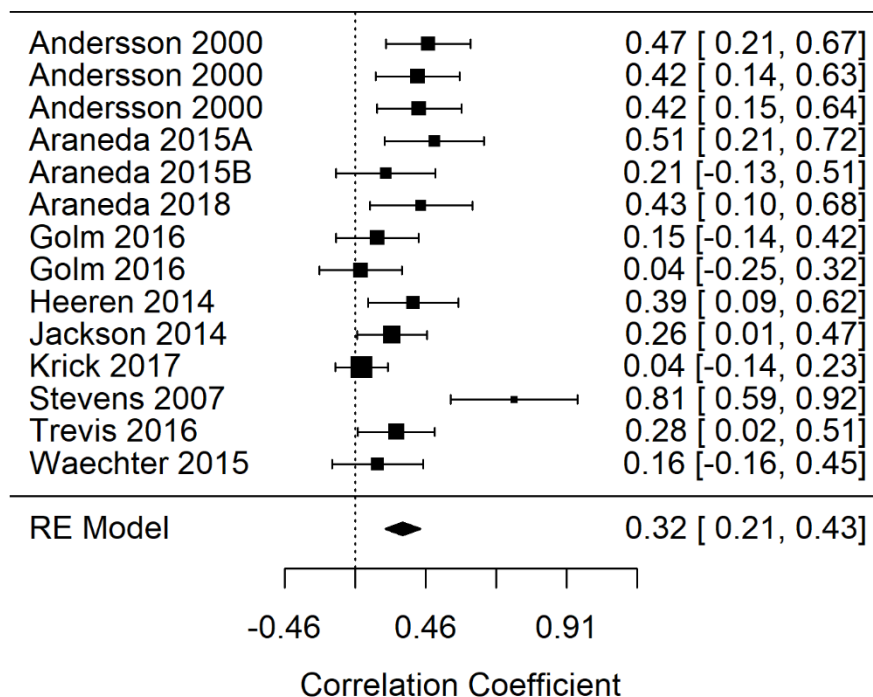


Figure 2—7. Forest plot showing the correlation between tinnitus and response times in Inhibition tasks (Inhibition-rt). Each study is represented by a point estimate bounded by a 95% CI, with the area of each square proportional the study's weight within the model. Summary effect size is displayed as a polygon at the bottom of the plot, with the width of the polygon representing the 95% CI.

Random-effects meta-analyses for Inhibition response time (Inhibition-rt) and errors (Inhibition-errors) were performed. A statistically significant summary correlation was identified for Inhibition-rt [$r = 0.32$; 95% CI (0.21, 0.43), $p < 0.001$]. Figure 2.7 shows the Inhibition-rt summary correlation. Inhibition-error was not statistically significant. These results demonstrate longer response times for tasks measuring Inhibition as tinnitus severity increases. Results for this narrow factor include only three studies. RVE for Inhibition-rt supported the statistically significant finding with a significant point estimate [0.34; 95% CI (0.16, 0.51), $p = 0.002$]. Heterogeneity was reduced in the

Inhibition-rt model, indicating that the cognitive sub-grouping had reduced heterogeneity between studies from a large to a moderate amount [$I^2 = 58.94$].

A Baujat plot and outlier diagnostics indicated that Stevens et al. (2007) may have been a potential outlier, having an overly influential effect on the overall result. A random-effects model was run with the removal of this study. The correlation remained statistically significant, but was reduced from $r = 0.32$ to $r = 0.28$. However, removal of this study significantly reduced heterogeneity of the model from a large to moderate amount [I^2 reduced from 58.94 to 37.86]. Combined with a non-significant Q statistic, this may indicate that the remaining studies share a common effect size.

2.4.3.3 EF - Shifting

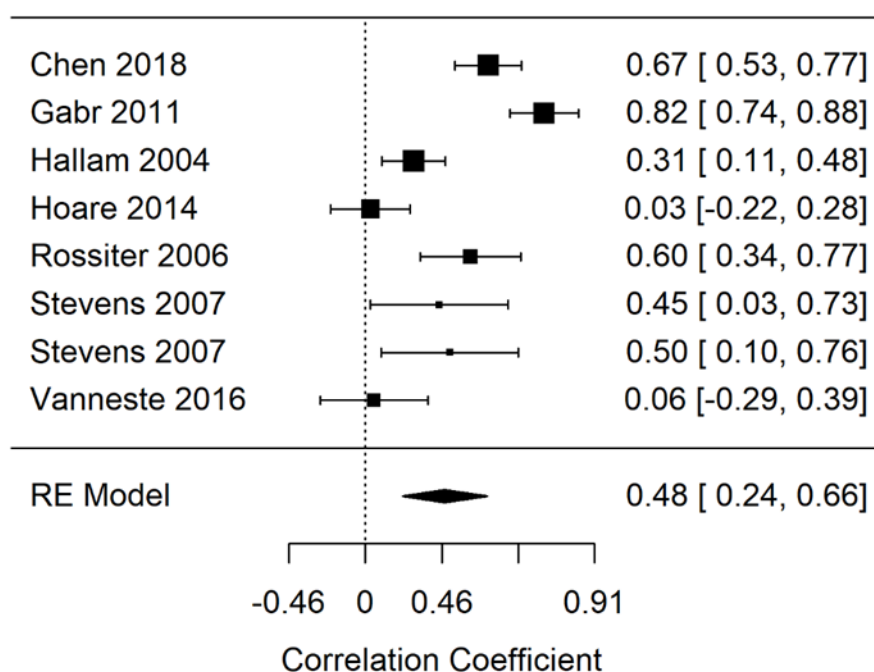


Figure 2—8. Forest plot showing the correlation between tinnitus and response times in Shifting tasks (Shifting-rt). Each study is represented by a point estimate bounded by a 95% CI, with the area of each square proportional the study's weight within the model. Summary effect size is displayed as a polygon at the bottom of the plot, with the width of the polygon representing the 95% CI.

Random-effects meta-analyses for Shifting response time (Shifting-rt) and errors (Shifting-errors) were performed. A statistically significant summary correlation was identified for Shifting-rt [$r = 0.48$; 95% CI (0.24, 0.66), $p < 0.001$] and Shifting-errors [$r = 0.37$; 95% CI (0.13, 0.57), $p = 0.003$]. Figure 2.8 shows the Shifting-rt summary correlation. RVE for Shifting-rt supported the statistically significant finding with a significant point estimate [0.52; 95% CI (0.17, 0.86), $p = 0.011$]. Response times and error rates for Shifting tasks increased as a function of tinnitus severity.

A Baujat and outlier diagnostics indicated that Gabr et al. (2011) may have been an outlier. The summary correlation remained statistically significant after the removal of Gabr et al. (2011) but with a smaller point estimate [$r = 0.39$; 95% CI (0.18, 0.57), $p < 0.001$]. Although removal of this study reduced heterogeneity of the model [$I^2 =$ reduced from 87 to 76.86], a large amount of heterogeneity remained across included studies.

2.4.3.4 EF - Updating

Only two studies included Updating effect sizes, using response time outcomes (Updating-rt). The summary correlation for Updating-rt was not statistically significant.

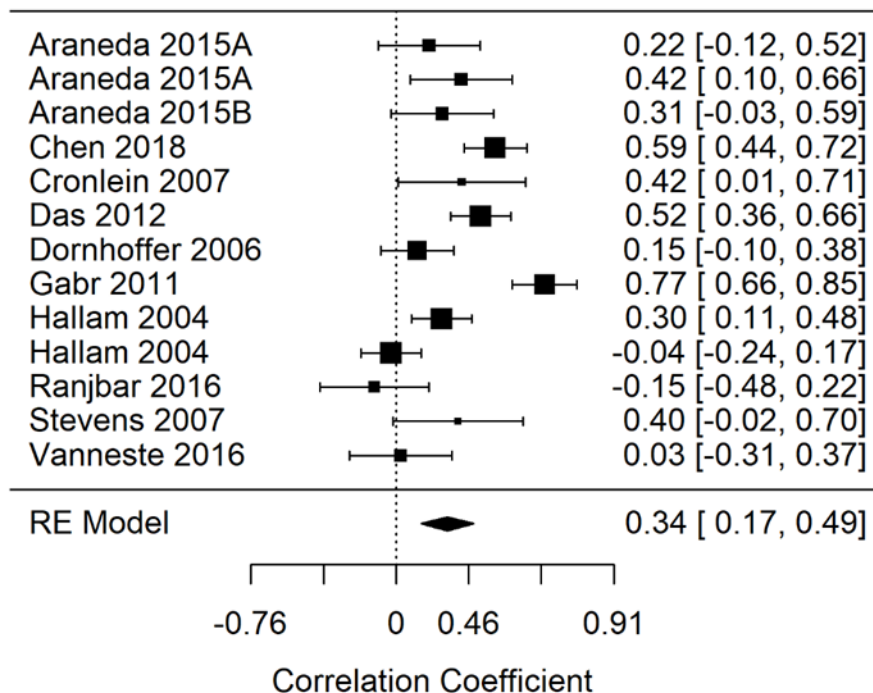


Figure 2—9. Forest plot showing the correlation between tinnitus and response times in processing speed tasks (Gs-rt). Each study is represented by a point estimate bounded by a 95% CI, with the area of each square proportional the study's weight within the model. Summary effect size is displayed as a polygon at the bottom of the plot, with the width of the polygon representing the 95% CI.

Meta-analyses were performed on Gs response time (Gs-rt), errors (Gs-errors), and correct (Gs-correct). Figure 2.9 shows the Gs-rt summary correlation. Statistically significant summary correlations were identified in models for Gs-rt [$r = 0.34$; 95% CI (0.17, 0.49), $p < 0.001$] and Gs-errors [$r = 0.27$; 95% CI (0.1, 0.43), $p = 0.002$]. These results demonstrate longer response times and increased error rates in processing speed tasks as tinnitus severity increased.

The summary correlation for Gs-correct was not statistically significant. RVE for Gs-rt supported the statistically significant finding with a significant point estimate [0.35; 95% CI (0.14, 0.56), $p = 0.004$]. RVE could not be performed for Gs-errors due

to lack of available degrees of freedom. A leave-one-out analysis for Gs-rt, which remained statistically significant in every iteration, indicating that the result was not being driven by the effect of a single study.

A Baujat plot indicated that Gabr et al. (2011) may be overly influential on the Gs-rt model. Gs-rt was rerun with the removal of this study. The summary correlation remained statistically significant but with a smaller point estimate [$r = 0.28$; 95% CI (0.14, 0.42), $p < 0.001$]. RVE for Gs-rt supported the statistically significant finding with a significant point estimate [0.29; 95% CI (0.12, 0.46), $p = 0.005$]. Although removal of this study significantly reduced heterogeneity of the model [I^2 reduced from 81.6 to 69.72], moderate heterogeneity remained across included studies.

Various moderator variables were tested in the Gs-rt model to explain observed heterogeneity. These included presence/absence of hearing measurement, presence/absence of depression measurement, presence/absence of anxiety measurement, risk of bias quality score, and mean hearing threshold of tinnitus participants included in the record. The model including presence/absence of depression measurement produced a statically significant moderator coefficient [$p = 0.037$]. This suggests that a proportion of the slower response times observed on processing speed tasks were confounded by inclusion of participants with depression, with I^2 being reduced from 81.63 to 75.12.

2.4.3.6 General short-term memory (Gsm)

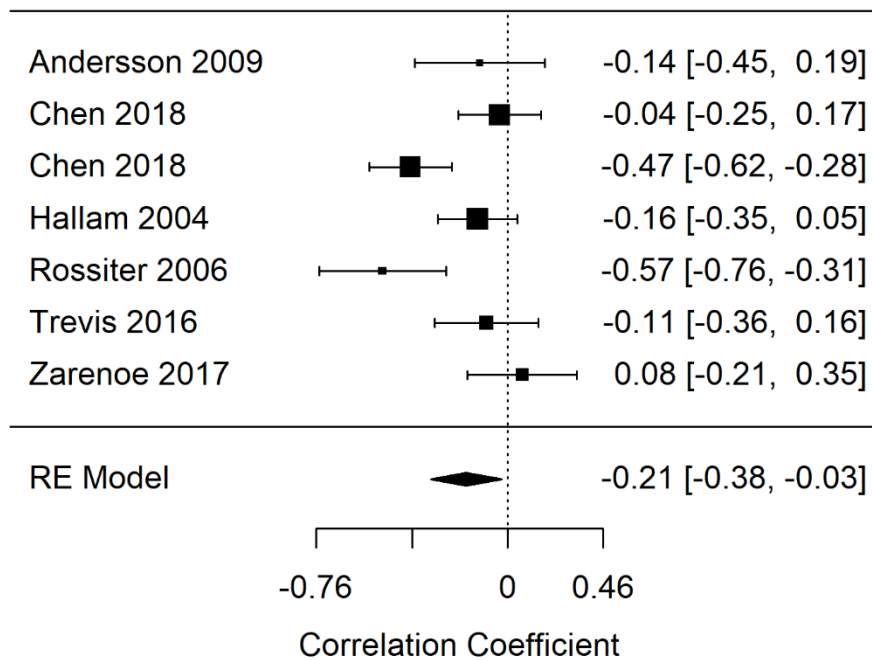


Figure 2—10. Forest plot showing the correlation between tinnitus and correct responses in General Short-term Memory tasks (Gsm-correct). Each study is represented by a point estimate bounded by a 95% CI, with the area of each square proportional the study's weight within the model. Summary effect size is displayed as a polygon at the bottom of the plot, with the width of the polygon representing the 95% CI.

Gudwani et al. (2017) reported errors as an outcome measure; it was therefore excluded from the analysis as the remaining studies reported number of correct responses. A random-effects meta-analysis was performed on the remaining studies (Gsm-correct). Figure 2.10 shows the Gsm-correct summary correlation. A statistically significant summary correlation was calculated for Gsm-correct [$r = -0.21$; 95% CI (-0.38, -0.03), $p = 0.024$]. The results demonstrate less correct answers on general short-term memory tasks as tinnitus severity increased. RVE supported the finding with a significant point estimate [-0.22; 95% CI (-0.43, -0.001), $p = 0.049$]. Moderate heterogeneity was present in the Gsm-correct model [$I^2 = 72.67$].

In a leave-one-out analysis, statistical significance was not reached in two iterations, indicating that the result was being driven by the effects of two influential studies. Regression diagnostics indicated no evidence of outliers in Gsm-correct.

A model including presence/absence of hearing measurement as a moderating variable produced a statistically significant moderator coefficient [$p = 0.047$]. This suggests that the fewer correct responses observed in general short-term memory tasks were confounded by inclusion of participants with hearing loss. I^2 in the models with hearing impairment reduced from 72.69 to 62.71.

2.4.3.7 Short-term memory (STM), Low working memory (LWM), and High working memory (HWM)

Random-effects meta-analyses were conducted on Gsm narrow factors for short-term memory (Gsm-STM), low working memory (Gsm-LWM), and high working memory (Gsm-HWM). None of the narrow factor models were statistically significant.

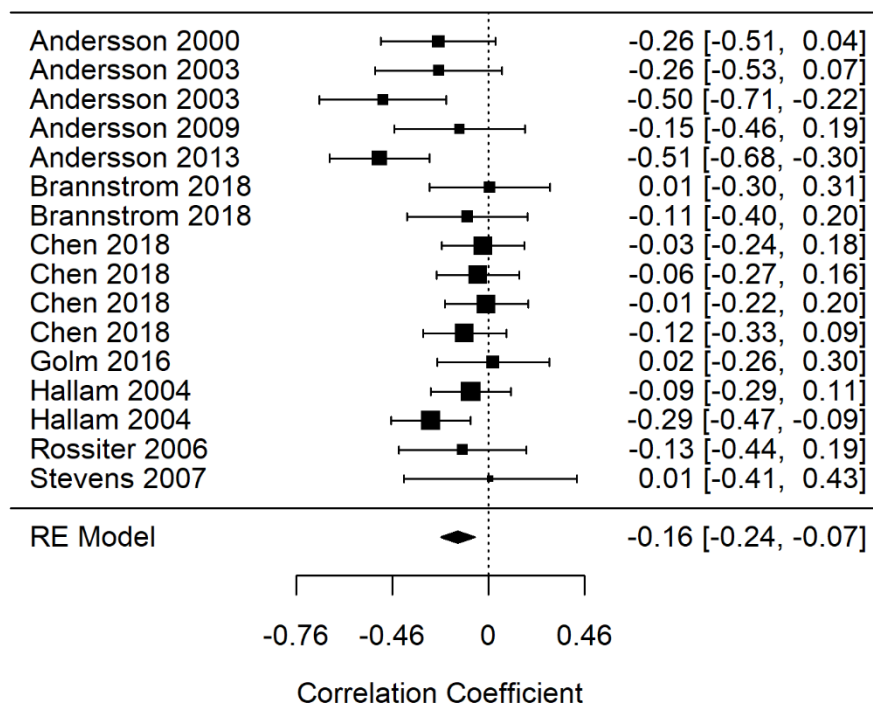


Figure 2—11. Forest plot showing the correlation between tinnitus and correct responses in General Learning and Retrieval tasks (Glr-correct). Each study is represented by a point estimate bounded by a 95% CI, with the area of each square proportional the study's weight within the model. Summary effect size is displayed as a polygon at the bottom of the plot, with the width of the polygon representing the 95% CI.

A random effects meta-analysis was performed for Glr (Glr-correct). Figure 2.11 shows the Glr-correct summary correlation. A statistically significant summary correlation was calculated for Glr-correct [$r = -0.16$; 95% CI $(-0.24, -0.07)$, $p < 0.001$]. The results demonstrate less correct answers on general learning and retrieval tasks as tinnitus severity increased. RVE supported the finding with a significant point estimate $[-0.16$; 95% CI $(-0.31, -0.01)$, $p = 0.04$]. Moderate heterogeneity was present in the Glr-correct model [$I^2 = 41.31$]. A leave-one-out analysis for Glr-correct remained statistically significant in every iteration, indicating that the result was not driven by the effect of a single study.

Moderator analysis for Glr-correct found that no variable accounted for a statistically significant amount of heterogeneity. Random-effects meta-analyses were ran on Glr narrow factors for learning (Glr-learning) and retrieval (Glr-retrieval).

2.4.3.9 General learning (Gl)

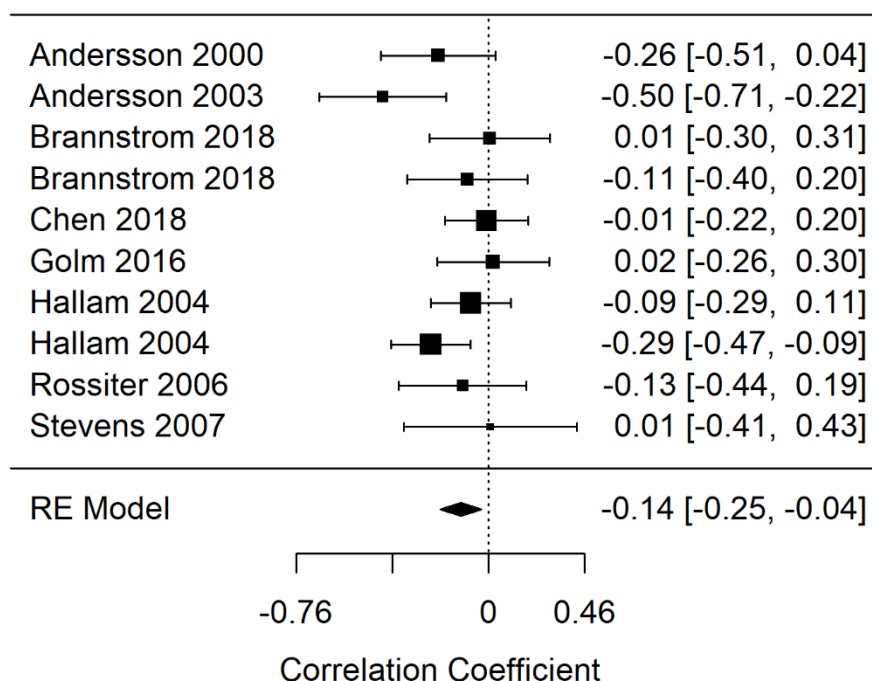


Figure 2—12. Forest plot showing the correlation between tinnitus and correct responses in General Learning tasks (Glr-GL-correct). Each study is represented by a point estimate bounded by a 95% CI, with the area of each square proportional the study's weight within the model. Summary effect size is displayed as a polygon at the bottom of the plot, with the width of the polygon representing the 95% CI.

Figure 2.12 shows the Glr-GL-correct summary correlation. The summary correlation for Glr-learning was statistically significant [$r = -0.14$; 95% CI (-0.25, -0.04), $p = 0.03$]. The results demonstrate less correct answers when performing memory retrieval tasks as tinnitus severity increased. Moderate heterogeneity was present in the Glr-learning model [$I^2 = 29.2$].

2.4.3.10 General retrieval (Gr)

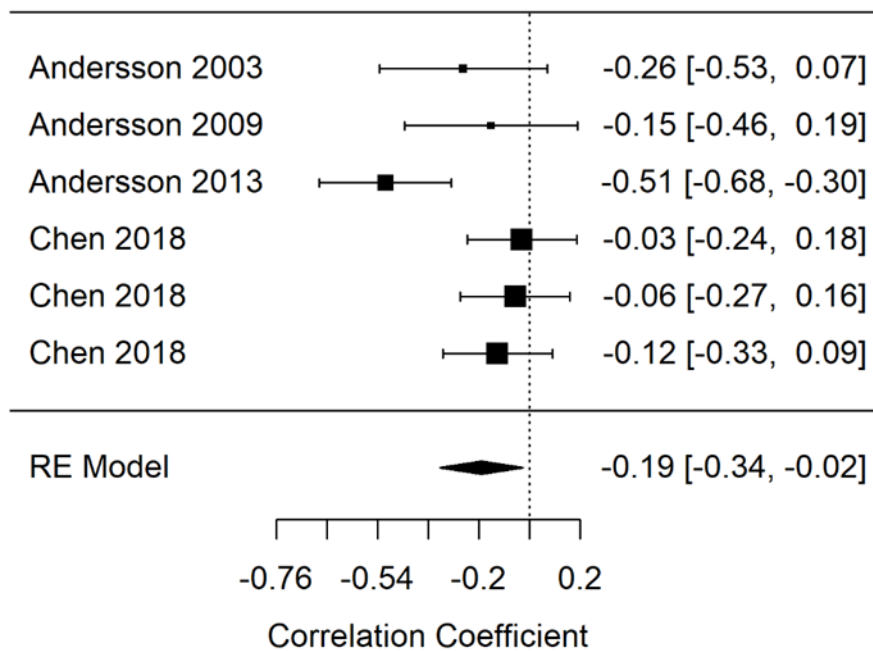


Figure 2—13. Forest plot showing the correlation between tinnitus and correct responses in General Retrieval tasks (Glr-GR-correct). Each study is represented by a point estimate bounded by a 95% CI, with the area of each square proportional the study's weight within the model. Summary effect size is displayed as a polygon at the bottom of the plot, with the width of the polygon representing the 95% CI.

Figure 2.13 shows the Glr-GR-correct summary correlation. The summary correlation for Glr-retrieval was statistically significant [$r = -0.19$; 95% CI (-0.34, -0.03), $p = 0.024$]. The results demonstrate less correct answers when performing memory retrieval tasks as tinnitus severity increases. Moderate heterogeneity was present in the Glr-retrieval model [$I^2 = 59.84$].

2.4.3.11 Visual processing (Gv)

A random effects meta-analysis was performed for the Gv (Gv-rt). The summary correlation for Gv-rt was not statistically significant.

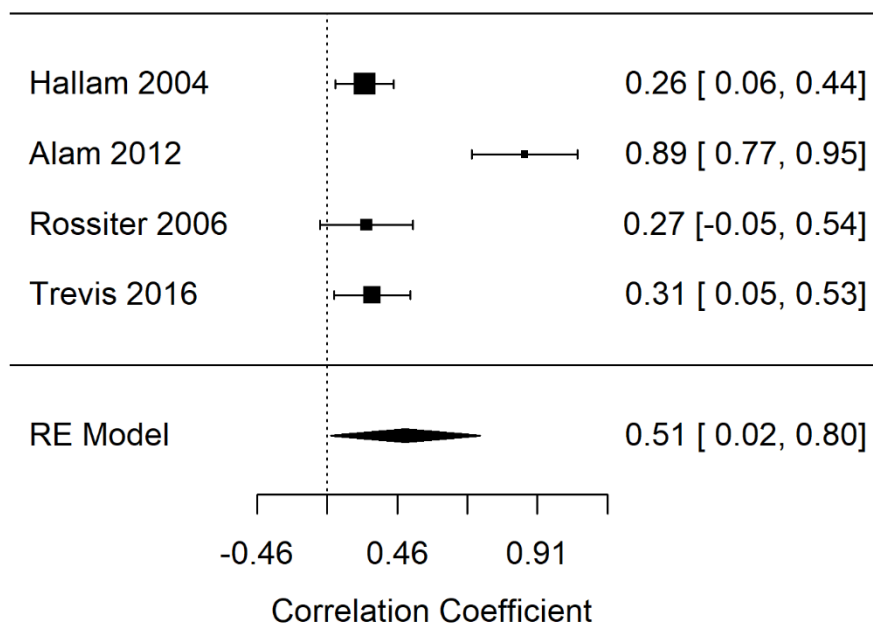


Figure 2—14. Forest plot showing the correlation between tinnitus and cognitive self-report measures (Self-report). Each study is represented by a point estimate bounded by a 95% CI, with the area of each square proportional the study's weight within the model. Summary effect size is displayed as a polygon at the bottom of the plot, with the width of the polygon representing the 95% CI.

Figure 2.14 shows the Self-report summary correlation. The summary correlation for Cognitive Self-Report was statistically significant [$r = 0.51$; 95% CI (0.03, 0.8), $p = 0.04$]. The results demonstrate poorer self-reported cognition as tinnitus severity increased. High heterogeneity was present in the Cognitive Self-Report model [$I^2 = 92.73$]. A leave-one-out analysis showed that the observed effect was likely driven by a single study. The effect size without Alam et al. (2012) showed a smaller, but still statistically significant point estimate [$r = 0.28$; 95% CI (0.14, 0.41), $p < 0.001$]. Moderator analysis showed that no single variable was significant for any Cognitive Self-Report model.

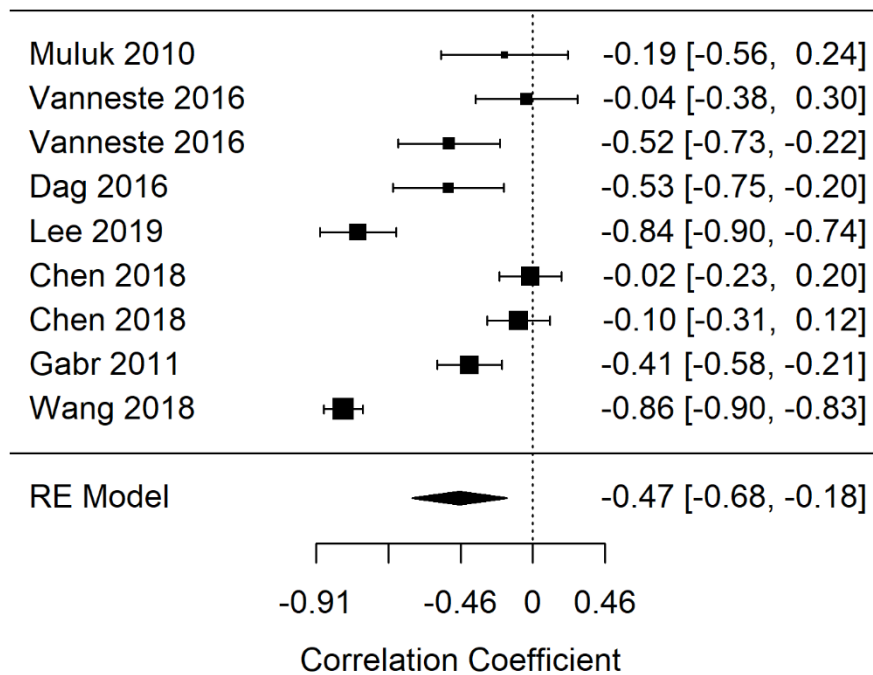


Figure 2—15. . Forest plot showing the correlation between tinnitus and cognitive screening measures. Each study is represented by a point estimate bounded by a 95% CI, with the area of each square proportional the study's weight within the model. Summary effect size is displayed as a polygon at the bottom of the plot, with the width of the polygon representing the 95% CI.

Figure 2.15 shows the MCI summary correlation. The summary correlation for MCI was statistically significant [$r = -0.47$; 95% CI (-0.68, -0.18), $p = 0.002$]. The results show poorer performance on cognitive screening tasks for participants with tinnitus. RVE supported the finding, with a statistically significant point estimate [-0.5; 95% CI (-0.97, -0.04), $p = 0.04$]. High heterogeneity was present in the Cognitive Screening model [$I^2 = 93.42$]. In a leave-one-out analysis of Cognitive Impairment Screening all models remained statistically significant, i.e. the effect was not driven by a single study.

2.5 Discussion: is subjective tinnitus associated with poorer cognitive performance?

This review provides a profile of cognitive performance associated with tinnitus through different types of outcome measures used in tasks measuring cognitive performance (i.e. response time, accuracy, and error rates) across various cognitive domains. Having tinnitus is associated with longer responses times and increased error rates in processing speed (Gs) and executive function (EF) tasks, and less correct answers on general short-term memory (Gsm) and general learning and retrieval (Glr) tasks. Being able to incorporate the different outcome measures used across the varying studies meant that this review was comprehensive in its scope, providing a broad synthesis of the evidence of tasks that have investigated any aspect of tinnitus and cognitive performance. In doing so, the analysis illuminates a general trend of significantly poorer cognitive performance in participants with tinnitus, as well as estimates of the relative size of this effect within each cognitive domain.

The pattern of associations outlined in this analysis provides a framework for theorising about various cognitive domains and their relationship to tinnitus. This is important as evidence suggests that cognitive training, particularly of executive functions, may improve aspects of cognition (Diamond and Ling, 2016); this could serve to remediate poorer cognitive performance, but its impact on perceived severity of the tinnitus is unknown. The following sections consider associations between tinnitus and specific broad and narrow cognitive factors within the CHC-M taxonomy (i.e. strata I and II).

2.5.1 Which broad cognitive domains are associated with tinnitus?

Subjective tinnitus is associated with poorer performance in tasks measuring executive function, general short-term memory, long-term storage and retrieval, and processing speed. There were no studies that featured tasks measuring the broad cognitive factors of fluid intelligence (Gf) and crystallised intelligence (Gc).

2.5.1.1 Executive Function (EF)

This review demonstrates consistently poorer executive function performance by individuals with tinnitus. This is evidenced through increased response times and errors rates in EF-rt and EF-error models. No significant moderating variables were identified within the domain of executive functioning. The models applied within this synthesis likely reflect high heterogeneity that is typical within the construct of executive functioning, given their proposed role in providing cognitive control and coordination of “lower level” cognitive systems (Miyake et al., 2000). Subsequently, executive function has been a theoretically challenging construct to research for various methodological reasons (c.f. Rabbitt, 2004 for a review). Despite this, recent theorising has used psychometric techniques such as factor analysis, which have facilitated consensus regarding three “core” executive functions: Updating, Inhibition, and Shifting (Diamond, 2013; Miyake and Friedman, 2012). Associations between tinnitus and individual core executive functions are considered later when discussing narrow cognitive domains.

2.5.1.2 Processing speed (Gs)

Increased errors rates and response times in processing speed tasks were demonstrated through the Gs-rt and Gs-error models. The moderate correlation remained significant despite the removal of a potential outlier. Analysis of moderator variables revealed a significant effect of whether depression was measured as a confounding variable within the study. This suggests that some of the observed effect could be confounded by inclusion of participants with depression, which is known to be associated with impaired cognitive performance (Rock et al., 2014). Due to depression having an established association with severe tinnitus (Bhatt et al., 2017), it is necessary to consider it as a potentially confounding factor in the demonstrated association between tinnitus and cognitive performance. Adequate measurement of depression is therefore important to consider as a statistical “control” within analyses in primary studies.

Interestingly, summary correlations for both processing speed and executive function response time tasks were similar in size. This may be due to executive functions

being required to perform even seemingly simple tasks. Processing speed is thought to represent the speed with which an individual processes information, and tasks described as measuring processing speed are typically claimed to reflect underlying neural speed, efficiency, and capacity. While slower processing speed is associated with tinnitus, authors have noted that processing speed tasks are seldom free of cognitive control and therefore the association found between tinnitus and processing speed may reflect an artefact of an association between tinnitus and executive function (Cepeda et al., 2013; Duncan et al., 2005).

2.5.1.3 General short-term memory (Gsm)

The Gsm-correct model demonstrated poorer performance on general short-term memory tasks for participants who have tinnitus. However, there were only seven effect sizes used in this analysis. Furthermore, a leave-one-out analysis indicated that the statistically significant finding was not as reliable as the findings of executive function analyses and was driven by two studies.

A moderator model that included whether hearing thresholds were measured was statistically significant, and explained over 30% more variance. Hearing impairment associated with tinnitus and a known potential confounding factor for cognitive performance, due to a known association with poorer overall cognition (Dawes et al., 2015). It is therefore difficult to arrive at firm conclusions with the available data concerning the association between tinnitus and short-term memory function. Previous reviews have reported mixed findings and an inability to form conclusions based on the available evidence. They attribute this to variability in tasks and controls featured within studies (Mohamad et al., 2016; Trevis et al., 2017a). This review builds on this conclusion by demonstrating that hearing impairment is likely to be a confounding factor in the available evidence.

2.5.1.4 General learning and retrieval (Glr)

The Glr-correct model demonstrated an association between tinnitus and performance in general learning and retrieval tasks. This is an interesting finding as several theories

have suggested that long-term memory retrieval could be a significant feature of tinnitus generation and maintenance (De Ridder et al., 2014b; Sedley, 2019). Previous theoretical reviews of tinnitus generation have posited a potentially important role for memory, however, this has typically focused on short-term and working memory capacity. This review demonstrates an association between tinnitus and long-term memory retrieval. Furthermore, I was able to analyse specific narrow cognitive factors of general learning/encoding (Gl) and general retrieval (Gr), which are both considered below.

2.5.2 Which narrow cognitive domains are associated with tinnitus?

Tinnitus is associated with EF-Inhibition and EF-Shifting narrow factors within executive functioning (i.e. within Stratum I). More data are required to form conclusions regarding EF-Updating. While general short-term memory was found to be significantly associated with tinnitus, none of its constituent narrow factors were significant in isolation. This is likely due to insufficient data or the task choice not being specific to the theoretical aspect of general short-term memory under investigation. EF-Updating represents a promising cognitive factor for further research. Both general learning and general retrieval were found to be significantly associated with tinnitus. This cognitive domain is of emerging interest in cognitive performance research related to tinnitus.

2.5.2.1 EF-Inhibition, EF-Shifting, and EF-Updating

While no significant correlation was calculated for the EF-Updating factor, only two studies were available to contribute data. Due to the significant association with other executive function constructs, EF-Updating remains an important theoretical construct to consider (Ecker et al., 2014, 2010). For example, Trevis et al. (2016) used the N-back task and found statistically significant poorer performance between a group with subjective tinnitus compared to a control group. This suggests that all “core” executive functions are associated with tinnitus. Updating is intricately related to general short-term memory through being a theoretical functional component of working memory. Although several studies have included working memory measures, they have used

capacity measures, which attempt to measure limitations of short-term memory storage while performing operations, not the ability to update or refresh short-term memory.

2.5.2.2 Short-term memory (STM), Low working memory (LWM), and High working memory (HWM)

While the broad general short-term memory factor showed significantly poorer performance in participants with tinnitus, analysis of the narrow factors found that none were significant in isolation. This could be due to few available data points within each model, however, it may also result from the use of capacity tasks in previous study designs. These tasks may be insensitive measures of the Updating aspect of working memory, which can be conceptualised as a “refreshing” of a memory buffer and which may be less efficient in people with tinnitus. Poorer performance in narrow executive functioning factors of Inhibition and Shifting suggest tasks that operationalise Updating are more specific in measuring working memory deficits in people with tinnitus. This has empirical support from a study by Trevis et al. (2016) that used an N-back task, which better conceptually represents Updating; the authors reported a significant difference on performance on this task between participants with and without tinnitus. Webb et al. (2018) note that the literature distinguishes between two types of working memory task: those that deal with one stream of stimuli (for retrieval) and those that deal with multi-task coordination (i.e. two or more streams of information). It is interesting to note that many tasks analysed in this review were capacity tasks (Schmiedek et al., 2009).

2.5.2.3 General learning (Gl) and general retrieval (Gr)

Analysis of these narrow factors demonstrated a significant association between tinnitus and both Gl and Gr. Participants with tinnitus performed worse on tasks of long-term storage and retrieval of information from long-term memory. General learning and retrieval tasks have typically featured as control tasks in studies that were primarily interested in other domains of cognitive performance linked to tinnitus. The different sized summary correlations and confidence intervals of general learning and general retrieval indicate that poorer general retrieval performance is more likely (i.e. narrow

confidence interval) than general learning. It is also important to note that tasks investigating the ability to perform 'rapid retrieval' (i.e. pressured retrieval within a small timeframe) will necessarily involve executing strategy via executive functioning, which may moderate the observed association.

2.5.2.4 Cognitive self-report

Previous reviews have suggested more self-reported cognitive failures for participants who have tinnitus. This review revealed a significant association between tinnitus and self-reported cognitive performance, however, this appears to be driven by the large effect reported in Alam et al. (2012). According to the results of this analysis, the association between tinnitus severity and perception of cognitive performance is likely to be more modest. All studies in this analysis used the Cognitive Failures Questionnaire (Broadbent et al., 1982). Self-reported cognitive difficulties may be better assessed in a clinical tinnitus population through a specific screening tool such as the cognitive scale of the Tinnitus Functional Index.

2.5.2.5 Cognitive impairment screening

Both objective and subjective measures of cognitive performance were considered in this review as per our published protocol (Clarke et al., 2018). However, during the review process, it became evident that a distinct category was required for tests that screen for cognitive impairment. Such paper and pencil tests include the Montreal Cognitive Assessment (MoCA), Mini State Mental Examination (MMSE), and the Cognitive Abilities Screening Questionnaire (CASI) (Devenney and Hodges, 2017; Nasreddine et al., 2005; Teng et al., 1994). These measures are typically used in a clinical setting to screen for cognitive impairment in diseases such as dementia. It is important to note that the moderate correlation demonstrated with this analysis is being driven by two studies that feature reported correlations (i.e. correlations reported in the study manuscript, not calculated by the review team), which are unusually large (Lee et al., 2019; Wang et al., 2018).

The correlation reported by Lee et al. (2019) is unusually large and the reported association is likely confounded by having an older mean sample of tinnitus participants. While some studies have suggested that cognitive screening tools have clinical utility (Lee et al., 2019), it is uncertain how sensitive they are for patients with cognitive problems associated with tinnitus in addition to MCI per se. Use of both a cognitive impairment screening tool and a specific tinnitus-related concentration measure might therefore be afforded due consideration. Furthermore, choice of cognitive screening instrument is an important consideration, as with tasks of cognitive performance. In this review, there was large association between tinnitus and the MoCA cognitive screening tool. This provides limited support for the main study findings of an association between tinnitus and executive function because the MoCA includes short tasks that test executive functions, and was designed to investigate cognitive processes not tapped by the MMSE (Ciesielska et al., 2016).

2.5.3 Quality of available evidence

Although this analysis has demonstrated associations between tinnitus and several domains of cognition, there are methodological and conceptual considerations inherent in the available data that preclude a conclusion of tinnitus per se causing poorer memory or cognitive control, as has been suggested in previous studies (Heeren et al., 2014; Pierce et al., 2012; Schultz et al., 2018; Tegg-Quinn et al., 2016). Understanding the quality of the evidence currently available concerning associations between tinnitus and cognitive performance is important for developing models of tinnitus generation, its maintenance, and its perceived severity; this understanding will facilitate the identification of targets for intervening on elements within prospective models. For example, both cognitive training (broadly construed) and executive function training specifically are potential interventions for treating tinnitus (Diamond and Ling, 2016; Krings et al., 2015).

This review initially appeared to suggest that cognitive factors such as processing speed, short-term memory, long-term memory retrieval, and executive functioning are all appropriate targets for a training intervention; however, the quality of the available

evidence suggests that executive function training is a more promising interventional target due to its larger associations with tinnitus, while processing speed and short-term memory domains are confounded by depression and age. Furthermore, with an increased research interest in cognitive aspects of tinnitus being displayed in recent years, it is crucial that researchers are aware of various methodological and confounding factors that will affect conclusions that are drawn from primary research. The following sections discuss the pervasive issue of confounding, methodological issues within currently available evidence, and specific limitations to the conclusions that can be reached as a result of this synthesis.

2.5.3.1 Confounding variables

Although these analyses did not find age to be a statistically significant moderator within any model, this is likely due to the unit of analysis within models being studies not individuals, which may mask the effects of age on cognitive performance at the within-study level. Many studies within this review featured similarly aged samples, with approximately 70% of the meta-analysed records reporting a mean sample age of 40 to 50-years old. It is well established that absolute response time when performing cognitive tasks increases as a function of age, so much so that the utility of a null model within the cognitive gerontological literature has been dubbed the “dull hypothesis” (Perfect and Maylor, 2000; Salthouse, 1996; Verhaeghen, 2013). Future research should therefore ensure age is accounted for by matching groups or including age as a covariate in data analyses.

Participant hearing acuity is a further potential confounding factor requiring appropriate consideration in future research, with current evidence demonstrating associations between poorer cognition and reduced hearing thresholds (Wayne and Johnsrude, 2015). Reduced hearing thresholds are also associated with an increased prevalence of tinnitus. Although 31 studies within the review measured hearing using pure tone audiometry, the range of test frequencies measured, and subsequent reporting of mean average hearing thresholds were inconsistent, precluding detailed

coding of this variable. To facilitate future syntheses and comparisons between studies it is important to use a uniform criteria such as those provided by the World Health Organisation or British Society of Audiology (British Society of Audiology, 2018).

Prevalence of depression and anxiety symptoms increases with tinnitus severity (Bhatt et al., 2017) and both are associated with cognitive performance; 27 and 24 studies respectively included measurements of depression and anxiety. Sparseness of studies that did so within each cognitive domain, combined with differences in the questionnaires that were used (e.g. state or trait measures), made it difficult to synthesise the studies and understand their potential confounding influence. Future research would benefit from considering adequate measurement of depression and anxiety or featuring clinical levels as an exclusion criteria to exclude the contribution of these factors.

2.5.3.2 Methodological considerations

Few studies within this review report an *a priori* power calculation. A lack of statistical power limits the conclusions that can be drawn from published research. This synthesis makes clear that accurate estimates of associations between tinnitus and cognitive performance require adequately powered studies. This is particularly relevant in psychological research given recent debate concerning a 'replicability crisis' (Open Science Collaboration, 2015). Future research investigating links between tinnitus and cognitive performance would benefit from consideration of how best to obtain adequate inferential statistical power, such as sufficient sample size, and could power studies based on findings within this meta-analysis. For example, given that the summary correlations we typically observed are of the order of 0.3 or less, at least 85 participants would be required for 80% power at a significance level of 0.05 (Cohen, 1988, p. 100, Table 3.4.1). Further to the issue of sufficiently powering studies through appropriate sample size, we also note that approximately one in three studies within the reviewed literature failed to report the severity of tinnitus in sampled participants, while definitions of tinnitus were notable in their absence and/or variability.

A crucial methodological consideration for cognitive performance studies is the specific cognitive task used. Mohamad et al. (2016) suggested that multiple tasks should be used to assess the cognitive domain of interest. They also noted the importance of using validated tasks. Due to task impurity, “validation” in a cognitive sense is tantamount to majority consensus within the literature. It is therefore important to consider the underlying theory when investigating any potential behavioural impact of tinnitus using cognitive tasks. For example, the go/no-go task paradigm has been used as a measure of EF-Inhibition, but Diamond (2013) notes that this appears to tap a special case of Inhibition, with the construct being arguably better operationalised with the colour-word Stroop paradigm. Theoretical specificity of tasks assessing working memory and its constituent components should also be given due consideration in future work, such as complex span tasks versus tasks of Updating (Ecker et al., 2010; Schmiedek et al., 2009). A major area of study under such models has been trying to attain an understanding of the “capacity” of working memory, with many tasks that are used as working memory measures being capacity measures when Updating measures may be more appropriate in a tinnitus population (Oberauer, 2019; Schmiedek et al., 2009; Wilhelm et al., 2013).

Given the findings of this review, cognitive tasks that tap fluid ability (Gf) may be of interest in future tinnitus research. Crystallised intelligence (Gc) is the sum of knowledge acquired through the lifespan, and as such typically increases over the course of a person’s life, while Gf encompasses the ability to solve novel problems (McGrew, 2009). Diamond (2013) suggests that Gf is essentially the joint functioning of core executive functions (Updating, Inhibition, and Shifting). Therefore, Gf measures may be of particular interest in assessing clinical tinnitus populations; for example, although the MMSE has been the more popular MCI screening measure, the MoCA taps executive functioning and may therefore be more sensitive to concentration difficulties accompanying distressing tinnitus.

2.5.3.3 Synthesis strengths and limitations

A primary strength of this review is its bias limiting methods. The first was registration of the protocol on PROSPERO database (CRD42018085528). Second, a detailed protocol of the planned analysis was published in a peer-reviewed journal in advance of the review being completed (Clarke et al., 2018). Third, we limited bias produced through excessive researcher degrees of freedom through task impurity and subsequent categorisation of cognitive tasks, which may enable authors to pool studies in a manner that would produce a statistically significant finding, justified in a post-hoc manner. Using an existing cognitive taxonomy removed the pooling decision from the review team, limiting opportunities to introduce author bias. In addition to reducing author bias, CHC-M taxonomy provides a cogent, theoretically-derived analytic framework by which to structure a broad analysis of associations between tinnitus and cognitive performance. It also provided a theoretically-driven means of focussing meta-analyses to obtain maximum value from existing data available within the literature.

A limitation of this review is the varying amount of data available across the various broad cognitive factors. It is clear from the counts of observations across broad cognitive categories that the majority of research to date concerning tinnitus and cognition has focussed on attention (Acrani and Pereira, 2010; Cuny et al., 2004; Heeren et al., 2014; Husain et al., 2015; Mohamad et al., 2016; Roberts et al., 2013; Rossiter et al., 2006; Searchfield et al., 2007; Spiegel et al., 2015; Tegg-Quinn et al., 2016). This is evident as we were able to calculate three times as many effect sizes for executive functioning compared to processing speed (the cognitive category with the second largest number of observations).

High heterogeneity was evident across studies in many of the models in this review. This is unsurprising as many different kinds of studies were included. There were many anticipated sources of non-random variation, such as how tinnitus was defined and measured, hearing thresholds of participants, measurement of depression, or anxiety. This analysis anticipated sources of variation through *a priori* specification of data items

to extract as moderating variables (Clarke et al., 2018), but few of these were found to be statistically significant within the models tested. This suggests that other factors accounted for the high heterogeneity noted in many models. We suggest that a major contributor to the unexplained variance may be the type of task that was used in each study. While included studies cover many of the broad cognitive domains, some tasks operationalise these theoretical domains better than others. For instance, some tasks such as Test of Everyday Attention have been used in good quality studies but failed to observe findings of poorer performance in participants with tinnitus (Hoare et al., 2014). While such tasks have high face validity of measuring 'attention', they are unlikely to tap a specific theoretical factor within executive functioning, but instead put minimal load on several narrow factors, however, the data in this review do not allow for testing of this possibility.

2.6 Conclusions

The data presented in this chapter demonstrate that tinnitus is associated with poorer performance across a variety of broad cognitive domains including executive functioning (EF), processing speed (Gs), general short-term memory (Gsm), and general learning and retrieval (Glr). This analysis also indicates that specific narrow cognitive domains are associated with tinnitus to varying extents; these include Inhibition and Shifting (within EF) and learning and retrieval (within Glr). Associations between tinnitus and executive functioning therefore warrant further investigation because important unknowns remain concerning the nature of these associations. For example, experimental designs have typically treated participants with tinnitus as a homogeneous group and compared them with participants without tinnitus, and have therefore essentially considered only the presence of tinnitus in relation to cognitive performance and not the impact of tinnitus distress.

3 Chapter 3 – A comparison of executive functioning in distressing and non-distressing tinnitus groups (EFiT study)

3.1 Introduction

The purpose of this chapter is to build on the findings reported in Clarke et al. (2020) by investigating cognitive performance in participants with distressing and non-distressing tinnitus. First, I aim to provide an overview of the central cognitive concepts (i.e. executive functioning and fluid ability), followed by a description of the rationale and resultant novel experimental design that was used to compare distressing and non-distressing tinnitus groups. I will then present experimental data from the study, along with a pre-planned analysis that employs statistical and analytical techniques that have been typical within the field to date. Finally, I will present a critique of the inferential conclusions that can be made using these methods, and demonstrate that they may lead to specious conclusions concerning the relationship between tinnitus and cognitive performance.

3.2 What is Executive Functioning?

First described during the 1970s, executive functions (or executive abilities) are typically synonymous with cognitive control, which originated during the 1840s (Goldstein and Naglieri, 2014, p. 3; Nigg, 2017). This coincides with the infamous case of Phineas Gage, whose “disinhibited” behaviour after surviving an iron rod being propelled through his skull had a far-reaching influence on biomedical and psychological theorising (c.f. Macmillan, 2002). Executive functions have since been strongly linked to the frontal lobes (Alvarez and Emory, 2006). Many theorised executive functions have proliferated in the literature (Goldstein and Naglieri, 2014, p. 4), but Banich and Compton (2018, p. 334) suggest three broad theoretical perspectives:

3.2.1 Controlled Versus Automatic Processes Models

Various prominent theories distinguish between controlled and automatic processes (or bottom-up and top-down actions), and view the frontal lobes as playing a critical role in facilitating controlled behaviour. Shallice (1982) proposed a two-part system, consisting of contention scheduling and the Supervisory Attentional System (SAS). Contention scheduling facilitates relatively automatic processing through schemas. Schemas are a chain of actions, routines, or processing schemes that have become linked together; for example seeing a roundabout while driving may activate a schema consisting of actions chained together such as the application of brakes, turning of head towards wing-mirrors, and switching on an indicator light. Once activated, schemas continue to be active until inhibited by activation of a mutually incompatible process. In the second part of this model, the SAS is enlisted when no pre-existing schemas are available; this is typical of novel situations, or when tasks are technically difficult, or prepotent responses must be overridden (or inhibited). While originally conceived of as a unitary entity, more recent versions of the model consider the role of the SAS to be fractionated into various subcomponents that activate, inhibit, and monitor ongoing schemas (Shallice and Burgess, 1996).

3.2.2 Goal-Centred Processing Models

Goal-centred processing models have drawn from work in artificial intelligence, following from the observation that achieving complex, planned behaviours entail the specification of a goal, and the subsequent creation of various hierarchical sub-goals that serve the superordinate goal. Duncan (2013) posits an attentional hierarchy, where primate behaviour (including that of humans) is goal-directed and goal-orientated to optimally perform in response to environmental or internal demands, with a domain-general 'Multiple Demand System' that is distributed throughout various parts of the brain, including the parietal lobe, basal ganglia, and cerebellum. The Multiple Demand system hierarchically engages with various goals and sub-goals. The Goal-Neglect theory suggests that an impairment of executive functioning arises from an inability to specify

these goals, resulting in the poor planning and disorganised behaviour typically seen in people with frontal lobe damage (Duncan et al., 1996).

3.2.3 Multifactor Models

Many early models conceived of executive functioning as a unitary construct, however, multifactor models argue that it is best conceived of as separate factors. A prominent multifactor model is Miyake et al.'s (2000) Unity-Diversity model (Figure 3.1), which suggests that executive functioning consists of one common factor (Unity), with various specialised functions that are only required in specific situations (Diversity). One specialised factor is the ability to switch between tasks or sub-goals to appropriately guide behaviour; the other is to update information held in working memory, allowing novel information to be incorporated into ongoing processing. Similar models have been proposed that draw on patterns of impaired behaviour following brain damage (Stuss, 2011); separable basic sub-processes have been suggested including initiation, task-setting, and monitoring.

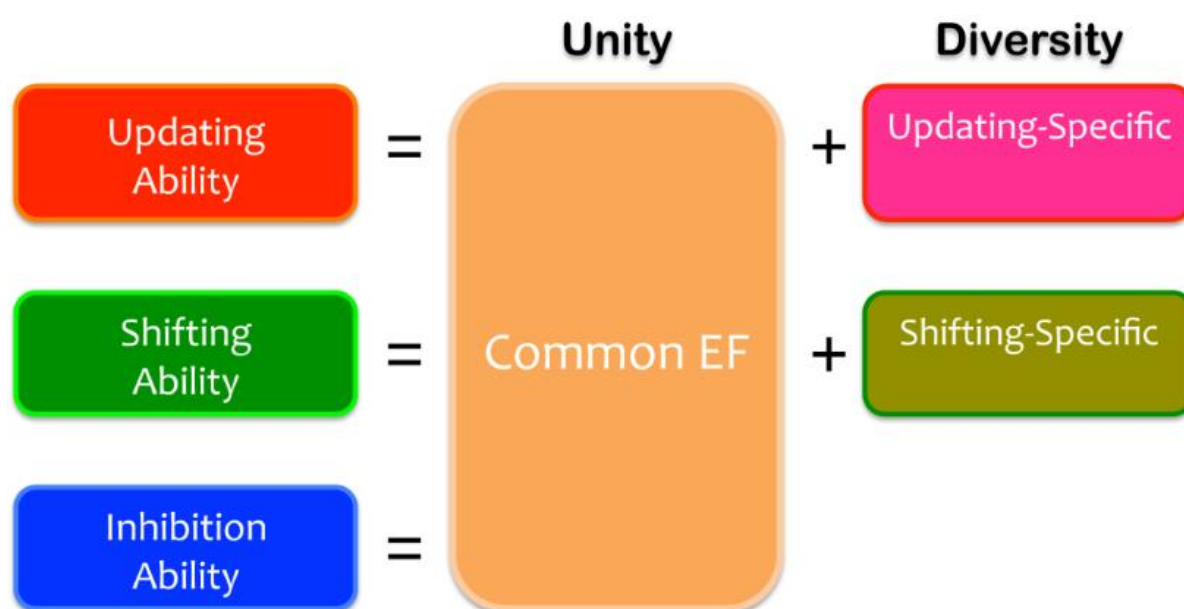


Figure 3—1. A schematic representation of the unity and diversity of three core executive functions from Miyake and Friedman (2012).

This brief overview highlights the multiplicity of viewpoints that have attempted, and continue to attempt, to understand the nature of executive functions, resulting in a complex and often difficult to parse taxonomy of related terminology (Nigg, 2017). However, the various theoretical perspectives generally converge on the following description of executive functioning as: a set of domain general, partially independent top-down functions that support goal-directed actions to facilitate complex cognition and perform novel tasks (i.e. tasks that are not “overlearned”). Webb et al. (2018) note that although executive functioning was not part of the original CHC model, many contemporary theorists view it as separable and necessary, arguing that it has utility in predicting a wide variety of quality of life and wellbeing outcomes (Diamond, 2013; Friedman and Miyake, 2017).

3.2.4 Core executive functions: Updating (of working memory), Inhibition, and Shifting

In recent years, multiple influential conceptualisations of executive functioning have emerged (Diamond, 2013; Friedman and Miyake, 2017). These conceptualisations converge in conceiving of ‘core’ executive functions that overlap and give rise to more complex behaviours. These core executive functions are Updating (of working memory), Inhibition, and Shifting. Working memory is distinguished from short-term memory in referring to information that is subject to manipulation while being held in memory. Updating of working memory specifically refers to the ability to rapidly update/delete or “refresh” the contents of working memory with task-specific information (Miyake and Friedman, 2012).

Inhibition (also referred to as inhibitory control) is a term generally used to describe the ability to deliberately override dominant, automatic, prepotent responses, or external lures in the service of a goal or action (Diamond, 2013; Miyake et al., 2000). It entails the overt stopping or inhibiting of a learned action, and when this fails individuals have been described as having executive dysfunction, or are described as exhibiting ‘perseveration’, which is continuing to perform a learned action or response that should be inhibited. An important feature of updating working memory and

inhibitory control is their co-occurrence and synergistic nature. Diamond (2013) argues that rarely, if ever, is one employed without the other.

The third core executive function of Shifting refers to the ability to switch effortlessly between tasks or mental sets (Miyake and Friedman, 2012), which Diamond (2013) refers to as 'cognitive flexibility'. There is a well described increased latency in having to jump between tasks, termed 'switch-cost' (Grange and Becker, 2019), which is "difficult if not impossible to eliminate".

3.2.5 Fluid intelligence (or fluid ability)

Contemporary psychological theorising conceives of executive functioning as separable from general intelligence (Friedman et al., 2006), however, there is noteworthy overlap in the domain of fluid intelligence (or fluid ability). Diamond (2013) describes fluid ability as synonymous with higher-level executive functions that facilitate reasoning, problem-solving, and planning (Figure 3.2). However, Duncan (2010) likens executive functioning to 'mental programs' that people create in order to complete various specific goals. In this analogy, fluid ability is not emergent, but is akin to an individuals' ability to "compile" specific mental programs.

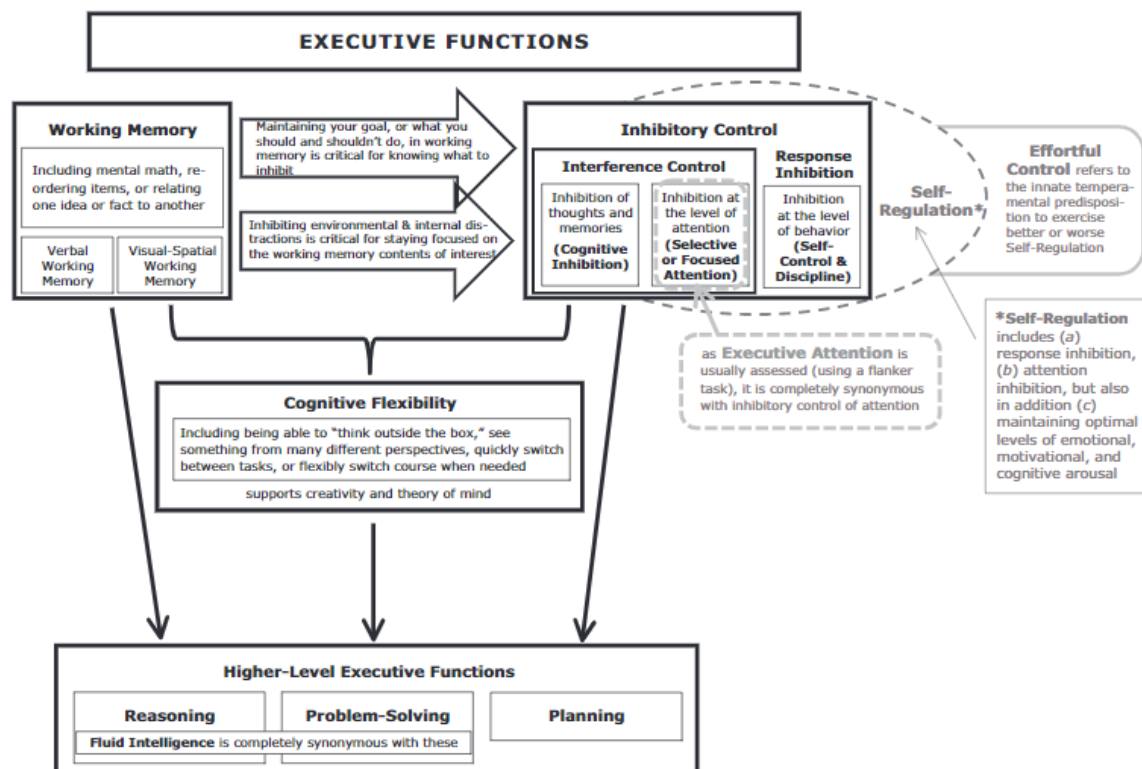


Figure 3—2. A schematic diagram of interrelations between Updating, Inhibition, Switching, Fluid ability, and higher-level functions from Diamond (2013).

3.3 EFIt Study (Executive Functioning in People with Tinnitus)

The following sections provide an overview of the focal study of this chapter entitled 'Executive Functioning in Adults with Tinnitus' or 'EFIt', by describing novel aspects the and rationale for the study design.

3.3.1.1 Novelty of Study Design

Novel aspects of the subsequent study design were based on insights from the systematic review and meta-analyses described in the previous chapter, as well as other reviews of the literature (Clarke et al., 2020; Mohamad et al., 2016).

The majority of previous experimental designs have grouped participants with 'tinnitus' as a single category and compared them with participants without tinnitus, resulting in experimental groups that frequently contained a mixture of participants with distressing and non-distressing tinnitus, and have therefore essentially only considered

the presence of tinnitus in relation to cognitive performance. It remains unknown as to whether only participants with distressing tinnitus have poorer executive functioning. This study explicitly considered perceived tinnitus distress and related clinical categories by featuring two tinnitus groups, determined using a common clinical measure of tinnitus distress. Additionally, confounding variables have generally not been well accounted for in previous experimental design (e.g. hearing loss); this has occurred through confounds not being measured, or when measured, having not been included in the inferential analysis of the data. Therefore, this study considered and measured various known confounding factors.

Many previous studies had not considered expected effect sizes through the inclusion of *a priori* power calculations (Clarke et al., 2020), making it problematic to interpret studies that report statistically non-significant results because such outcomes could be due to the absence of an effect or data insensitivity. Therefore, this study considered appropriate statistical power to find an effect through *a priori* power calculations, informed by the systematic review and meta-analyses described in the previous chapter. Furthermore, supplemental analytical methods were employed to enhance inferences regarding the experimental data.

Although Clarke et al. (2020) showed that tinnitus is associated with executive functioning (considered broadly), few studies have investigated its constituent factors. Therefore this study simultaneously investigated 'core' executive functions with tasks of varying complexity to further current understanding of the association between tinnitus and executive functions. Furthermore, many studies have used only a single measure to assess a cognitive domain of interest. Mohamad et al. (2016) note that doing so may introduce analysis bias due to task impurity, which is inherent in cognitive performance tasks. Therefore, this study used multiple cognitive performance measures to reduce analysis bias, enhance the attribution of performance to a specific construct, and thereby bolster inferences regarding associations between executive functioning and cognitive performance.

Many tasks identified in Clarke et al. (2020) essentially represented a “best fit” within each cognitive domain (i.e. a task was more representative of EF tasks than, for example Gsm, but may significantly tap both). Moreover, previous studies had either used bespoke paradigms that adapted established tasks based on psychological theorising specific to tinnitus (Waechter and Brännström, 2015), and/or tasks that may have not been optimal for assessing the cognitive construct they reported measuring (e.g. capacity measures of working memory, instead of Updating specific measures). Therefore this study used theoretical ‘exemplar’ tasks of the executive functioning narrow factors, that were based on established theoretical frameworks and operational suggestions (Diamond, 2013; Friedman and Miyake, 2017).

3.3.1.2 Study Rationale

The EFIT study featured a prospective, cross-sectional, between-groups design that included a battery of cognitive tasks that were designed to tap various executive functions. Three experimental groups were defined for this study: participants who did not experience tinnitus (No Tinnitus), participants with chronic tinnitus who were not distressed by the percept (Non-Distressing Tinnitus), and participants with chronic tinnitus who were distressed by the percept (Distressing Tinnitus). The experimental groups were established using the Tinnitus Handicap Inventory (THI), a unidimensional measure of tinnitus distress, frequently used in clinical settings, which has been recommended for use in a research setting (McCombe et al., 2001; Fackrell et al. in Signorelli and Turjman, 2014, p. 47). The Non-Distressing Tinnitus group was defined as having a mild handicap or no handicap (i.e. a THI < 38). While the Distressing Tinnitus group was comprised of participants with a moderate, severe, or catastrophic tinnitus handicap (i.e. a THI ≥ 38). Research concerning the validity of the THI shows it to be a unidimensional measure of tinnitus distress (Fackrell et al. in Signorelli and Turjman, 2014, p. 47).

Categorisation of the Distressing Tinnitus and Non-Distressing Tinnitus groups was chosen at ≥38 with a clinical descriptor of “Moderate”. The clinical descriptors that

accompany the Moderate category state that the tinnitus is “noticed in presence of background noise, although daily activities can still be performed”. If allocating of participants to groups is solely based on the THI original scoring manual, it could be argued that that distressing tinnitus would be mapped to a clinical descriptor of at least “Severe” tinnitus (i.e., THI score ≥ 58); the accompanying suggested clinical descriptor for Severe describes this grading of tinnitus as: ‘Almost always heard, leads to disturbed sleep patterns and can interfere with daily activities’.

However, given scrutiny of the item content, it should be noted that the clinical descriptors that accompany the numerical thresholds should serve only as a guide and are in no way a strict criteria mapping of distress; for example, despite its descriptor suggesting that Severe tinnitus “leads to disturbed sleep patterns”, only 1 of 25 questions present on the THI asks about sleep (Q7). This allows someone to obtain a clinical descriptor of Severe/Catastrophic tinnitus (≥ 58 points), while answering ‘No’ to the only question regarding difficulties with sleep. Conversely, someone may have Slight tinnitus (< 16), while answering ‘Yes’ to this question. Furthermore, relying solely on the clinical descriptors would suggest that the presence of a chronic tinnitus percept (as every participant had) is insufficient to cause distress, which was an assumption I avoided when planning the EFiT study. Therefore the grouping at a Moderate level on a unidimensional measure of tinnitus distress, frequently used within a clinical setting, was considered a pragmatic solution to assigning study groups.

Various representative tasks were chosen for each of the core executive functions described earlier, including Updating, Inhibition, and Switching (Section 3.2.4). A complete description of each task is provided in the ‘Methods’ and ‘Materials’ sections. Fluid ability was also included within the experimental design because it has not been investigated previously in a population with distressing tinnitus. Furthermore, because Diamond (2013) theorises that fluid ability is emergent from core executive functions; it was reasoned that this would be a strong theoretical candidate for further investigation, and may potentially exhibit the largest association with distressing tinnitus.

Clarke et al. (2020) noted that representative long-term memory tasks have typically featured as control tasks in studies that were primarily interested in other domains of cognitive performance linked to tinnitus, and this domain would therefore benefit from specific analysis. Long-term memory intersects with the core executive function factor of Inhibition through a reliable psychological effect called retrieval-induced forgetting (Section 3.3.1.3).

Related self-reported measures of theoretical interest were also included the study design. Self-report of cognitive failures was included because it has been used in numerous previous studies investigating tinnitus and cognitive performance (Alam et al., 2012; Hallam et al., 2004; Rossiter et al., 2006; Trevis et al., 2016). Furthermore, Clarke et al. (2020) estimated a self-reported cognitive failures effect size based on meta-analysis results, which facilitates comparison with EFiT results. Trait rumination was measured because it may be a causative factor in the perception of distressing tinnitus (McKenna et al., 2014), and it is also linked to experimental findings related to the retrieval-induced forgetting effect (Whitmer and Banich, 2010; Whitmer and Gotlib, 2013), which is discussed further in Section 3.3.1.3. Rumination has been linked to tinnitus through mechanistic accounts that posit that depressed thoughts about tinnitus are central to increased functional handicap, potentially even distorting the percept per se – making it seem “louder than it really is” (McKenna et al., 2014).

Depression, anxiety, and stress were measured because they are known covariates of distressing tinnitus. Their inclusion served to inform statistical models given anticipated differences in these measures in the distressing tinnitus when compared to the other two groups. After consideration of the wider literature, I also hypothesised that ‘cognitive reserve’ may be a pertinent construct to measure. This concept initially developed in the dementia literature to account for inter-individual variation in the ability to maintain cognitive functionality following brain damage, but has since been used in other research domains, and has been linked to executive functioning and concepts highly relevant to tinnitus such as maladaptive neuroplasticity (Barulli and Stern, 2013;

Opdebeeck et al., 2016; Roldán-Tapia et al., 2012; Stern, 2013, 2009). Cognitive reserve essentially represents an individual's exposure to cognitively enriching activity that may provide a neuroprotective effect.

3.3.1.3 Retrieval-Induced Forgetting (RIF) and Tinnitus

Clarke et al. (2020) also identified associations between tinnitus and general learning and retrieval (Glr). This finding was interesting because various theories of tinnitus generation and maintenance have implicated long-term memory (De Ridder et al., 2014a; Sedley, 2019), therefore behavioural evidence is required to understand the relationship between Glr and tinnitus distress. A retrieval-practise paradigm was therefore incorporated into the EFiT study to investigate the intersection between executive functioning and long-term memory (Anderson et al., 1994; Murayama et al., 2014). This paradigm elicits the retrieval-induced forgetting (RIF) effect - a phenomenon where remembering causes forgetting of other information in memory - which is linked to executive functions through inhibitory control (Anderson et al., 1994; Román et al., 2009).

When employing retrieval-practise paradigms, two findings have been consistently found: First, participants are typically able to recall more items that they practise (Rp+) compared to both those they do not. This is a common sense finding and arises from having practised these items in the Retrieval Practise stage of the paradigm. The second, and perhaps unintuitive finding, has demonstrated that participants typically show a poorer recall of unpractised items from practised categories (Rp-) compared to items that were not practised at all (Np). Said differently, the act of recalling some items during the Retrieval Practise phase causes participants to forget non-practised items from the same categories.

A leading theory postulates that the RIF effect arises as a consequence of typical inhibitory control (Anderson, 2003; Storm and Levy, 2012). Therefore, the absence of the RIF in clinical populations has been interpreted as evidence for compromised inhibitory control. Several studies have reported a reduced or absent RIF effect in clinical

populations, including people with frontal lobe damage, clinical depression, attention-deficit/hyperactivity disorder, and schizophrenia. If individuals with chronic tinnitus have impaired inhibitory control, then this may be evident through a reduced or absent RIF effect. The retrieval-practise paradigm therefore simultaneously provides a measure of inhibitory control through the RIF effect, and long-term memory through analysis of its subcategories (Section 4.3.4).

3.3.1.4 Study Objectives

Primary objectives of this study were established *a priori* to test the following hypotheses:

- i. Performance on measures of core EFs of Inhibition and Shifting are poorer in individuals with increased tinnitus distress; those individuals with distressing tinnitus perform worse on Inhibition and Shifting tasks than those without distressing tinnitus.
- ii. The performance on core EF measures of Updating is negatively associated with the presence of tinnitus (i.e. those with tinnitus perform worse on measures of Updating than participants without tinnitus).
- iii. Updating is poorer in individuals with increased tinnitus distress; those with distressing tinnitus perform worse on Updating tasks than those without distressing tinnitus.

Secondary objectives were:

- i. To identify which measures of core EFs are most sensitive to differences between those with and without distressing tinnitus.
- ii. To identify which EFs display the most potential for development as therapeutic targets for people with tinnitus.

3.3.2 Methods

This study was given a favourable review and granted approval by University of Nottingham School of Medicine Research Ethics Committee (Ethics reference number:

141-1811) and NHS Liverpool East Research Ethics Committee (IRAS project ID: 242476), and was subsequently adopted in the NIHR research portfolio.

3.3.2.1 Recruitment

Participants were recruited from the local community using online research participant recruitment websites, poster advertisement in public areas, and via a database of volunteers that is managed and maintained by the National Institute of Health Research Nottingham Biomedical Research Centre. Following a press release detailing the study by the University of Nottingham, participants were also solicited on a radio interview on BBC Radio Nottingham (March 27th 2019). Participants were also recruited following attendance at Nottingham Audiology Services and ENT services between April 2019 and June 2020. Although non-probability sampling methods (i.e. convenience and snowball sampling) were used to obtain research participants, the variety and nature of recruitment strategies/sites was aimed at obtaining a broadly representative sample of the population with tinnitus.

3.3.2.2 Inclusion criteria

In order to be eligible to take part in this study, participants had to satisfy the following inclusion criteria:

- Be fluent in English
- Be capable of providing informed consent
- Be literate and capable of comprehending questionnaires and automated, computerised, testing instructions
- Provide satisfactory evidence of meeting tinnitus-specific inclusion criteria (i.e. through satisfying investigators screening questions regarding the condition).
- Be aged 18-years up to 80-years, as age-related hearing impairment impacts a majority of individuals and cognitive functions are reported to significantly decline after this age (Gates and Cooper, 1991; Rosenhall et al., 2013; Treitz et al., 2007; Wattamwar et al., 2017)

- Have no more than a mild hearing loss (defined as no more than average hearing thresholds of > 40 dB HL in the better ear at 500, 1000, 2000, and 4000 Hz)
- Must not have been taking any psychoactive medications or other medication likely to interfere with performance in cognitive tasks
- Must have normal or corrected-to-normal vision
- Must have been able to give informed consent

3.3.2.3 Power calculation

Statistical power was calculated *a priori* using the Pwr package in R (Champely, 2020; R Core Team, 2018) based on preliminary results of the previously outlined meta-analysis (Clarke et al., 2020). Expected poorer EF cognitive performance was based on a combination of the lower boundary of the confidence interval for EF-Inhibition and a theoretically-driven, and a conservative effect size estimate ($d = 0.5$) given incorporation of theoretically optimal measures in the EFiT study. A power calculation for an independent one-tailed t-test suggested sufficient participants ($n = 48$ per group) to detect between-group differences at a typical experimental significance (0.05) and power (0.8).

3.3.2.4 Study measures

3.3.2.4.1 Pure tone audiometry (PTA)

Pure-tone audiometry (PTA) was performed using an adaptive staircase technique in 5 dB HL steps. Air conduction thresholds (0.25, 0.5, 1, 2, 3, 4, and 8 kHz) were obtained for each ear following the procedure recommended by the British Society of Audiology (British Society of Audiology, 2018), using a Siemens (Crawley, West Sussex, UK) Unity PC audiometer, Sennheiser (Hanover, Germany) HDA-200 headphones, and B71 Radioear (New Eagle, PA) transducer in a sound-attenuating booth. For inclusion in statistical analysis, the arithmetic mean value of thresholds values at .5, 1, 2, and 5 kHz was calculated to obtain the better-ear-average (BEA).

3.3.2.4.2 Objective cognitive measures

Table 3.1 summarises the objective cognitive measures used in this study. With the exception of the matrix reasoning task, all measures were generated and implemented

in the Python programming language by the author using PsychoPy Version 1.9 (Peirce et al., 2019). PsychoPy is an open-source application for the creation of experiments in behavioural sciences. Scripts containing code to implement the tasks used in this experiment are available at Open Science Framework (https://osf.io/gk5zh/?view_only=d2fa693eced945ea9e12c5dcce2d3957).

Table 3—1. Overview of objective cognitive measures, outcome type, and cognitive construct measured.

Objective Measure	Outcome type	Construct
Letter-number sequencing	Correct responses	Updating
Keep track task	Correct responses	Updating
Stroop task	Response time and error rates	Inhibition
Rule-switching task	Response time and error rates	Switching
Dual task	Correct responses	Switching/Listening effort
Retrieval-induced forgetting	Correct responses	Inhibition/Long-term storage and retrieval
Matrix reasoning	Correct responses	Fluid Intelligence

3.3.2.4.2.1 Letter-number sequencing

The letter-number sequencing task is generally considered to tap working memory, with Diamond (2013) suggesting that this paradigm is a strong theoretical exemplar of Updating within EF. During the task, participants are presented with a stimulus string containing a mixture of letters and numbers (e.g. C71). The stimulus must then be reordered and reported in the correct sequence, where numbers are reported first in ascending order, followed by letters in alphabetical order (e.g. 17C). The task increases in difficulty as letters and numbers are incrementally added to the stimulus string. Stimuli for this task were taken from the fourth edition of the Weschler Adult Intelligence Scale – Fourth Edition (WAIS-IV; Weschler, 2010). The task began with five practise trials featuring stimuli strings of two or three alphanumeric characters; the practise trials familiarised participants with the task and allowed the investigator to check that the task instructions were understood by the participant. The participant then completed thirty experimental trials containing two to eight alphanumeric characters. The stimulus string was white on a grey background, and was presented on screen for 2 seconds. This was

followed with a response box where participants had an unlimited amount of time to type their response. A total of 30 correct answers could be obtained in this task.

3.3.2.4.2.2 Keep track task

The Keep Track task is a measure of Updating where participants must update working memory through a process of remembering, forgetting, and updating various target words (Miyake et al., 2000; Yntema, 1963). In this task there are six categories (animals, colours, countries, distances, metals, and relatives); participants are randomly assigned 3-5 categories as 'target' categories at the start of each trial. During each trial, a sequence of words from all of the categories is randomly displayed on the screen for two seconds. Participants must recall the last word they see from their target categories and type them into the computer at the end of each trial. To verify conceptual understanding of the task, participants completed two practise trials with the investigator in the room before completing the main trials with no supervision. Participants completed three trials each with three target categories (3 trials), four target categories (3 trials), and five target categories (3 trials). A total of 36 correct target words could be achieved in this task.

3.3.2.4.2.3 Colour-word Stroop

The Stroop task is generally considered to be a measure of Inhibition and is based on one of the most well-known effects in psychology, originally published in 1935 by John Ridley Stroop (Diamond, 2013; MacLeod, 2014). In this task, the 'Stroop effect' is elicited through attempting to name the colour that a word is printed in, while ignoring the word itself; however, many variants of the Stroop task exist, with a wide variety already applied to researching cognitive performance in tinnitus (Andersson et al., 2005, 2000; Clarke et al., 2020; Jackson et al., 2014; Waechter and Brännström, 2015). The paradigm described in Jackson et al. (2014) was used in this study because it detected an effect between participants who had not sought help for their tinnitus, suggesting that it may be more sensitive to between-group effects. During trials, the words red, green, or blue were presented centrally on the screen. Words randomly

appeared in congruent (e.g. green printed in green), incongruent (e.g. green printed in red), or neutral conditions (e.g. the word character string 'xxxxx' printed in red, green, or blue). Participants were then required to use left, down, or right arrow keys to report the colour that each word was printed in. Participants first completed 15 practise trials with the investigator in the room to ensure their understanding. Participants then performed the main task with no supervision, completing a total of 150 trials of this task. Scoring for the task was based on the number of errors made by a participant and incongruent trial response time (see Section 3.4.2 for response time data processing details).

3.3.2.4.2.4 Rule switching

The Rule Switching task is described by Diamond (2013) as a strong theoretical exemplar of Switching. During each trial, participants are shown a stimuli pair consisting of both a letter and a number (e.g. A 1). A rule stating either 'Letters' or 'Numbers' also appears above the stimuli pair during each trial and served to direct participant attention to either the letter or the number of the pair. If the participant is directed to the letter, they must decide whether the letter is a consonant or a vowel; if the participant is directed to the number, they must decide whether the number is odd or even. The participant is instructed to press either the left or right arrow key respectively during either rule condition (i.e. left arrow = consonant or odd number, right arrow key = vowel or even number). The conflict caused by having to change rules during the task (i.e. response to a 'Number' condition then 'Letter' condition, as opposed to 'Number' then 'Number') reliably induces longer response times and increased error rates; this has been termed 'switch-cost' (Grange and Becker, 2019). Scoring for the task was based on the number of errors made by a participant and correct-only trial switch-cost response times (see Section 3.4.2 for response time data processing details).

3.3.2.4.2.5 Matrix reasoning

This Matrix Reasoning task is generally considered to be a measure of fluid intelligence and non-verbal problem solving (Duncan, 2010). Stimuli were used from the Stimulus

Book accompanying the WAIS-IV (Weschler, 2010). During the task, participants viewed a matrix of patterned shapes that contained one missing element from a matrix that should contain a logically complete sequence. Participants are provided with a selection of five responses, with one response from the selection correctly completing the logical sequence implied by the patterned shapes that were present in the matrix. Participants received one point for successfully selecting the correct answer, with a total of twenty-six points being available. As per the accompanying instructions with the WAIS-IV, participants completed the task after selecting answers for all available matrices, or if they selected an incorrect answer on three consecutive matrices.

3.3.2.4.2.6 Retrieval-induced forgetting

A retrieval-practise paradigm (Murayama et al., 2014) was used in this study with test items consisting of category-exemplar pairs used in the original paradigm describe by Anderson et al. (1994). The task comprised of three stages: Study, Retrieval-practise, and Final test (Figure 3.3).

- Study: during this phase, participants were presented with 48 category-exemplar pairs (e.g. fruit-lemon) from 8 categories (see Appendix B for full list of category-exemplar pairs). Pairs were presented one at a time in a randomised order for 5 seconds each. Participants were instructed to study the pairs for a subsequent test or think about the association between the category-exemplar pair.
- Retrieval-practise: participants were guided to retrieve half of the exemplars from half of the categories. Participants were provided with category-stem cues (e.g. fruits-le____). Participants had 10 seconds to type in the word, with each category-stem cue being shown randomly three times each.
- Final test: following a delay of at least 10 minutes (where participants completed the self-report measures included in this study), participants were tested on their ability to recall all 48 exemplars, prompted by a clue about which exemplar to recall (e.g. fruit-l). Participants had 5 seconds to say their answer out loud, while the investigator noted whether the correct word had been said. Requiring

participants to recall the category-exemplar pairs in a specified order is referred to as controlling for 'output interference', which limits the influence of potential non-inhibitory mechanisms (i.e. well known recency recall effects arising from recalling all of the Rp+ words first), and is therefore typically considered to allow results to be attributable to an inhibitory mechanism (Murayama et al., 2014).

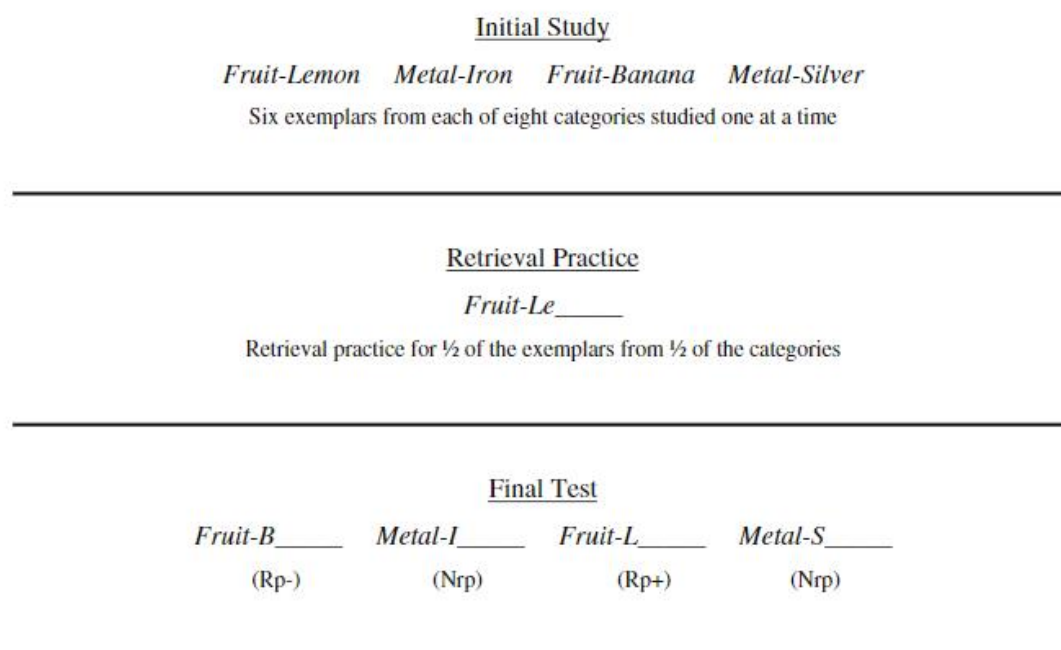


Figure 3—3. Schematic representation of the retrieval-practise paradigm from Anderson et al. (1994).

Practising only half of the categories and half of the words from those categories created the following stimuli descriptors: practised words from practised categories (Rp+), unpractised words from practised categories (Rp-), and unpractised words from unpractised categories (Np). The retrieval-induced forgetting (RIF) effect is defined as the difference in exemplar recall between Rp- and Np. There were 24 Np, 12 Rp+, and 12 Rp- items, providing 48 items in total. The main outcome variable of interest in this chapter was the RIF effect ($RIF = Np - Rp$) with a larger score indicating a greater effect.

3.3.2.4.2.7 Dual-task: digit recall and words-in-noise

Webb et al. (2018) consider dual-task paradigms as a measure of Shifting. This dual-task paradigm featured a primary speech-in-noise task combined with a secondary short-term memory task. In the primary task, single words were spoken over speech-babble noise. During the brief pause that followed each word, participants had to repeat the word they heard out loud, while an investigator made note of their answer. The concurrent secondary task entailed participants being shown a 5-digit string (e.g. 48790), and then being asked to remember and type this into the computer immediately after the primary task concluded. A sound level meter was used to set audio stimuli at 65 dB SPL (approximately conversational level). The primary task then featured an easy condition (where the stimuli words and babble had a +4 dB signal-to-noise ratio) and a difficult condition (with a 0 dB signal-to-noise ratio). Participants completed eight trials (four easy and four difficult). In the primary task participants could achieve a total of 40 correct words (20 in each condition). In the secondary task, participants were awarded one point for exact recall of the digit string, with a maximum of 8 points being available.

3.3.2.4.3 Self-report measures

3.3.2.4.3.1 Tinnitus sample case history questionnaire (TSCHQ)

The TSCHQ was developed as a standardised instrument for obtaining patient case histories, with the aim of being able to categorise them into meaningful subgroups (Langguth et al., 2007). It asks various questions about the tinnitus percept and associative information. Due to its aims of subgroup categorisation, no score or scale exists for the TSCHQ.

3.3.2.4.3.2 Tinnitus Handicap Inventory (THI)

THI features 25 items and was designed to measure individual's perceived handicap resulting from their tinnitus (Newman et al., 1998). However, scrutiny of the item content validity and various studies of its factor structure demonstrate the THI to be a unidimensional measure of tinnitus distress (Fackrell et al. in Signorelli and Turjman, 2014, p. 47). Each item features a 'Yes', 'No', or 'Sometimes' response, with a

numerical equivalent of 4, 0, and 2 points scored respectively. Scores are totalled and equate to five clinical categories: Grade 1 – slight or no handicap (0-6), Grade 2 – mild handicap (18-36), Grade 3 – moderate handicap (38-56), Grade 4 – severe (56-76), and Grade 5 – catastrophic (78-100). A total of 100 can be scored on the THI; larger scores represent increased perceived tinnitus handicap.

3.3.2.4.3.3 Cognitive failures questionnaire (CFQ)

CFQ features 25 items concerning failures in perceptions, memory, and motor function (Broadbent et al., 1982). It presents participants with various cognitive failures and asks them to rate their perceived frequency over the most recent six-month period (e.g. 'do you find you forget appointments?'). Participants can choose from five responses ('very often', 'quite often', 'occasionally', 'very rarely', and 'never'). A total of 100 can be scored on the CFQ; larger scores represent increased perceived frequency of cognitive failures.

3.3.2.4.3.4 Rumination response scale (RSS)

RRS is a 22-item measure asking respondents about the types of thoughts they typically experience when they are feeling sad, down, or depressed (Nolen-Hoeksema, 1987). Respondents rate the frequency of different types of thoughts they might experience in a sad, down, or depressed mood (e.g. 'think about how alone you feel') using a four-point scale (1 = almost never, 2 = sometimes, 3 = often, and 4 = almost always). A total of 88 can be scored on the RRS; larger scores represent increased tendency towards rumination.

3.3.2.4.3.5 Cognitive reserve index questionnaire (CRIq)

CRIq is an instrument for assessing cognitive reserve accumulated by individuals throughout their lifespan (Nucci et al., 2012). It is administered as a semi-structured interview featuring 24 items; for each item, participants are asked to provide estimates of the number of years that they have spent undertaking each activity relating to education, undertaking various categories of job role, and the amount of time spent undertaking various leisure activities. In doing so, the CRIq aims to obtain an estimate

of the number of years spent undertaking cognitively enriching activities with larger scores representing increased cognitive reserve.

3.3.2.4.3.6 Depression, anxiety, and stress 21 (DASS21)

The DASS21 is a shortened version of DASS questionnaire containing 21 items, which was psychometrically validated in clinical groups and a community sample (Antony et al., 1998). It comprises of three subscales: Depression (i.e. dysphoric mood such as feelings of sadness or worthlessness), Anxiety (i.e. symptoms of physical arousal such as panic attacks and fear responses), and Stress (i.e. symptoms of tension and irritability such as tendencies to overreact to stressful events). Participants are asked to rate the applicability of example symptoms during the last 7 days (e.g. 'I found it hard to wind down' for stress) using a 0-3 scale. Larger scores on the DASS21 represent increased depression, anxiety, or stress, with an accompanying categorisation system provided for 'normal', 'mild', 'moderate', 'severe', and 'extremely severe' (Table 3.2).

Table 3—2. Clinical categorisation scoring system provided with the DASS21.

Clinical Category	Depression	Anxiety	Stress
Normal	0-9	0-7	0-14
Mild	10-13	8-9	15-18
Moderate	14-20	10-14	19-25
Severe	21-27	15-19	26-33
Extremely Severe	28+	20+	34+

3.3.2.5 Experimental procedure

All participant testing sessions took place in a quiet, well-lit room, and took approximately 2 hours and 30 minutes to complete. Participants underwent PTA screening in a sound-treated booth to ensure that hearing thresholds met the study criteria of having no more than a mild hearing loss. Participants then completed the objective cognitive measures. To attenuate order effects (such as fatigue and context/contrast effects), a unique sequence of cognitive measures was assigned to each of the participants. This was achieved through a pseudorandom process in R statistical programming environment, where four unique testing sequences were generated with a seeded Latin-square function, and a random number generator then assigned each

participant a testing sequence. All objective cognitive tasks were performed on a HP laptop computer with a 15-inch screen running Windows 7.

Before completing the objective cognitive measures, participants sat directly in front of the computer while the investigator confirmed that they were comfortable and could see the written instructions on the screen. Each objective cognitive measure featured practise trials (with the exception of the dual-task); during all practise trials, the investigator sat with the participant to ensure that participants understood task instructions and were comfortable being left to complete the main trials without further supervision. Participants with tinnitus completed the TSCHQ and THI at the beginning of the experimental session. All participants completed DASS21, CFQ, and RRS during a planned break in the retrieval-practise paradigm (i.e. between the Retrieval-Practise and Final Test sections). The CRIq was completed by all participants as the final measure of the session. Refreshment and comfort breaks were offered to participants and provided as requested during the experimental session.

3.4 Data analysis

3.4.1 Data checking

All analyses were performed in R statistical programming environment (R Core Team, 2018), using the following additional packages: 'effectsize', 'tidyverse', 'bayestestR' (Ben-Shachar M et al., 2020; Makowski et al., 2019; Wickham et al., 2019). Summary statistics and graphical plots of the data were created to provide an overview of the dataset and distribution of variables. Histograms were created to assess distributional normality, skew, and kurtosis. Scatter plots were created to visualise bivariate relationships between variables.

3.4.2 Processing response time data

The Stroop and Rule-switching tasks used response time (RT) as an outcome measure and assume that this reflects differences in discrete processes of Inhibition and Shifting.

The inference typically relies on the mean RT of the individuals over many trials, however, this method may not be optimal on raw data because it is likely to be affected by being skewed, containing outliers, and being heteroscedastic, resulting in a decreased power to detect differences in their comparison (Whelan, 2008). Furthermore, a small proportion of RTs lie at the extremes of the RT distribution and can influence the estimate of central tendency, “contaminating” the estimate of the psychological process under investigation. This arises because RTs are not normally distributed; most of their mass is distributed to the left of the curve, with a long positive tail to the right. Practically, this means that they are better described by distributions such as the shifted-log normal, or ex-Gaussian distribution; the latter is a convolution of a normal and exponential distribution defined by parameters μ , τ , and σ (Figure 3.4). It is often appropriate to treat RT before using the mean of the distribution to limit the influence of non-normality of the distribution prior to analysis (Whelan, 2008).

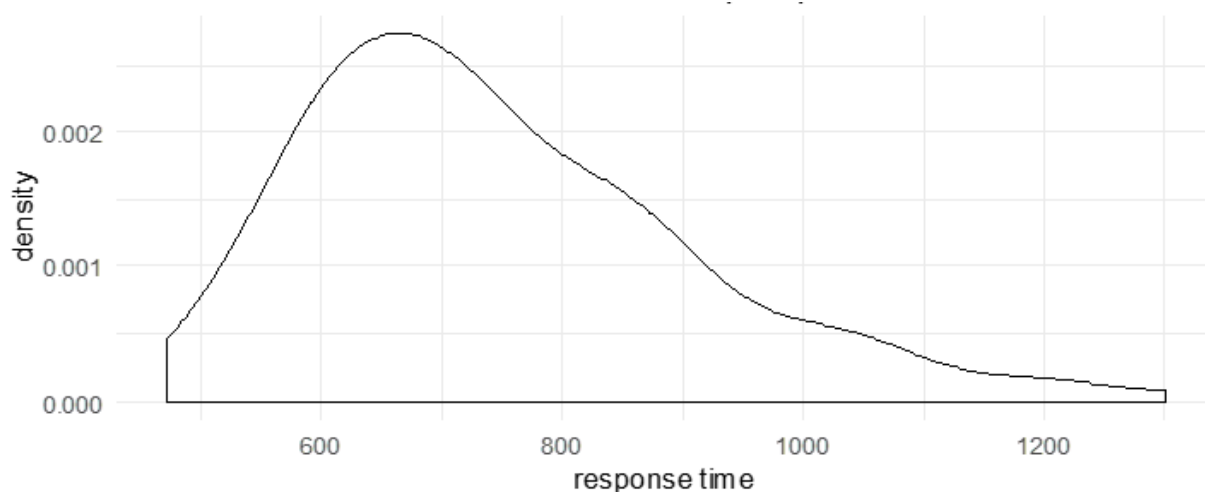


Figure 3—4. Simulated Stroop task ex-Gaussian density function of response time over 144 trials ($\mu = 600$, $\sigma = 80$, $\tau = 150$).

There are several options for the pre-averaging treatment of RTs, including the removal of outliers to provide a trimmed mean. Two common methods for obtaining a trimmed mean were used in this study: Absolute Value Criterion and Standard Deviation Criterion. Absolute Value Criterion involves identifying an absolute upper- and lower-limit for response times, while Standard Deviation Criterion removes any RTs that fall outside of

a specified number of standard deviations. The Standard Deviation Criterion (set at 2.5 standard deviations) was chosen for the Stroop task trimming, with a lower absolute limit of 200 milliseconds. As large RTs are possible in the Rule-switching task, no upper limit was specified, only a lower limit of 200 milliseconds to remove erroneous values arising from rapid anticipatory response. Standard Deviation Criterion was implemented using 'trimr: an R package of response time trimming methods' (Grange, 2015).

3.4.3 Statistical inference

There is still much debate concerning appropriate methodology for statistical inference, particularly concerning limitations, misapplications, and misunderstanding of null hypothesis significance testing (NHST), which remains the dominant paradigm in many applied sciences (Benjamin et al., 2018; Betensky, 2019; Cumming, 2017; Haig, 2017; Johnson and Whisman, 2013; Johnson, 2013; Mayo, 2018). In light of this, planned between-group inferential analyses were made using typical NHST t-tests, Bayesian t-tests, and standardised effect sizes (i.e. standardised mean differences).

3.4.3.1 Frequentist t-tests

The statistical significance of differences between experimental groups was tested in a frequentist framework using independent samples, one-tailed t-tests. One-tailed tests were chosen because of the *a priori* directional nature of hypotheses (i.e. poorer scores: Distressing Tinnitus < Non-Distressing Tinnitus < No Tinnitus). Welch's t-test was used because this version has been shown to provide better control of Type 1 error rates when the statistical assumption of homogeneity of variance between groups is not met, as is typically the case in psychological research (Delacre et al., 2017).

The aim of statistical significance within the prevailing Neyman-Pearson hypothesis testing paradigm is to avoid making inferential errors in the long run of experiments; therefore it has been argued that correction for multiple statistical tests within a study (or family-wise error rates) should be employed. The simplest way of achieving this is by using a Bonferroni correction, which divides the alpha level by the number of tests performed but may be overly strict. However, there is no consensus

regarding the necessity of family-wise error correction for statistical tests, and there are noted difficulties defining what a “family” of tests is (Althouse, 2016; Perneger, 1998). It was therefore decided to adopt a conventional statistical significance level for this study (i.e. $\alpha = 0.05$); this also provides congruence with previous studies, and ensures adequate power within the study based on the presented power calculations.

3.4.3.2 Bayesian t-tests

Bayesian statistical methods have been gaining popularity in the cognitive sciences in recent years (Dienes, 2011; Etz et al., 2018; Kruschke and Liddell, 2018).

Fundamentally, Bayesian methods consider the sampled data to be fixed, with a prior belief regarding parameter estimates being updated to a posterior uncertainty. This is contrasted with frequentist methods that consider ‘true’ parameter values over an infinite number of hypothetical experimental repetitions.

Bayesian methods are becoming more widely applied within biobehavioural sciences, and examples can also be found within hearing and tinnitus research (McMillan et al., 2014; McMillan and Cannon, 2019; Quintana and Williams, 2018). Bayesian methods are advantageous in the intuitive interpretation of model results. They enable substantive claims to be made about alternative models per se, while often NHST cannot (Dienes and Mclatchie, 2018). Quintana and Williams (2018) note that Bayes factors can typically be expected to fall within the range of 0.1 to 100 in the biobehavioural sciences (Figure 3.5).

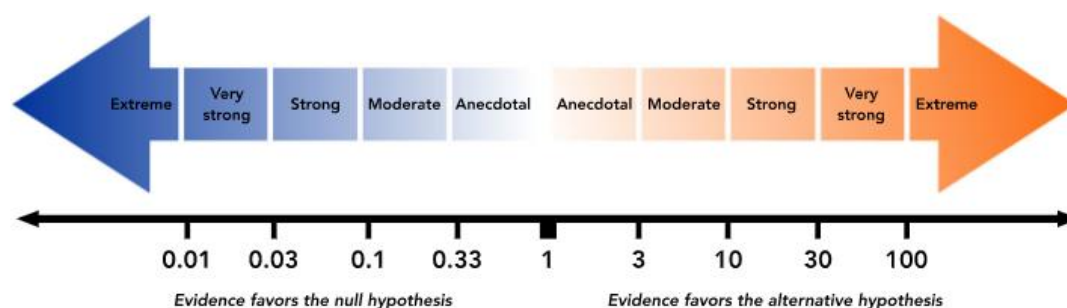


Figure 3—5. Spectrum of Bayes factor values that are typical within the biobehavioural sciences (Quintana and Williams, 2018).

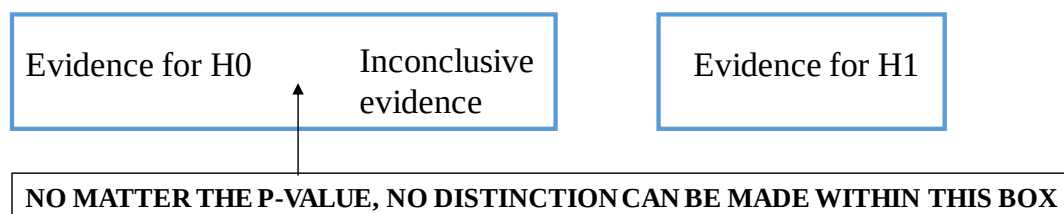
NHST can only provide information concerning the probability of the data given a hypothesis (Dienes and Mclatchie, 2018; Wetzels et al., 2011). However, we would typically like to understand the probability of our hypothesis given the data. These statements are not equivalent; i.e. $P(D|H) \neq P(H|D)$. 'Bayesian t-tests' provide information for $P(H|D)$ through the provision of Bayes factors.

NHST therefore arguably provides a weaker form of statistical inference, because no matter how large the p-value, it cannot distinguish between good evidence for H_0 from not much evidence at all (Dienes and Mclatchie, 2018). As such, NHST tests and p-values can only make a two-way distinction between evidential states and cannot provide direct evidence for H_1 . In contrast, Bayes factors can provide evidential support for H_1 , facilitating a three-way distinction (Figure 3.6).

a) States of evidence

Evidence for H_0	Inconclusive evidence	Evidence for H_1
--------------------	-----------------------	--------------------

b) What p-values provide



c) What Bayes factors provide

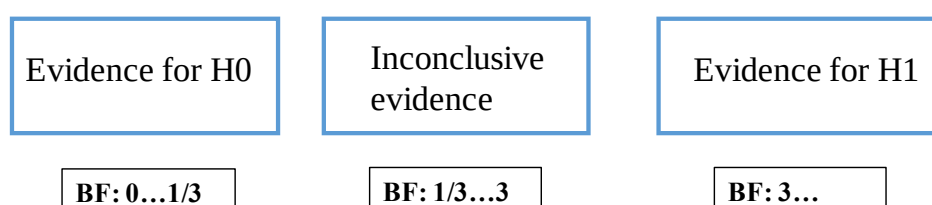


Figure 3—6. a) States of evidence. b) What p-values provide. c) What Bayes factors provide. (Adapted from Dienes and Mclatchie, 2018)

Technically, a Bayes factor is the ratio between the marginal likelihood of the null model and the alternative model (Quintana and Williams, 2018). Therefore, according to Bayes theorem, when comparing the model and data:

$$P(M|D) = P(D|M) \times P(M)P(D)$$

Then comparison of two models gives:

$$\frac{P(M_1|D)}{P(M_2|D)} = \frac{P(D|M_1)}{P(D|M_2)} \times \frac{P(M_1)}{P(M_2)}$$

Where the middle term is the Bayes factor:

$$\frac{P(D|M_1)}{P(D|M_2)}$$

3.4.3.3 *Standardised mean differences*

Standardised effect sizes provide magnitude estimates that are free from conflating factors such as sample size, as well as facilitating dissimilar measurements to be compared both within and across studies. Measures of effect size therefore provide unique information and reporting them is becoming increasingly recommended, and indeed required by some academic journals (Ellis, 2010; Wetzels et al., 2011). Furthermore, standardised effect sizes have the additional benefit of enabling primary study results to be compared to quantitative syntheses of existing bodies of literature.

A standardised mean difference (SMD) is typically used to convey effect sizes differences in between-groups research (Ellis, 2010). SMDs convey information about the effect in terms of standard deviation units, and can be used to interpret the performance in the different tasks within the study through placing them on the same relative scale. A further desirable feature of standardised effect sizes is that they can be converted between one another. This allows group SMDs to be calculated as point-biserial correlations and compared with effect sizes described in Clarke et al. (2020).

The SMD used in this study, is the between groups divided by pooled standard deviation:

$$SDM = \frac{M_1 - M_2}{SD_{pooled}}$$

To summarise, three inferential methods (NHST, Bayes factors, and standardised effect sizes) were used to understand group differences between participants with non-distressing tinnitus, distressing tinnitus, and without tinnitus. Table 3.3 provides a comparative summary of the relative interpretation of evidence for each method.

Table 3—3. Evidence categories for *p* values, for effect sizes, and for Bayes factors (adapted for Wetzels et al., 2011).

Statistic	Interpretation
<i>p</i>-value	
<.001	Decisive evidence against H0
.001-.01	Substantive evidence against H0
0.1-.05	Positive evidence against H0
>.05	No evidence against H0
<i>Effect size</i>	
<0.2	Small effect size
0.2-0.5	Small to medium effect size
0.5-0.8	Medium to large effect size
0.8	Large to very large effect size
<i>Bayes factor</i>	
>100	Decisive evidence for H1
30-100	Very strong evidence for H1
10-30	Strong evidence for H1
3-10	Substantial evidence for H1
1-3	Anecdotal evidence for H1
1	No evidence
1/3-1	Anecdotal evidence for H0
1/10-1/3	Substantial evidence for H0
1/30-1/10	Strong evidence for H0
1/100-1/30	Very strong evidence for H0
<1/100	Decisive evidence for H0

3.5 Results

3.5.1 Overview of study groups

A total of 139 participants were recruited to the three groups in the study: 48 Distressing Tinnitus (THI $\geq$ 38), 48 Non-Distressing Tinnitus (THI < 38), and 43 No Tinnitus participants. The 5 outstanding participants from the No Tinnitus group were due to premature recruitment cessation resulting from the COVID-19 pandemic (Fauci et al., 2020).

As anticipated, there were noteworthy differences between the groups with regards to average demographic and self-reported measures (Table 3.4). The No Tinnitus group were youngest on average, however as indicated by the relatively large standard deviation for age in this group, the mean age was skewed by inclusion of more young adult participants in the lower end of the distribution (Figure 3.7). The median age of Distressing Tinnitus and No Tinnitus groups was the same (60-years-old). The Non-Distressing Tinnitus group was the oldest on average according to both mean and median by 5-7 years.

Table 3—4. Group counts, means and standard deviations of demographic and self-reported measures.

	No Tinnitus	Non-Distressing Tinnitus	Distressing Tinnitus
Group count	43	48	48
Median age	60	67	60.5
Mean age	54 (21)	65.3 (10)	59.8 (9.8)
Mean hearing threshold (dB SPL)	15 (11)	22.8 (10.3)	18.4 (11.4)
Mean CFQ	38 (13)	35.5 (11)	46 (14.1)
Mean RRS	41 (14)	36.3 (10.5)	48.1 (13.5)
Mean CRIq	440 (220)	527.5 (162.3)	486.5 (140.8)
Mean DASS_D	6.1 (8.4)	3.5 (4.7)	9.7 (8.6)
Mean DASS_A	4.9 (6.2)	2.6 (2.8)	7 (6.7)
Mean DASS_S	9.7 (8.6)	7.6 (6.8)	16.9 (8)

Although all of the participants in this study had a better-ear-average (BEA) hearing threshold of no more than 40 dB HL, the mean hearing thresholds of the groups reflect the distributional variation in age. The Distressing Tinnitus and Non-Distressing Tinnitus groups had mean BEA hearing thresholds that were 3 and 8 dB HL poorer respectively; unlike mean group age, standard deviations for BEA hearing thresholds were similar between groups.

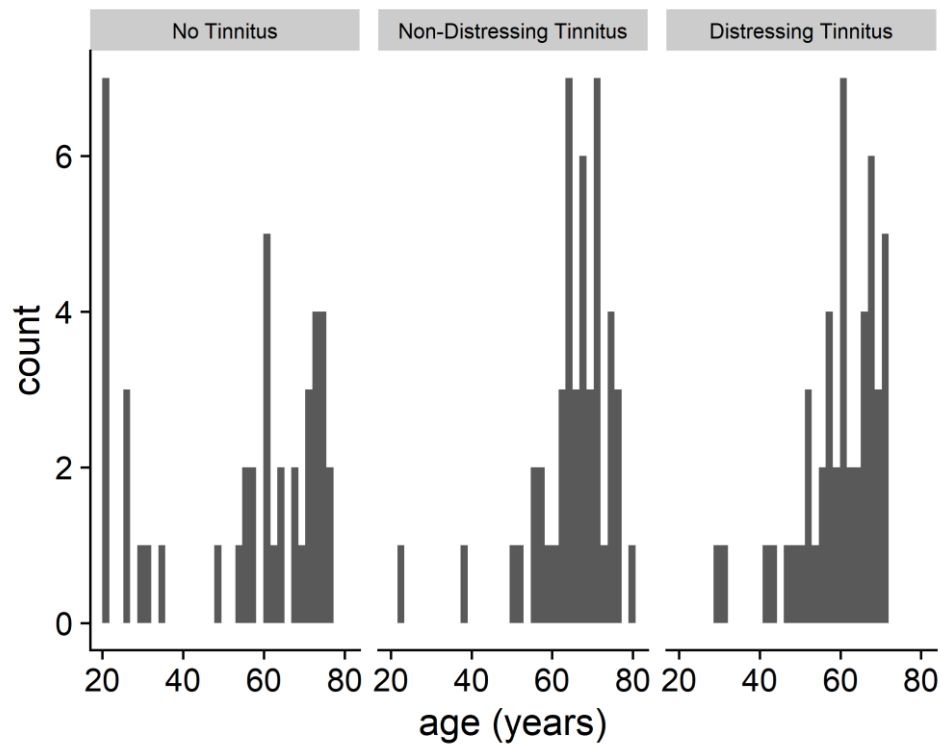


Figure 3—7. Distribution of ages for No Tinnitus, Non-Distressing Tinnitus, and Distressing Tinnitus groups.

Figures 3.8, 3.9, and 3.10 show stacked bar charts displaying the proportion of each category within each group. As expected from previous research, the Distressing Tinnitus group featured the largest proportion of participants with greater than normal levels of depression, anxiety, and stress.

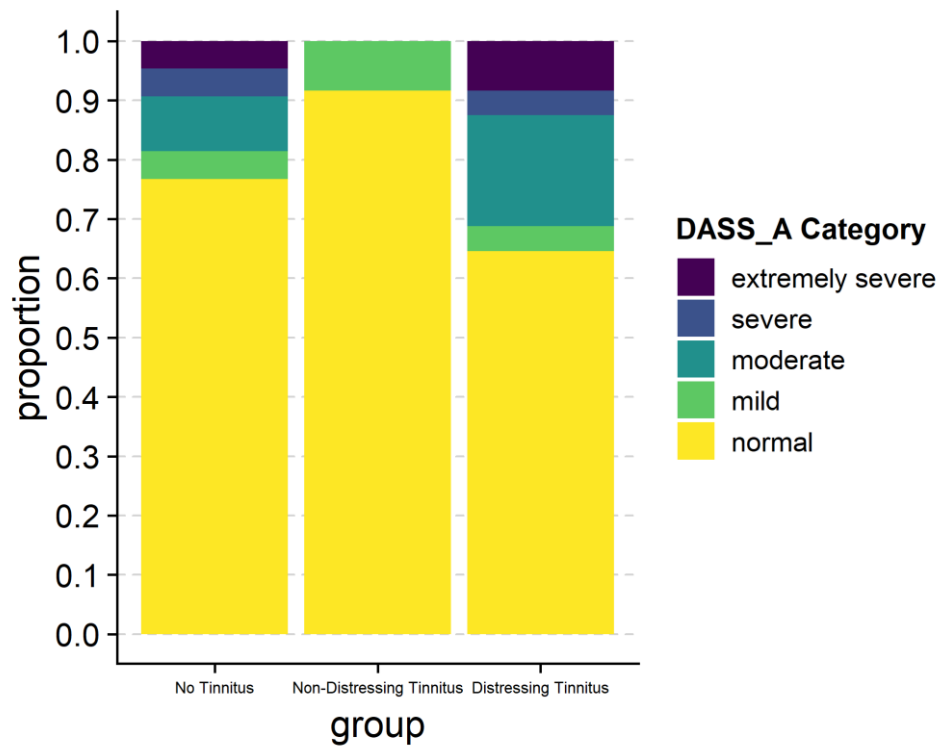


Figure 3—8. Proportion of DASS Anxiety score categories by group.

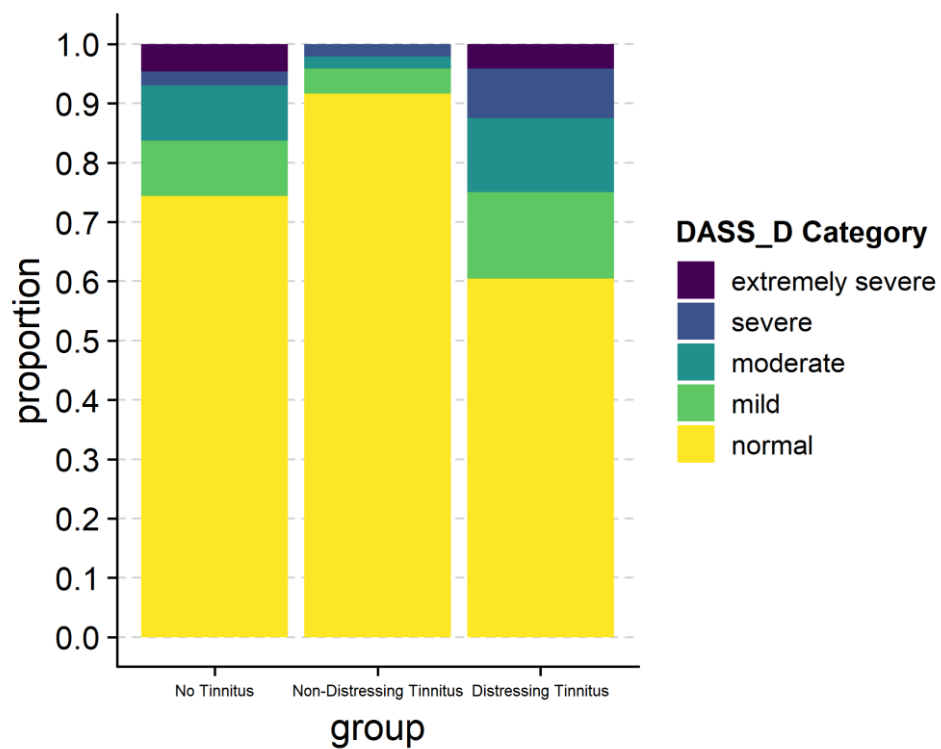


Figure 3—9. Proportion of DASS Depression score categories by group.

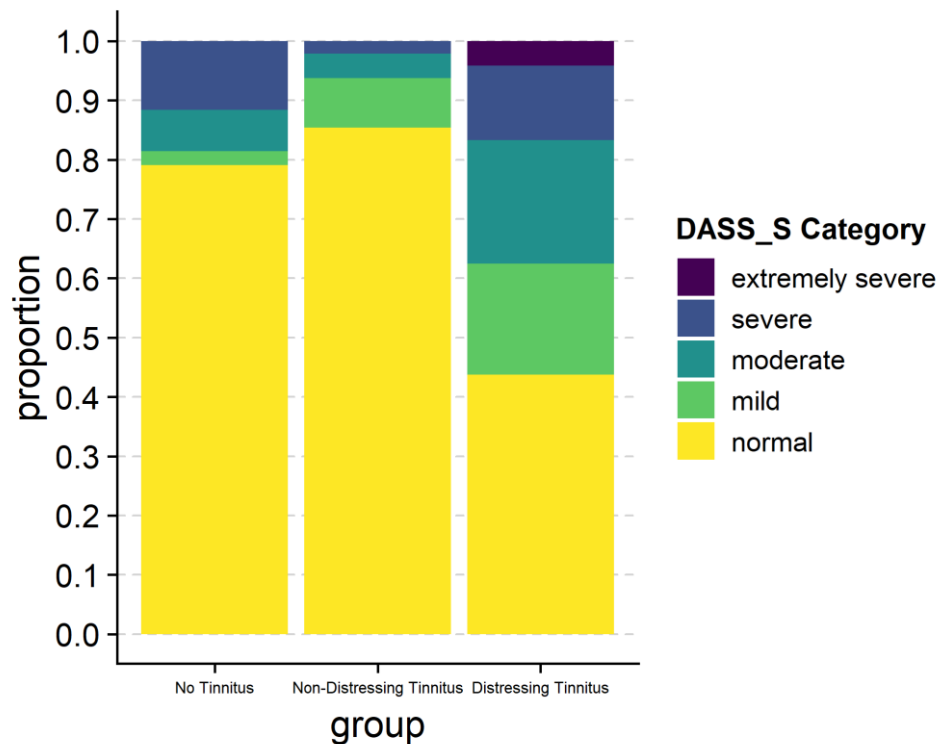


Figure 3—10. Proportion of DASS Stress score categories by group.

Figure 3.11 displays the proportion of participants in each group that have a moderate score on either depression, anxiety, or stress subscales. Less than 10% of the Non-Distressing Tinnitus group had a moderate score on at least one of these scales; this is contrasted with the Distressing Tinnitus group that had over 50% of participants, reporting a moderate score on at least one of the scales.

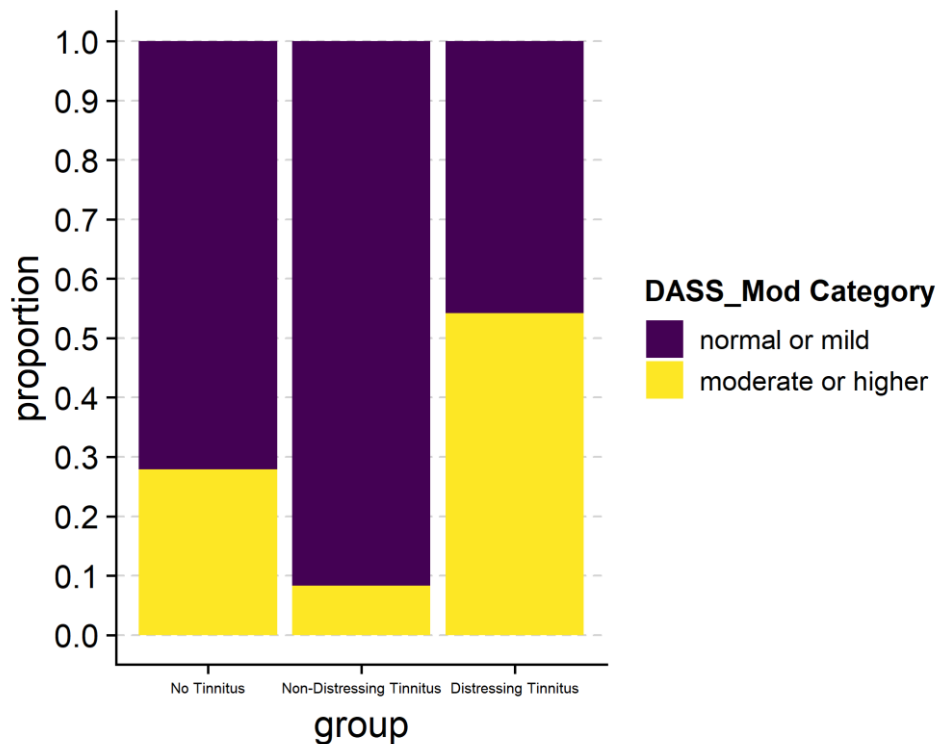


Figure 3—11. Proportion of moderate or higher DASS subscale score categories by group.

With regards to the remaining self-reported measures: the Distressing Tinnitus group had the highest scores on the CFQ, while the Non-Distressing Tinnitus group had the lowest scores. This indicates that the Distressing Tinnitus group had the most perceived cognitive failures, while those in the Tinnitus group had the least. The Distressing Tinnitus group also scored the highest in RRS, indicating that this group had the most ruminative thoughts. The Non-Distressing Tinnitus and No Tinnitus groups had similar mean RRS scores.

The Non-Distressing Tinnitus group had the largest score on the CRIq, indicating that they had the greatest cognitive reserve. However, this score is confounded by age (i.e. the longer you are alive, the more years you have to build cognitive reserve). A simple way to account for this is to divide the mean CRIq score by mean group age; however, this method does not model variation within groups. A simple linear model regressing CRIq scores on participant age and tinnitus group membership confirms the confounding, and reveals that increasing age is associated with increasing CRIq scores

($B = 9.5$), while being in the Non-Distressing Tinnitus ($B = -21.5$) or Distressing Tinnitus groups ($B = -9.7$) was associated with poorer cognitive reserve compared to the No Tinnitus group after holding age constant.

3.5.2 Descriptive group differences in objective cognitive measures

Table 3.4 shows that the Distressing Tinnitus Group had lower mean scores on every objective cognitive measure than the No Tinnitus and Non-Distressing Tinnitus groups (i.e. less correct answers and increased response time and errors depending on the task). With the exception of switch cost (tied score), switch errors (less), Stroop errors (less), and total of words remembered in the RIF task (more). The Non-Distressing Tinnitus group generally had lower mean scores than the No Tinnitus group.

Table 3—5. Group means and standard deviations of objective cognitive measures.

Cognitive Task	No Tinnitus	Non-Distressing Tinnitus	Distressing Tinnitus
Dual Task Words (easy)	11.8 (3.9)	9.5 (3.4)	10.9 (3.8)
Dual Task Words (hard)	9.6 (4.5)	6.5 (4.4)	7.8 (4)
Dual Numbers (total correct)	5.5 (2)	4.6 (2.1)	4.3 (2.1)
Matrix Reasoning (total correct)	20.4 (3.8)	20 (3.4)	17.6 (4.4)
Keep Track Correct (total)	24.7 (4.1)	22.5 (4.4)	22.4 (4.4)
Rule Switching - Switch Trials (seconds)	2.3 (0.7)	2.4 (0.7)	2.6 (0.8)
Rule Switching - Non-switch Trials (seconds)	1.8 (0.6)	1.9 (0.6)	2 (0.6)
Switch Cost	0.5 (0.4)	0.5 (0.3)	0.6 (0.4)
Switch Errors (total)	4 (4.1)	3.1 (3.1)	4.6 (4.5)
Stroop – Congruent Trials (seconds)	1 (0.3)	1 (0.3)	1 (0.2)
Stroop - Incongruent Trials (seconds)	1.1 (0.4)	1.3 (0.4)	1.3 (0.3)
Stroop - Neutral Trials (seconds)	0.9 (0.3)	1 (0.2)	1 (0.2)
Stroop Errors (total)	2.1 (2.3)	2.1 (2.9)	2.7 (4)
Number-Letter Sequencing (total correct)	21.7 (2.9)	20.9 (3)	20.5 (3.6)
RIF Rp+ (percent correct)	87.4 (12.8)	87.5 (13.2)	79.7 (17.8)
RIF Rp- (percent correct)	51.4 (16.3)	55.7 (17)	47.9 (16.9)
RIF Np (percent correct)	54.8 (15.9)	55.9 (16.8)	46.3 (16.6)
RIF Difference (percent correct)	3.5 (13.7)	0.2 (17.4)	-1.6 (14.8)
RIF Total (percent correct)	62.1 (12.7)	63.8 (13.3)	55 (14.5)

3.5.3 Standardised Mean Differences (SMD)

Table 3.6 contains SMDs and their respective 95% confidence intervals for all objective cognitive measures for the No Tinnitus and Non-Distressing Tinnitus groups, No Tinnitus

and Distressing Tinnitus groups, and Non-Distressing Tinnitus and Distressing Tinnitus groups. This information is plotted in Figure 3.12, using task performance from the perspective of the No Tinnitus group as a reference.

Figure 3.12 clearly displays poorer group performance across all tasks for the Distressing Tinnitus Group. These points generally sit between the Distressing Tinnitus and No Tinnitus groups, giving a linear impression of declining task performance as function of group membership. Notable exceptions are the number of Stroop errors made, and switch-cost, where there is no difference between the Non-Distressing Tinnitus and No Tinnitus Groups, the number of rule-switching task errors made, where the Non-Distressing Tinnitus group perform better (i.e. have less switch-cost) than the No Tinnitus group by .2 of a standard deviation. The largest SMD is for the Matrix reasoning ($d = .68$); this indicates that people in the Distressing Tinnitus group perform nearly .7 of a standard deviation worse on this task, which would be classified as a 'medium to large' effect according to Table 3.3. The majority of effects fall within the 'small to medium' and 'small' range according to typical classification, with the exception of the keep track task, which would also be considered 'medium to large' for both the Non-Distressing Tinnitus and Distressing Tinnitus group when compared to the No Tinnitus group.

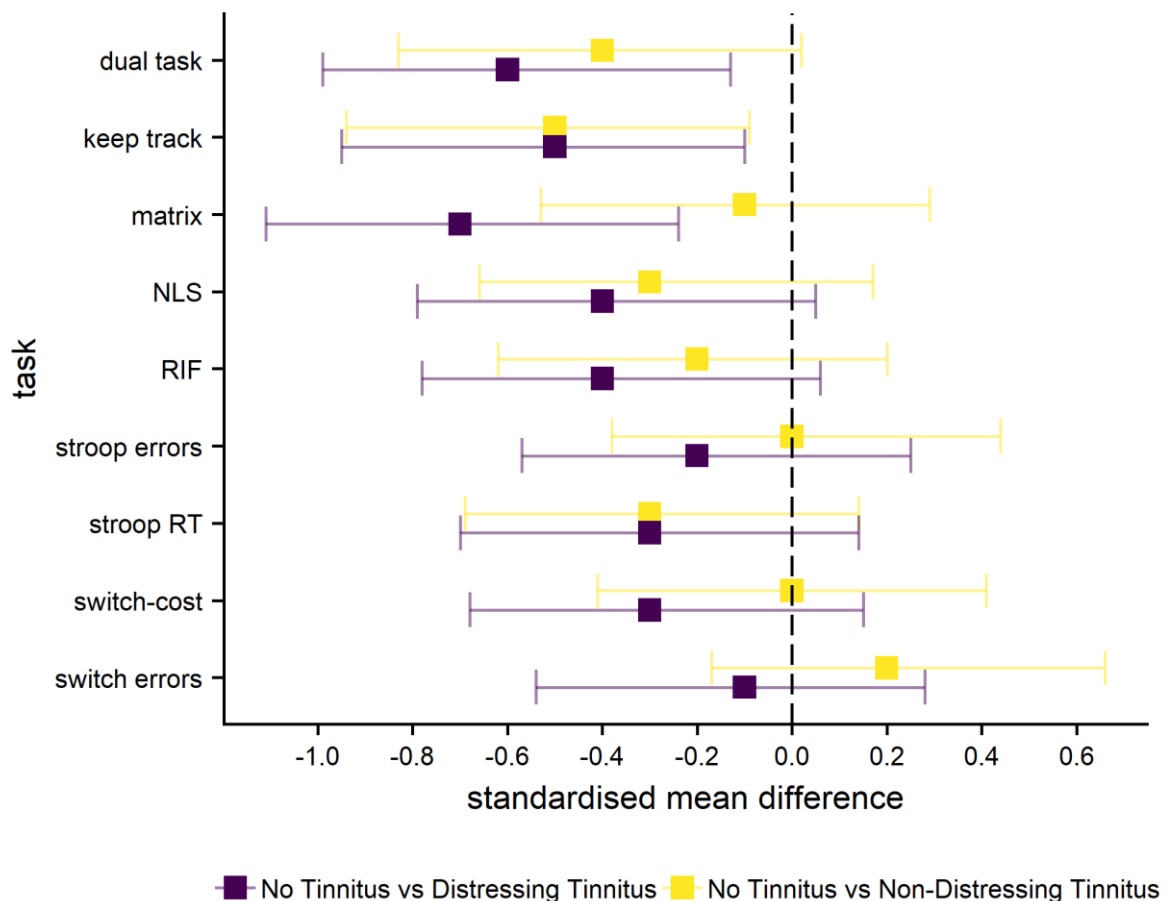


Figure 3—12. Plot displaying standardised mean differences (SMDs) for objective cognitive measures between No Tinnitus, Non-Distressing Tinnitus, and Distressing Tinnitus Groups. Positive SMDs represent superior performance to No Tinnitus group, while Negative SMDs show poorer performance. In this plot, the dashed vertical line represents the performance of the No Tinnitus group, while the purple and yellow squares represent the relative performance of the Distressing Tinnitus and Non-Distressing Tinnitus groups respectively. The signs of the SMDs have been transformed so that negative SMDs (i.e. squares to the left of dashed line) represent worse task performance than the No Tinnitus group, while positive SMDs (i.e. squares to the right of the dashed line) represent relatively better task performance.

Figure 3.13 essentially displays the same information, but shifts the point of reference, with the dashed line now representing the Tinnitus group. From this perspective, the fundamental pattern that was evident in the above plot becomes easier to see: purple dots, representing the SMD with the Distressing Tinnitus group, generally fall to the left of the line, while yellow dots generally fall to the right, representing the SMDs with the No Tinnitus group. Figure 3.13 also emphasises that there was essentially no difference in performance between the Non-Distressing Tinnitus and Distressing Tinnitus groups on the Stroop task (response time) or keep track task; there was also no difference in task

performance between the Non-Distressing Tinnitus and No Tinnitus groups on Stroop (errors) and switch-cost.

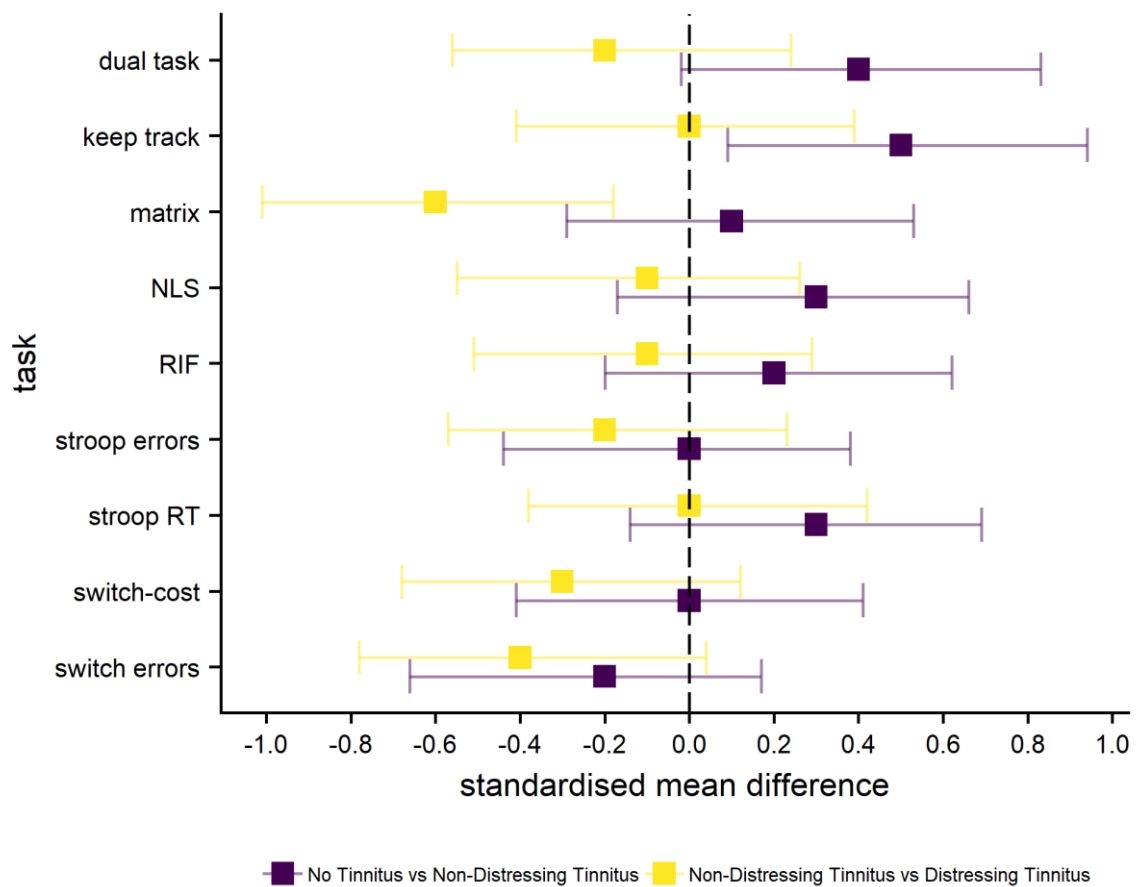


Figure 3—13. Plot displaying standardised mean differences (SMDs) for objective cognitive measures between Non-Distressing Tinnitus, No Tinnitus, and Distressing Tinnitus Groups. Positive SMDs represent superior performance to Tinnitus group, while Negative SMDs show poorer performance than the No Tinnitus group. In this plot, the dashed vertical line represents the performance of the Non-Distressing Tinnitus Group, while the purple and yellow squares represent the relative performance of the No Tinnitus and Non-Distressing Tinnitus groups respectively. The signs of the SMDs have been transformed so that negative SMDs (i.e. squares to the left of dashed line) represent worse task performance than the Non-Distressing Tinnitus group, while positive SMDs (i.e. squares to the right of the dashed line) represent relatively better task performance.

Both plots also make it easy to see that this analysis can only demonstrate clear none-zero effects with 95% confidence intervals (i.e. whether confidence intervals pass the dashed reference line); clear none-zero effects were evident for the matrix reasoning and dual task between the Distressing Tinnitus and No Tinnitus groups, and the keep track task between the No Tinnitus and both Non-Distressing Tinnitus and Distressing Tinnitus groups.

Table 3—6. Standardised mean difference between and confidence intervals for objective cognitive measures between No Tinnitus, Non-Distressing Tinnitus, and Distressing Tinnitus groups.

Task	Cohen's D	CI (lower)	CI (upper)	Groups
RIF	-0.21	-0.62	0.20	control vs tinnitus
NLS	-0.25	-0.66	0.17	control vs tinnitus
keep track	-0.52	-0.94	-0.09	control vs tinnitus
matrix	-0.12	-0.53	0.29	control vs tinnitus
stroop RT	0.27	-0.14	0.69	control vs tinnitus
stroop errors	-0.03	-0.44	0.38	control vs tinnitus
switch-cost	0.00	-0.41	0.41	control vs tinnitus
switch errors	-0.25	-0.66	0.17	control vs tinnitus
dual task	-0.41	-0.83	0.02	control vs tinnitus
RIF	-0.36	-0.78	0.06	control vs distressing tinnitus
NLS	-0.37	-0.79	0.05	control vs distressing tinnitus
keep track	-0.53	-0.95	-0.10	control vs distressing tinnitus
matrix	-0.68	-1.11	-0.24	control vs distressing tinnitus
stroop RT	0.28	-0.14	0.70	control vs distressing tinnitus
stroop errors	0.16	-0.25	0.57	control vs distressing tinnitus
switch-cost	0.27	-0.15	0.68	control vs distressing tinnitus
switch errors	0.13	-0.28	0.54	control vs distressing tinnitus
dual task	-0.56	-0.99	-0.13	control vs distressing tinnitus
RIF	-0.11	-0.51	0.29	tinnitus vs distressing tinnitus
NLS	-0.15	-0.55	0.26	tinnitus vs distressing tinnitus
keep track	-0.01	-0.41	0.39	tinnitus vs distressing tinnitus
matrix	-0.60	-1.01	-0.18	tinnitus vs distressing tinnitus
stroop RT	-0.02	-0.42	0.38	tinnitus vs distressing tinnitus
stroop errors	0.17	-0.23	0.57	tinnitus vs distressing tinnitus
switch-cost	0.28	-0.12	0.68	tinnitus vs distressing tinnitus
switch errors	0.37	-0.04	0.78	tinnitus vs distressing tinnitus
dual task	-0.16	-0.56	0.24	tinnitus vs distressing tinnitus

This SMDs were also converted to correlation (r) coefficients to facilitate a more direct comparison with the results obtained in the meta-analysis from the previous chapter (Tables 3.7 & 3.8). Clarke et al. (2020) noted in the Section 2.5.3.2 that approximately 1 in 3 studies within the reviewed literature failed to report the mean severity of tinnitus in sampled participants and many more studies failed to define the tinnitus population they were interested in, making it problematic to infer whether presence or severity of tinnitus were driving observed effects. Converting SMDs to correlations and comparing

results across groups provides some evidence regarding whether the meta-analysis results were being driven by the inclusion of participants with distressing tinnitus.

Table 3—7. Objective cognitive measure standardised mean differences between No Tinnitus and Non-Distressing Tinnitus groups converted to correlations.

Task	r	Groups	Meta-Analysis Category
keep track	-0.24	control vs tinnitus	updating
dual task	-0.20	control vs tinnitus	switching
NLS	-0.15	control vs tinnitus	updating
stroop RT	-0.15	control vs tinnitus	inhibition
RIF	-0.10	control vs tinnitus	inhibition/glr
matrix	-0.05	control vs tinnitus	fluid intelligence
stroop errors	0.00	control vs tinnitus	inhibition
switch-cost	0.00	control vs tinnitus	switching
switch errors	0.10	control vs tinnitus	switching

Table 3—8. Objective cognitive measure standardised mean differences between No Tinnitus and Distressing Tinnitus groups converted to correlations.

Task	r	Groups	Meta-analysis Category
matrix	-0.33	control vs distressing tinnitus	fluid intelligence
dual task	-0.29	control vs distressing tinnitus	switching
keep track	-0.24	control vs distressing tinnitus	updating
NLS	-0.20	control vs distressing tinnitus	updating
RIF	-0.20	control vs distressing tinnitus	inhibition/glr
stroop RT	-0.15	control vs distressing tinnitus	inhibition
switch-cost	-0.15	control vs distressing tinnitus	switching
stroop errors	-0.10	control vs distressing tinnitus	inhibition
switch errors	-0.05	control vs distressing tinnitus	switching

The strength of correlations suggests that the results in Clarke et al. (2020) were being driven by participants with distressing tinnitus. However, the effect sizes were smaller; this is consistent with Camerer et al. (2018), who reported that replicating meta-analytic effect sizes typically yields effect sizes that are smaller by 75%.

3.5.4 Frequentist t-tests

Tables 3.9-3.11 detail the frequentist t-test results across groups. Statistically significant differences (i.e. $p < 0.05$) were observed in the Keep Track and dual task between the No Tinnitus and Non-Distressing Tinnitus group. Differences between these two groups

from all other objective cognitive tasks were not statistically significant. Statistically significant differences were observed between the No Tinnitus and Distressing Tinnitus groups for the dual task, matrix reasoning, keep track, number-letter sequencing, and retrieval-induced forgetting difference scores, supporting H1. Differences were not statistically significant for Stroop task (response times or the number of errors) or the rule-switching task (number of errors or switch-cost). Finally, differences between the Non-Distressing Tinnitus and Distressing Tinnitus group were statistically significant for the matrix reasoning and rule switching task (number of errors), supporting H1. Differences for all other tasks between these groups were not statistically significant.

Table 3—9. One-tailed, independent samples t-test results for No Tinnitus and Non-Distressing Tinnitus groups.

Task	M1-M2	M1	M2	CI (lower)	CI (upper)	t-statistic	df	p-value	H1
keeptrack	2.20	24.67	22.48	0.72	Inf	2.48	89	0.01	greater
dual_task	0.84	5.47	4.62	0.12	Inf	1.94	88	0.03	greater
stroopRT	-0.11	1.15	1.26	-Inf	0.03	-1.30	84	0.10	less
nls	0.74	21.67	20.94	-0.29	Inf	1.19	88	0.12	greater
RIFdiff	3.31	3.49	0.17	-2.12	Inf	1.01	88	0.16	greater
matrix_reason	0.44	20.40	19.96	-0.82	Inf	0.58	85	0.28	greater
switch_cost	0.00	0.52	0.52	-Inf	0.12	0.00	85	0.50	less
stroopErr	0.08	2.14	2.06	-Inf	0.98	0.14	88	0.56	less
switch_error	0.90	4.00	3.10	-Inf	2.18	1.16	78	0.88	less ¹

Table 3—10. One-tailed, independent samples t-test results for No Tinnitus and Distressing Tinnitus Groups.

Task	M1-M2	M1	M2	CI (lower)	CI (upper)	t-statistic	df	p-value	H1
dual_task	1.17	5.47	4.29	0.45	Inf	2.69	88	< 0.01	greater
matrix_reason	2.79	20.40	17.60	1.36	Inf	3.25	89	< 0.01	greater
keeptrack	2.26	24.67	22.42	0.77	Inf	2.53	89	0.01	greater
nls	1.22	21.67	20.46	0.09	Inf	1.79	88	0.04	greater
RIFdiff	5.14	3.49	-1.65	0.17	Inf	1.72	89	0.04	greater
stroopRT	-0.11	1.15	1.26	-Inf	0.03	-1.33	73	0.09	less
switch_cost	-0.09	0.52	0.61	-Inf	0.03	-1.28	89	0.10	less
stroopErr	-0.53	2.14	2.67	-Inf	0.59	-0.78	76	0.22	less
switch_error	-0.56	4.00	4.56	-Inf	0.95	-0.62	89	0.27	less

¹ M1 = Group 1 Mean, M2 = Group 2 Mean, CI = Confidence Interval, df = Degrees of Freedom, H1 = Alternative Hypothesis

Table 3—11. One-tailed, independent samples t-test results for Non-Distressing Tinnitus and Distressing Tinnitus Groups

Task	M1-M2	M1	M2	CI (lower)	CI (upper)	t-statistic	df	p-value	H1
matrix_reason	2.35	19.96	17.60	1.02	Inf	2.94	88	< 0.01	greater
switch_error	-1.46	3.10	4.56	-Inf	-0.13	-1.83	83	0.04	less
switch_cost	-0.09	0.52	0.61	-Inf	0.02	-1.38	92	0.09	less
stroopErr	-0.60	2.06	2.67	-Inf	0.58	-0.85	86	0.20	less
dual_task	0.33	4.62	4.29	-0.38	Inf	0.78	94	0.22	greater
nls	0.48	20.94	20.46	-0.64	Inf	0.71	91	0.24	greater
RIFdiff	1.82	0.17	-1.65	-3.65	Inf	0.55	92	0.29	greater
keeptrack	0.06	22.48	22.42	-1.43	Inf	0.07	94	0.47	greater
stroopRT	0.01	1.26	1.26	-Inf	0.12	0.10	89	0.54	less

3.5.5 Bayesian t-tests

Tables 3.12-3.14 detail the frequentist t-test results across groups. Bayesian t-tests between the No Tinnitus and Non-Distressing Tinnitus groups providing support for H1 were consistent with frequentist t-test results, providing positive support for group differences for the keep track task (anecdotal) and dual task (moderate). Furthermore, self-report results showed anecdotal support for H1 for DASS_A and RRS. Results for the remaining objective cognitive and self-report measures showed support for H0 within the anecdotal range.

Table 3—12. Self-reported and objective cognitive measure Bayesian t-test results for No Tinnitus and Non-Distressing Tinnitus groups.

Task	Bayes Factor
Keep track	3.07
Anxiety	2.06
RRS	1.16
Dual task	1.13
Stroop-RT	0.47
Stress	0.46
Number letter sequencing	0.41
Switch errors	0.40
Retrieval-induced forgetting	0.34
CFQ	0.31
Matrix reasoning	0.26

Depression	0.26
Stroop-errors	0.22
Switch cost	0.22

Bayesian t-tests between the No Tinnitus and Distressing Tinnitus groups displayed some agreement with frequentist t-tests, providing support for H1 for the keep track, dual task, and matrix reasoning tasks. However, in contrast to frequentist t-tests, Bayesian t-tests for that number-letter and retrieval-induced forgetting tasks provided anecdotal support for H0. BFs for self-report measures provided support for H1 for RRS (anecdotal), CFQ (moderate), and DASS_S (extreme). Results for the remaining objective cognitive and self-report measures showed support for H0 ranging from anecdotal to moderate.

Table 3—13. Self-reported and objective cognitive measure Bayesian t-test results for No Tinnitus and Distressing Tinnitus groups.

Task	Bayes Factor
Stress	261.18
Matrix reasoning	18.17
CFQ	7.16
Dual task	4.95
Keep track	3.39
RRS	2.34
Number letter sequencing	0.87
Retrieval-induced forgetting	0.79
Anxiety	0.62
Stroop-RT	0.49
Switch cost	0.45
Stroop-errors	0.28
CRIq_aged	0.28
Depression	0.28
Switch errors	0.26

Bayesian t-tests between the Non-Distressing Tinnitus and Distressing Tinnitus groups displayed some agreement with frequentist t-tests matrix reasoning. However, in contrast to the frequentist t-test described in the previous section, the BF for the rule switching task (number of errors) provided anecdotal support for H0. BF for the self-

reported measures of CFQ, DASS_A, RRS, and DASS_S provide extreme support for H1. BFs for the remaining objective cognitive and self-report measures provided anecdotal to moderate support for H0.

Table 3—14. Self-reported and objective cognitive measure Bayesian t-test results for Non-Distressing Tinnitus and Distressing Tinnitus groups

Task	Bayes Factor
Stress	469697.31
RRS	2294.31
Anxiety	293.94
CFQ	228.22
Matrix reasoning	8.92
Switch errors	0.94
Switch cost	0.50
Stroop-errors	0.30
Dual task	0.28
Number letter sequencing	0.27
Retrieval-induced forgetting	0.25
Keep track	0.22
Stroop-RT	0.22
CRIq_aged	0.21
Depression	0.21

Both frequentist and Bayesian t-tests provide some evidence that differences in sampled data are attributable to expected to random variation, essentially helping to answer the question of *whether* differences between groups exist. However, it is important to understand the size of such differences. One way to do this is to consider standardised effect sizes.

3.5.6 Summary of results

3.5.6.1 Study objectives

The reported results fulfilled the primary *a priori* objectives of this study through testing the hypotheses outlined in Section 3.3.1.4):

- i. Inhibition and Shifting are poorer in individuals with increased tinnitus distress. Participants in the Distressing Tinnitus group performed worse on Inhibition and Shifting tasks than those in the Non-Distressing Tinnitus group.
- ii. Updating is negatively associated with the presence of tinnitus (i.e. those with tinnitus performed worse on measures of Updating than the No Tinnitus group).
- iii. Updating is poorer in individuals with increased tinnitus distress. Participants in the Distressing Tinnitus group performed worse on Updating tasks than those with Non-Distressing Tinnitus group.

Secondary study objectives were fulfilled by finding that:

- i. The matrix reasoning task was the most sensitive to overall differences between those with and without distressing tinnitus. But errors on the rule-switching task were most sensitive of the core executive functions.
- ii. All core executive functions appear to have potential for development as therapeutic targets for people with distressing tinnitus due to generalised poorer performance. This is similar for the Non-Distressing Tinnitus group, with the exception of Shifting.

3.5.6.2 No Tinnitus group contrasted with Distressing Tinnitus group

Taken together, results demonstrate poorer performance by the Distressing Tinnitus group in Updating, Inhibition, Shifting, and Fluid ability tasks compared to the No Tinnitus group. These differences are clearly evident in Figure 3.12, which displays poorer performance by the Distressing Tinnitus group across all task outcomes.

Frequentist t-tests generally supported statistically significant between-group differences in Updating measures (keep track and number-letter sequencing tasks) and fluid ability (matrix reasoning). Support was mixed for Inhibition with a statistically significant difference for RIF effect, but not for Stroop response time or errors. Support was also mixed for Switching tasks, with a statistically significant difference for the dual task but not for rule-switching (switch-cost or errors).

Bayesian t-tests provided additional information to frequentist t-test conclusions by offering direct support for H0 (of no group differences) for the Stroop and rule-switching tasks. Bayes factors displayed clear H1 support for matrix reasoning and dual task performance. However, Bayes factors for the RIF effect and number-letter sequencing conflicted with p-values, indicating anecdotal support for H0. Bayesian t-tests also highlighted the clear between-group differences in self-reported confounding measures detailed at the beginning of the analysis (Figures 3.8-3.11); this demonstrates the importance of including confounding factors within models. Confounding may reduce the power of t-test results and also make it difficult to attribute observed group differences to distressing tinnitus per se.

3.5.6.3 No Tinnitus group contrasted with Non-distressing Tinnitus group

Figure 3.13 shows that although Non-Distressing Tinnitus group performance generally falls between the No Tinnitus and Distressing Tinnitus groups, the size of differences between either groups is small and uncertain. Furthermore, conversion of SMDs to correlations displays a generally smaller and dissimilar correlations than found in Clarke et al. (2020), suggesting that the meta-analysis results were driven by participants with distressing tinnitus. Frequentist t-tests only supported H1 of worse performance than the No Tinnitus group for the keep track and dual tasks, but better performance than the Distressing Tinnitus group in the matrix reasoning task and rule-switch errors. Bayesian t-tests provided H1 support for keep track and dual tasks; and again provided additional information to frequentist t-test conclusions by providing direct anecdotal-to-moderate support for H0 for all other objective cognitive tasks. BFs between Non-Distressing Tinnitus and No Tinnitus, and Non-Distressing Tinnitus and Distressing Tinnitus groups highlighted the clear between-group differences in self-reported confounding measures, particularly in having lower scores for stress and anxiety.

3.5.6.4 Notes on inferential between-group tests

The analyses demonstrated several inferential incongruities between frequentist and Bayesian t-tests, where a statistically significant p-value was found to have a BF

providing anecdotal support for H0. These included number-letter and RIF effect for the No Tinnitus and Distressing Tinnitus groups, and rule-switching errors for the Non-Distressing Tinnitus and Distressing Tinnitus groups. The p-value for all of these frequentist tests was $p = .04$. This finding is consistent with an empirical comparison of 855 t-tests that found 70% of the datasets where the p-value fell between .01 and .05, had a corresponding default BF indicating that the evidence was only anecdotal or insensitive (Wetzels et al., 2011). The EFiT study highlights difficulties in applying NHST to studies of tinnitus participants employing cognitive measures, but also suggests that the primacy of NHST may help to explain previous mixed findings on this topic.

Bonferroni correction to adjust p-values in relation to tests of between-group differences was discussed in the Section 3.4.3, but was not applied for congruence with previous statistical analyses in this field. Conceptual difficulties related to NHST methodology such as defining a family of tests (Althouse, 2016; Perneger, 1998), and the inclusion of additional inferential support from Bayesian methods, also negated use of correction. However, a more stringent alpha value could have been selected through the application of such a correction or other arbitrary procedures. For example, a 0.01 alpha could have been selected for congruence between Bayes Factors and NHST methods in the wider published psychological literature (Wetzels et al., 2011). If this alpha was applied then additional between-group differences to those already reported would not have been statistically significant: i.e., the number-letter sequencing and RIF for the No Tinnitus and Distressing Tinnitus groups, and the dual task for the Non-Distressing Tinnitus and No Tinnitus group. If alpha of 0.001 was applied, then additional between-group differences to those reported and described at the 0.01 level would not have been statistically significant: i.e., the keep track task for both the No Tinnitus and Distressing Tinnitus groups, and the Non-Distressing Tinnitus and No Tinnitus groups. However, interpretation of the study results and related hypotheses were ultimately informed using a variety of analyses, statistics, and theoretical considerations, in addition to formal NHST and its documented challenges.

3.6 Discussion: Methodological and Inferential Limitations Concerning Associations

between Tinnitus Distress and Cognitive Performance

Figures 3.12 and 3.13 show clear between-group differences across all objective cognitive tasks, but inferential analyses provide mixed support for any conclusion that attributes this to H1. However, significant variation between the experimental groups was also identified that confounds inference concerning between-group differences, and negates a conclusion that the observed differences are the result of tinnitus distress groupings (Figures 3.8-3.11). Furthermore, these findings strongly suggest that the inability to decisively attribute poorer cognitive performance to having tinnitus is commonplace in the literature; this is because the inferential analyses for the EFiT study were selected based on typical practises within this literature (i.e. between-group statistical tests). This is important to scrutinise further because inferences based on such practises may have resulted in inconsistent and/or unsupported conclusions concerning individual study results.

Consequently, I revisited data analysis from all studies in Clarke et al. (2020) that analysed objective measures of cognitive performance and surveyed how the methodology of these studies accounted for the top confounding factors that were highlighted in the experimental data presented in this chapter (age, hearing, depression, anxiety, and stress). The survey revealed two fundamental issues relevant to previous research into tinnitus and cognitive performance: using p-values to conclude that experimental groups are equivalent, and (typically justified by the first practise) the omission of potentially important confounding variables from subsequent models when analysing experimental data.

The majority of studies investigating between-group cognitive performance differences describe the participants within their study as being matched. However, this is achieved through matching on group statistics/dichotomous inference based on p-

values, and not case-controlled matching (Rose and Laan, 2009). Commenting on similar practices in research concerning intellectual and developmental disabilities, Kover & Atwood (2013) note that 'methodological issues in group matching remain at the forefront of concerns regarding the progress of behavioural research'. Approximately two thirds of the studies in Clarke (2020) reported equivalence of groups across various confounds based solely on p-values being non-significant at a 0.05 alpha level. The practice of using p-values to establish group equivalence has been shown to not sufficient based on the standard .05 criteria, and generally been criticised for its high potential for false conclusions (Mervis and Klein-Tasman, 2004; Mervis and Robinson, 1999). Kover & Atwood (2013) concur with this because the results of a t-test 'depend on the observed mean difference between groups and the statistical power of the test'. They subsequently used a simulation study to demonstrate that smaller sample sizes will frequently lead to false conclusions regarding group equivalence because of sample size alone.

This is particularly relevant for research into tinnitus and cognitive performance because the variability in sample sizes (which are frequently small) leaves studies at risk of p-value variability bias. Reliance of these methods to infer group equivalence also accounts for the noted inconsistency between previous studies. Given that variability in potential confounds has been demonstrated in the data presented in this chapter, the pitfalls of assumptions inherent in p-value matching can be illustrated by considering the example of participant age: a variable that is well established as being associated with cognitive performance and thereby confounds a conclusion that group differences alone are attributable to tinnitus distress (Verhaeghen, 2013).

Figure 3.7 shows histograms of the distribution of participant age in each group. It is clear that the No Tinnitus group has more participants under 60-years-old, and variability exists between the two tinnitus groups in the distribution of elderly participants. However, when performing two-sided t-tests on age between groups, the tests yield non-significant results (i.e. $p > 0.05$). In line with common practise with this

literature, I may therefore conclude that age need not be modelled in the subsequent analyses and could not possibly be an explanatory variable of the observed results. Moreover, of the studies that did provide some statistical control within their modelling process, all of them selected covariates in subsequent models based on a similar significant p-value results (assuming a relevant confound had been measured in the first instance).

It can therefore be concluded that typical analytic methods within the field have likely lead to confounded, unverified conclusions concerning the attribution of poorer cognitive performance to tinnitus per se. More sophisticated modelling techniques are required to establish independent associations between poorer cognitive performance and distressing or non-distressing tinnitus. One such method is to model confounding variables explicitly via regression methods such as general linear models (Kover and Atwood, 2013).

In summary, it is commonplace to use p-values as evidence of group equivalence within the tinnitus and cognitive performance literature, and subsequently conclude that between-group variation in confounding factors could not have affected results. Similarly, because confounding group variability was observed in this study, it cannot be concluded that observed differences were due to tinnitus per se. However, the EFiT experimental design (Section 3.3) facilitates advanced analysis and modelling to establish associations between cognitive performance and tinnitus distress.

3.7 Conclusion

The experimental data presented in this chapter demonstrated a clear pattern of poorer cognitive performance in participants in the Distressing Tinnitus and Non-Distressing Tinnitus groups compared to the No Tinnitus group. While performance appeared to vary as a function of tinnitus distress, the results were also shown to be confounded by other variables that could impact cognitive performance (e.g. age, stress, depression, and

anxiety). The reported analysis, planned in advance, has been typical within the literature to date. However, inherent methodological limitations to such analyses were identified, which hinder subsequent inferences concerning the relationship between tinnitus and cognitive performance. These limitations may result in specious conclusions about the nature of this relationship. Consequently, previous evidential claims that tinnitus per se is associated with poorer cognitive performance are problematic. These findings highlight the need for advanced analyses and modelling techniques that adequately account for confounding variables between experimental groups, to establish whether cognitive performance and tinnitus distress are independently associated.

4 Chapter 4 – Advanced analyses and modelling to establish independent associations between cognitive performance and tinnitus

4.1 Introduction

The purpose of this chapter is to interrogate the conclusions of the previous chapter by performing advanced analyses and modelling of the experimental data it presented, to ascertain whether Distressing and Non-Distressing tinnitus are independently associated with poorer cognitive performance. I aim to employ advanced analytic and modelling methods that will include general linear modelling (GLM) of cognitive task performance, GLM analysis of item categories of the retrieval-induced forgetting task, and non-Gaussian hierarchical linear modelling of response time data. I then aim to discuss the findings of these analyses, with particular consideration of the size of the observed effects and how these considerations may bear on the relationship between tinnitus and cognitive performance.

4.2 Methods

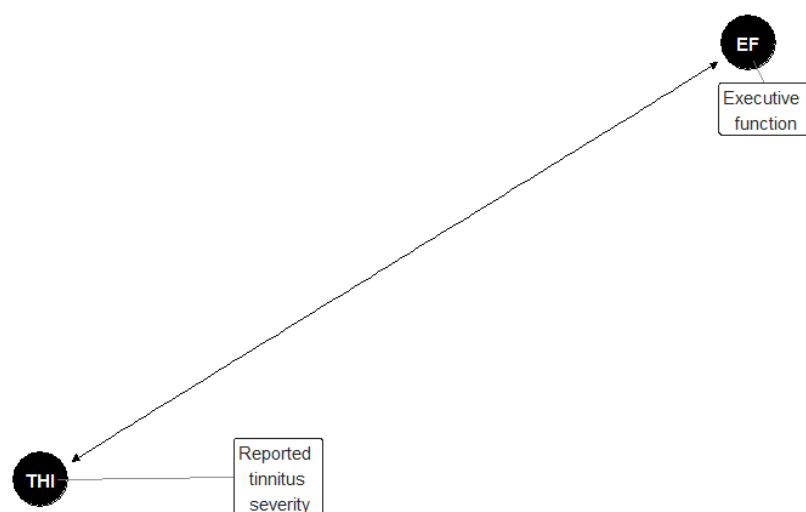


Figure 4—1. Simple associative model tested by between-groups analysis presented in Chapter 3, showing an association between THI (tinnitus severity/distress) and EF (executive function).

Due to the observational nature of this research, confounding factors were expected to exist between groups. For example, the group with distressing tinnitus was expected to have poorer hearing thresholds and higher scores for, depression, stress, and anxiety on average (Baguley et al., 2013; Mazurek et al., 2019). Between-group statistical tests (Sections 3.4.3.1 & 3.4.3.2) provide a very simple model of association between tinnitus and cognitive performance (Figure 4.1). However, it was anticipated *a priori* that significant confounding factors would be present that could account for differences in task performance between the study groups. This was confirmed in the previous chapter (Section 3.5.1), which showed notable between-group variation in age, hearing impairment, depression, anxiety, and stress.

All of these factors are thought to affect cognitive performance (Eysenck and Derakshan, 2011; Rock et al., 2014; Sandi, 2013; Verhaeghen, 2013; Wayne and Johnsrude, 2015). A more complex model is therefore required (Figure 4.2). General linear models enable the investigation of whether being in the Non-Distressing Tinnitus and Distressing Tinnitus groups are independently associated with poorer performance on the objective cognitive tasks.

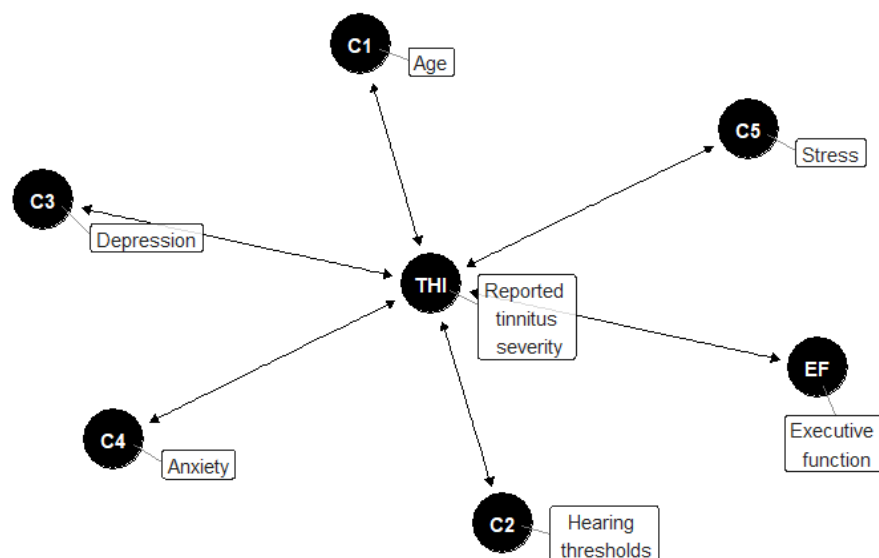


Figure 4—2. A more complex associative model of tinnitus distress showing an association between THI (tinnitus severity/distress), EF (executive function), and *a priori* confounders.

All of the subsequent analyses reported in this chapter were performed in R statistical programming environment (R Core Team, 2018), including the following additional packages: 'rstanarm', 'brms, and 'tidyverse' (Bürkner, 2017; Goodrich B et al., 2020; Wickham et al., 2019). In addition to data preparation and inspection described in Section 3.4, normal probability plots of standardised residuals (Chambers, 2018) were created to assess substantive departures from normality (i.e. outliers, skewness, or kurtosis).

4.2.1 General linear models (GLMs)

To infer whether group differences are associated with tinnitus per se, general multivariate linear regression models (GLMs) were constructed that included these covariates (Figure 4.2), and EFiT experimental groups (i.e. No Tinnitus, Non-Distressing Tinnitus, and Distressing Tinnitus). Including covariates in a GLM acts as pseudo-control to partial-out covariation (Figure 4.3), essentially leaving predictors interpretable as the unique association between the predictor and response while holding other variables constant (Cohen et al., 2003).

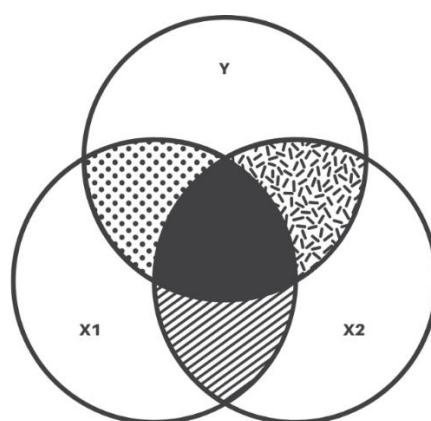


Figure 4—3. Diagram illustrating the conceptual rationale of estimated multiple linear relationship in multiple regression analyses (dotted region represents the joint variation between X1 and Y).

4.2.1.1 Statistical significance

GLMs were created in a frequentist framework for consistency with previous studies. However, the p-value was noted as a means of examining sources of inconsistency in previous research. McElreath (2016) argues that the most frequently employed strategy

for assessing the importance of variables within statistical models is a “stargazing” approach, which he describes as happening when a researcher essentially decides which variables are “significant” by assessing which coefficients have an asterisks next to them. While this indicates statistical significance as defined by a calculated p-value that falls below the specified alpha level, it conflates statistical significance with practical significance, as other statistics are likely to be of more interest to the research question being asked. Further, the statistical significance defined by the p-value is not a measure of effect size (Kruschke, 2010). Therefore the primary statistic of interest within these models was their standardised partial regression coefficients.

4.2.1.2 *Standardised partial regression coefficients (beta weights)*

Despite GLM flexibility, challenges arise regarding the interpretation of unstandardised partial regression coefficients because of difficulty in interpretation of the relative importance of each variable within the model. One method of addressing this interpretive challenge is by transforming predictor variables to the same scale through a process of standardisation. Standardising predictors entails subtracting the mean (μ) of the variable and dividing it by its standard deviation (σ), using the following formula:

$$Z_i = \frac{X_i - \mu}{\sigma}$$

where Z_i is the standardised predictor, X_i is the i th observation of a given variable, μ is mean of the variable, and σ is its standard deviation.

Multiple regression with standardised variables yields standardised partial regression coefficients (or beta weights), which provides a method of comparing the relative strength of predictors and allows them to be interpreted in standard deviation units while also statistically controlling for the influence of other variables within the model (Cohen et al., 2003, p. 82; Kline, 2016, pp. 31–32). Beta weights can be determined from unstandardized regression coefficients using the following formula:

$$\beta_i = B_i \frac{SD_i}{SD_y}$$

Where β_i is the beta weight, B_i is the unstandardised partial regression coefficient, and SD_i and SD_y are the standard deviations of the predictor and response variables respectively.

Because predictors within the model are transformed to the same standard deviation units, calculation of the relative strength of each predictor can be performed through taking a ratio of the beta weights; i.e. β_1/β_2 (Kline, 2016, p. 32). In addition to being able to obtain a ratio for the relative strength of association of each predictor within the model, standardised partial regression coefficients can also be ranked according to their size to give a sense of their 'importance', providing an ordinal metric to score variables within the model.

4.2.1.3 *Bayesian GLMs*

Analysis of the EFIT study data was also performed using Bayesian regression modelling. As with simple tests of group differences (t-tests and Bayes factors), multiple linear regression can be performed within a Bayesian framework. Within social and behavioural sciences, regression is most often estimated in a frequentist framework; however, using a Bayesian framework offers unique advantages concerning interpretability of parameter estimates and an increased flexibility that is afforded by being able to fit increasingly complex models (McElreath, 2016).

The key difference between frequentist and Bayesian regression models concerns the legitimacy of being able to describe uncertainty concerning parameters as probability estimates. The frequentist paradigm is concerned with the observed data given a fixed parameters and pertains to sequence of hypothetical datasets. In contrast, the Bayesian framework does not regard parameters as fixed and is interested in their probability distribution given the observed (fixed) data (Muth et al., 2018).

The primary distinction to be made from this contrast is the corollary interpretations that can be made from models fitted using each framework. While the frequentist approach can provide point-estimates and (occasionally) standard errors

based on long-run properties of estimators, it is unable to make probability statements about parameters.

In contrast, the Bayesian framework provides a probability distribution over all possible parameter values conditional on the model. Crucially, this permits probability statements to be made about parameters of interest that would be incorrect if interpreted in a Bayesian context; this helps to avoid intuitive misinterpretations of model parameters that are made by students, researchers, and statisticians alike (Gigerenzer et al., 2004).

Bayesian regression therefore permits a logically coherent answer to our substantive question of interest: 'what is the probability that the partial regression coefficient for distressing tinnitus is positive/negative?' Bayesian analysis can be decomposed into four key steps, which were followed during analysis of the EFiT study data (McElreath, 2016; Muth et al., 2018):

1. Specify the data model and prior
2. Estimating the model parameters
3. Checking sampling quality and model fit
4. Summarising and interpreting the results

The frequentist GLMs described above, and their associated confidence intervals or p-values, do not permit a straightforward interpretation of the statistical question of interest: what's the probability that having tinnitus or distressing tinnitus causes poorer task performance (i.e. that they have zero effect after controlling for other variables in the model). Therefore, Bayesian GLMs including the same predictors as the frequentist GLMs were fitted for all objective cognitive measures.

Bayesian GLMs were fit using `rstanarm` (Goodrich B et al., 2020), which focuses on applied Bayesian regression models using Markov Chain Monte Carlo (MCMC) sampling through an R interface to the Stan probabilistic programming language. By default, `rstanarm` uses weakly informative priors for parameters, which are generally preferable

to uninformative priors (typically normal and half-Cauchy respectively for predictors and standard deviations), particularly for fitting models where limited prior information is available (McElreath, 2016; Muth et al., 2018). All GLMs converged successfully, as assessed by trace plots and potential scale reduction factor statistics (i.e. R-hat values < 1.1 for all model parameters). Effective posterior sample size (ESS) was also sufficient for all GLMs with the number of effective samples achieving > 1000 for all parameters (Muth et al., 2018). Posterior predictive checks (PPC), where data is simulated according to the fitted models, were also performed on all GLMs for objective cognitive measures.

4.3 Results

4.3.1 Frequentist GLMs

Table 4.1 provides the GLMs predicting performance across all objective cognitive tasks with tinnitus group membership, age, hearing threshold, depression, anxiety, and stress as predictors.

Table 4—1. Unstandardised general linear model results predicting objective cognitive measures with tinnitus group, age, hearing threshold, depression, anxiety, and stress as predictors.

term	estimate	std.error	statistic	p.value	task
(Intercept)	25.658	1.298	19.763	< 0.001	nls_correct
tinnitus	0.177	0.683	0.260	0.796	nls_correct
distressing_tinnitus	-0.824	0.705	-1.170	0.244	nls_correct
age	-0.074	0.023	-3.227	0.002	nls_correct
hearThresh	-0.007	0.028	-0.252	0.801	nls_correct
DASS_A	-0.056	0.063	-0.887	0.377	nls_correct
DASS_D	0.034	0.046	0.753	0.453	nls_correct
DASS_S	0.012	0.043	0.281	0.779	nls_correct
(Intercept)	7.251	0.851	8.516	< 0.001	dual_task_numbers_total
tinnitus	-0.334	0.448	-0.745	0.457	dual_task_numbers_total
distressing_tinnitus	-1.057	0.462	-2.287	0.024	dual_task_numbers_total
age	-0.025	0.015	-1.673	0.097	dual_task_numbers_total
hearThresh	-0.034	0.018	-1.864	0.064	dual_task_numbers_total
DASS_A	-0.088	0.041	-2.143	0.034	dual_task_numbers_total
DASS_D	0.030	0.030	1.005	0.317	dual_task_numbers_total
DASS_S	0.032	0.028	1.121	0.264	dual_task_numbers_total

(Intercept)	30.400	1.653	18.394	< 0.001	keep_track_correct
tinnitus	-0.573	0.869	-0.660	0.511	keep_track_correct
distressing_tinnitus	-1.805	0.897	-2.013	0.046	keep_track_correct
age	-0.084	0.029	-2.893	0.004	keep_track_correct
hearThresh	-0.098	0.035	-2.772	0.006	keep_track_correct
DASS_A	-0.136	0.080	-1.697	0.092	keep_track_correct
DASS_D	-0.018	0.058	-0.315	0.753	keep_track_correct
DASS_S	0.104	0.055	1.886	0.061	keep_track_correct
(Intercept)	23.788	1.603	14.843	< 0.001	matrix_reasoning
tinnitus	0.413	0.843	0.490	0.625	matrix_reasoning
distressing_tinnitus	-2.489	0.870	-2.863	0.005	matrix_reasoning
age	-0.045	0.028	-1.600	0.112	matrix_reasoning
hearThresh	-0.063	0.034	-1.828	0.070	matrix_reasoning
DASS_A	-0.158	0.077	-2.046	0.043	matrix_reasoning
DASS_D	0.023	0.056	0.415	0.679	matrix_reasoning
DASS_S	0.061	0.054	1.131	0.260	matrix_reasoning
(Intercept)	-1.568	6.575	-0.238	0.812	RIF_difference
tinnitus	-4.548	3.459	-1.315	0.191	RIF_difference
distressing_tinnitus	-6.610	3.568	-1.853	0.066	RIF_difference
age	0.110	0.115	0.951	0.343	RIF_difference
hearThresh	-0.086	0.141	-0.609	0.544	RIF_difference
DASS_A	-0.308	0.318	-0.971	0.333	RIF_difference
DASS_D	-0.258	0.231	-1.117	0.266	RIF_difference
DASS_S	0.365	0.220	1.663	0.099	RIF_difference
(Intercept)	2.623	1.706	1.537	0.127	rule_switch_errors
tinnitus	-0.916	0.898	-1.020	0.309	rule_switch_errors
distressing_tinnitus	0.156	0.926	0.169	0.866	rule_switch_errors
age	0.020	0.030	0.653	0.515	rule_switch_errors
hearThresh	-0.012	0.037	-0.321	0.749	rule_switch_errors
DASS_A	0.067	0.082	0.815	0.416	rule_switch_errors
DASS_D	-0.055	0.060	-0.912	0.364	rule_switch_errors
DASS_S	0.053	0.057	0.922	0.358	rule_switch_errors
(Intercept)	0.258	0.144	1.788	0.076	rule_switch_cost
tinnitus	-0.015	0.076	-0.192	0.848	rule_switch_cost
distressing_tinnitus	0.016	0.078	0.206	0.837	rule_switch_cost
age	0.003	0.003	1.346	0.181	rule_switch_cost
hearThresh	0.000	0.003	-0.156	0.876	rule_switch_cost
DASS_A	0.004	0.007	0.528	0.598	rule_switch_cost
DASS_D	-0.001	0.005	-0.126	0.900	rule_switch_cost

DASS_S	0.007	0.005	1.537	0.127	rule_switch_cost
(Intercept)	2.620	1.358	1.929	0.056	stroop_errors
tinnitus	-0.043	0.715	-0.060	0.952	stroop_errors
distressing_tinnitus	0.529	0.737	0.717	0.474	stroop_errors
age	-0.021	0.024	-0.897	0.371	stroop_errors
hearThresh	0.037	0.029	1.265	0.208	stroop_errors
DASS_A	0.060	0.066	0.911	0.364	stroop_errors
DASS_D	-0.023	0.048	-0.472	0.638	stroop_errors
DASS_S	-0.005	0.045	-0.101	0.920	stroop_errors
(Intercept)	0.700	0.152	4.605	< 0.001	stroop_incongruent_RT
tinnitus	0.003	0.080	0.034	0.973	stroop_incongruent_RT
distressing_tinnitus	-0.005	0.083	-0.056	0.955	stroop_incongruent_RT
age	0.004	0.003	1.669	0.097	stroop_incongruent_RT
hearThresh	0.009	0.003	2.853	0.005	stroop_incongruent_RT
DASS_A	0.001	0.007	0.202	0.840	stroop_incongruent_RT
DASS_D	-0.003	0.005	-0.499	0.619	stroop_incongruent_RT
DASS_S	0.008	0.005	1.624	0.107	stroop_incongruent_RT

Table 4.2 details the predictors in that reached conventional levels of statistical significance (i.e. $p < 0.05$).

Table 4—2. Statistically significant predictors from general linear models predicting objective cognitive measures with tinnitus group, age, hearing threshold, depression, anxiety, and stress as predictors.

term	estimate	std.error	statistic	p.value	partial_beta_rank	task
age_scaled	-0.342	0.106	-3.227	0.002	1	nls_correct
distressing_tinnitus	-0.498	0.218	-2.287	0.024	1	dual_task_numbers_to tal
DASS_A_scaled	-0.238	0.111	-2.143	0.034	2	dual_task_numbers_to tal
distressing_tinnitus	-0.411	0.204	-2.013	0.046	1	keep_track_correct
age_scaled	-0.284	0.098	-2.893	0.004	2	keep_track_correct
hearThresh_scaled	-0.255	0.092	-2.772	0.006	3	keep_track_correct
distressing_tinnitus	-0.615	0.215	-2.863	0.005	1	matrix_reasoning
DASS_A_scaled	-0.224	0.110	-2.046	0.043	2	matrix_reasoning
hearThresh_scaled	0.279	0.098	2.853	0.005	1	stroop_incongruent_rt

Table 4.3 shows the GLMs from Table 4.1 with the addition of standardised partial regression coefficients or beta weights (β), which enable comparison of all beta weights as standard deviation units (Kline, 2016). Beta weights were then ranked by their relative size to create an ordinal measure of the importance of each predictor within the model as defined by its effect size.

Table 4—3. Standardised general linear model results predicting objective cognitive measures with tinnitus group, age, hearing threshold, depression, anxiety, and stress as predictors.

term	estimate	std.error	statistic	p.value	partial_beta_rank	task
tinnitus	0.056	0.214	0.260	0.796	6	nls_correct
distressing_tinnitus	-0.258	0.221	-1.170	0.244	2	nls_correct
age_scaled	-0.342	0.106	-3.227	0.002	1	nls_correct
hearThresh_scaled	-0.025	0.100	-0.252	0.801	4	nls_correct
DASS_A_scaled	-0.100	0.113	-0.887	0.377	3	nls_correct
DASS_D_scaled	0.084	0.112	0.753	0.453	7	nls_correct
DASS_S_scaled	0.033	0.119	0.281	0.779	5	nls_correct
tinnitus	-0.157	0.211	-0.745	0.457	5	dual_task_numbers_total
distressing_tinnitus	-0.498	0.218	-2.287	0.024	1	dual_task_numbers_total
age_scaled	-0.175	0.105	-1.673	0.097	4	dual_task_numbers_total
hearThresh_scaled	-0.183	0.098	-1.864	0.064	3	dual_task_numbers_total
DASS_A_scaled	-0.238	0.111	-2.143	0.034	2	dual_task_numbers_total
DASS_D_scaled	0.111	0.110	1.005	0.317	6	dual_task_numbers_total
DASS_S_scaled	0.132	0.117	1.121	0.264	7	dual_task_numbers_total
tinnitus	-0.130	0.198	-0.660	0.511	5	keep_track_correct
distressing_tinnitus	-0.411	0.204	-2.013	0.046	1	keep_track_correct
age_scaled	-0.284	0.098	-2.893	0.004	2	keep_track_correct
hearThresh_scaled	-0.255	0.092	-2.772	0.006	3	keep_track_correct
DASS_A_scaled	-0.177	0.104	-1.697	0.092	4	keep_track_correct
DASS_D_scaled	-0.033	0.103	-0.315	0.753	6	keep_track_correct
DASS_S_scaled	0.208	0.110	1.886	0.061	7	keep_track_correct
tinnitus	0.102	0.208	0.490	0.625	6	matrix_reasoning
distressing_tinnitus	-0.615	0.215	-2.863	0.005	1	matrix_reasoning
age_scaled	-0.165	0.103	-1.600	0.112	4	matrix_reasoning

hearThresh_scaled	-0.177	0.097	-1.828	0.070	3	matrix_reasoning
DASS_A_scaled	-0.224	0.110	-2.046	0.043	2	matrix_reasoning
DASS_D_scaled	0.045	0.109	0.415	0.679	5	matrix_reasoning
DASS_S_scaled	0.131	0.116	1.131	0.260	7	matrix_reasoning
tinnitus	-0.294	0.224	-1.315	0.191	2	RIF_difference
distressing_tinnitus	-0.427	0.231	-1.853	0.066	1	RIF_difference
age_scaled	0.105	0.111	0.951	0.343	6	RIF_difference
hearThresh_scaled	-0.063	0.104	-0.609	0.544	5	RIF_difference
DASS_A_scaled	-0.114	0.118	-0.971	0.333	4	RIF_difference
DASS_D_scaled	-0.131	0.117	-1.117	0.266	3	RIF_difference
DASS_S_scaled	0.207	0.124	1.663	0.099	7	RIF_difference
tinnitus	-0.230	0.225	-1.020	0.309	7	rule_switch_errors
distressing_tinnitus	0.039	0.232	0.169	0.866	4	rule_switch_errors
age_scaled	0.073	0.112	0.653	0.515	3	rule_switch_errors
hearThresh_scaled	-0.034	0.105	-0.321	0.749	5	rule_switch_errors
DASS_A_scaled	0.097	0.119	0.815	0.416	2	rule_switch_errors
DASS_D_scaled	-0.107	0.118	-0.912	0.364	6	rule_switch_errors
DASS_S_scaled	0.116	0.125	0.922	0.358	1	rule_switch_errors
tinnitus	-0.043	0.223	-0.192	0.848	7	rule_switch_cost
distressing_tinnitus	0.047	0.230	0.206	0.837	4	rule_switch_cost
age_scaled	0.149	0.110	1.346	0.181	2	rule_switch_cost
hearThresh_scaled	-0.016	0.104	-0.156	0.876	6	rule_switch_cost
DASS_A_scaled	0.062	0.117	0.528	0.598	3	rule_switch_cost
DASS_D_scaled	-0.015	0.117	-0.126	0.900	5	rule_switch_cost
DASS_S_scaled	0.191	0.124	1.537	0.127	1	rule_switch_cost
tinnitus	-0.014	0.227	-0.060	0.952	5	stroop_errors
distressing_tinnitus	0.168	0.234	0.717	0.474	1	stroop_errors
age_scaled	-0.101	0.112	-0.897	0.371	7	stroop_errors
hearThresh_scaled	0.134	0.106	1.265	0.208	2	stroop_errors
DASS_A_scaled	0.109	0.119	0.911	0.364	3	stroop_errors
DASS_D_scaled	-0.056	0.118	-0.472	0.638	6	stroop_errors
DASS_S_scaled	-0.013	0.126	-0.101	0.920	4	stroop_errors
tinnitus	0.007	0.210	0.034	0.973	5	stroop_incongruent_rt
distressing_tinnitus	-0.012	0.217	-0.056	0.955	6	stroop_incongruent_rt
age_scaled	0.174	0.104	1.669	0.097	3	stroop_incongruent_rt
hearThresh_scaled	0.279	0.098	2.853	0.005	1	stroop_incongruent_rt

DASS_A_scaled	0.022	0.110	0.202	0.840	4	stroop_incongruent _rt
DASS_D_scaled	-0.055	0.110	-0.499	0.619	7	stroop_incongruent _rt
DASS_S_scaled	0.190	0.117	1.624	0.107	2	stroop_incongruent _rt

Table 4.4 displays the beta weights for the categorical variable denoting membership of the Distressing Tinnitus group, as well the beta weight ranking within each model. It shows that that partial beta weights for Distressing Tinnitus group were ranked first or second for every outcome measure except switch-cost, switch errors, and Stroop (response time). This can be interpreted that having distressing tinnitus was the variable associated with the largest effect size (poorer performance), after controlling for all other variables within the model. The relative effect size ranged from -.26 (number-letter sequencing) to -.62 (matrix reasoning) in standard deviation units when compared to the No Tinnitus group. Partial beta weights for switch-cost, switch errors, and Stroop (response time) were all essentially 0.

Table 4—4. Ranking of Distressing Tinnitus group partial beta coefficient size within general linear models of objective cognitive measures with tinnitus group, age, hearing threshold, depression, anxiety, and stress as predictors.

task	partial_beta_rank	estimate
matrix_reasoning	1	-0.62
dual_task_numbers_total	1	-0.50
RIF_difference	1	-0.43
keep_track_correct	1	-0.41
stroop_errors	1	0.17
nls_correct	2	-0.26
rule_switch_errors	4	0.04
rule_switch_cost	4	0.05
stroop_incongruent_rt	6	-0.01

Table 4.5 displays the beta weights for the categorical variable denoting membership of the Non-Distressing Tinnitus group. It shows that having tinnitus is only ranked as the second beta weight within any model (RIF effect), with an estimate of -.29 standard deviation units relative to the No Tinnitus group. Small negative estimates of -.16 and -

.13 were associated with dual task and keep task performance, however Non-Distressing Tinnitus group membership was the fifth ranked variable within each of these models. Partial beta weights for Stroop task (response time and errors), number-letter sequencing, and switch-cost were essentially 0. Interestingly, after controlling for *a priori* confounding variables, Non-Distressing Tinnitus group membership was associated with better performance in the matrix reasoning task and rule switching task (errors), with estimates of 0.1 and -.23 standard deviation units respectively.

Table 4—5. Ranking of Non-Distressing Tinnitus group partial beta coefficient size within general linear models of objective cognitive measures with tinnitus group, age, hearing threshold, depression, anxiety, and stress as predictors.

task	partial_beta_rank	estimate
RIF_difference	2	-0.29
dual_task_numbers_total	5	-0.16
keep_track_correct	5	-0.13
stroop_errors	5	-0.01
stroop_incongruent_rt	5	0.01
nls_correct	6	0.06
matrix_reasoning	6	0.10
rule_switch_errors	7	-0.23
rule_switch_cost	7	-0.04

Table 4.6 details the first and second ranked predictor variables for all objective cognitive measure GLMs. This view of the data highlights the frequent estimate ranking of the Distressing Tinnitus group membership variable. It highlights Non-Distressing and Distressing Tinnitus group membership as the being the largest predictors for the RIF effect model. Table 4.6 also highlights the varying importance of other *a priori* identified confounding variables in objective cognitive measure task performance. These include anxiety (dual task, matrix reasoning, switch errors), age (keep track, switch-cost), stress (switch-cost, switch errors, Stroop response time), and hearing thresholds (Stroop errors and response time).

Table 4—6. First and second ranked partial beta coefficients size within general linear models of objective cognitive measures with tinnitus group, age, hearing threshold, depression, anxiety, and stress as predictors.

task	term	partial_beta_rank	estimate
------	------	-------------------	----------

dual_task_numbers_total	distressing_tinnitus	1	-0.498
dual_task_numbers_total	DASS_A_scaled	2	-0.238
keep_track_correct	distressing_tinnitus	1	-0.411
keep_track_correct	age_scaled	2	-0.284
matrix_reasoning	distressing_tinnitus	1	-0.615
matrix_reasoning	DASS_A_scaled	2	-0.224
nls_correct	age_scaled	1	-0.342
nls_correct	distressing_tinnitus	2	-0.258
RIF_difference	distressing_tinnitus	1	-0.427
RIF_difference	tinnitus	2	-0.294
rule_switch_cost	DASS_S_scaled	1	0.191
rule_switch_cost	age_scaled	2	0.149
rule_switch_errors	DASS_S_scaled	1	0.116
rule_switch_errors	DASS_A_scaled	2	0.097
stroop_errors	distressing_tinnitus	1	0.168
stroop_errors	hearThresh_scaled	2	0.134
stroop_incongruent_rt	hearThresh_scaled	1	0.279
stroop_incongruent_rt	DASS_S_scaled	2	0.190

4.3.2 Bayesian GLMs

Table 4—7. Bayesian general linear model predicting objective cognitive measures for Distressing Tinnitus group with tinnitus group, age, hearing threshold, depression, anxiety, and stress as predictors, standardised beta weights by rank with probability that beta is less than, or greater than zero.

task	partial_beta_rank	estimate	prob_beta_less_than_0	prob_beta_more_than_0
matrix_reasoning	1	-0.61	1.00	0.00
dual_task_words_correct	1	-0.50	0.99	0.01
RIF_effect	1	-0.42	0.97	0.03
keep_track_correct	1	-0.41	0.97	0.03
nls_correct	2	-0.25	0.88	0.12
stroop_rt	6	-0.01	0.52	0.48
rule_errors	4	0.04	0.42	0.58
switch_cost	4	0.05	0.41	0.59
stroop_errors	1	0.17	0.24	0.76

Tables 4.7 and 4.8 display the partial beta weights for the categorical variables denoting membership of the Distressing Tinnitus and Non-Distressing Tinnitus groups. The resulting partial beta weights from the Bayesian GLMs and subsequent rankings within the model are identical to those of the frequentist GLMs (Tables 4.4 and 4.5). This is

typical when weakly informative priors are used (Kruschke, 2010; Muth et al., 2018) and provides validation of Bayesian GLM parameter estimation in the specified models.

However, a crucial distinction between the two regression methodologies is that the Bayesian GLMs permit sampling from the posterior distribution of each parameter within the model (Kruschke, 2010; McElreath, 2016). This enables a probabilistic estimation that a predictor has a partial beta weight that is not zero, enabling an estimate of how likely it is that being in the Non-Distressing Tinnitus or Distressing Tinnitus groups per se is associated with poorer performance on the objective cognitive measures after controlling for confounding variables.

For the Distressing Tinnitus group parameter (Table 4.7), probability of none-zero parameter estimates within each objective cognitive performance GLM ranged from virtual certainty (1) to uncertainty (.52). Being in the Distressing Tinnitus group was virtually certain to be negatively associated with matrix reasoning task performance after controlling for other variables with probabilities of 1 and .99 respectively; the most likely estimate was -.61 and -.5 standard deviation units poorer respectively. Distressing Tinnitus group membership was .97 probable to be negatively associated with the keep track task and RIF effect after controlling for other variables, with the most likely estimates being -.42 and -.41 standard deviation units respectively. The most likely estimate for Distressing Tinnitus group membership being negatively associated with the number-letter sequencing task was -.25 standard deviation units; the probability that this was none-zero was .88 after controlling for other variables. The most likely estimate for Distressing Tinnitus group membership being positively associated with the Stroop (errors) was .17 standard deviation units; the probability that this was none-zero was .76 after controlling for other variables. None-zero probability estimates indicate that that parameter estimates for Stroop (response time) and rule-switching tasks (switch-cost and errors) were very uncertain, ranging from .52 to .59.

For the Non-Distressing Tinnitus group parameter (Table 4.8), probability of none-zero parameter estimates was generally lower for each task than for the

Distressing Tinnitus group parameters. The highest probability was for the RIF effect (.91) with the most likely estimate being a negative association of -.29 standard deviation units compared to the No Tinnitus group after controlling for other variables. The probability that the Non-Distressing Tinnitus group membership parameter is negatively associated with dual task performance is .77, with the most likely estimate being -.16 after controlling for other variables. None-zero probability estimates indicate that that parameter estimates for Stroop (response time and errors) and rule-switching tasks (switch-cost) were very uncertain, ranging from .51 to .58. The probability that Non-Distressing Tinnitus group membership is positively associated with matrix reasoning performance compared to No Tinnitus group after controlling for other variables had a probability of .69, with the most likely estimate being .1 standard deviation units after controlling for other variables. The probability that Non-Distressing Tinnitus group membership is positively associated with errors on the rule switching task was .84, with the most likely estimate being -.23 standard deviation units compared to the No Tinnitus group. The most likely estimate for the number-letter sequencing task was 0.05 with a .6 probability.

Table 4—8. Bayesian general linear regression predicting objective cognitive measures for Non-Distressing Tinnitus group with tinnitus group, age, hearing threshold, depression, anxiety, and stress as predictors, standardised beta weights by rank with probability that beta is less than, or greater than zero

task	partial_beta_rank	estimate	prob_beta_less_than_0	prob_beta_more_than_0
RIF_effect	2	-0.29	0.91	0.09
rule_errors	7	-0.23	0.84	0.16
dual_task_words_correct	5	-0.16	0.77	0.23
keep_track_correct	5	-0.13	0.75	0.25
switch_cost	7	-0.04	0.58	0.42
stroop_errors	5	-0.02	0.53	0.47
stroop_rt	5	0.01	0.49	0.51
nls_correct	6	0.05	0.40	0.60
matrix_reasoning	6	0.10	0.31	0.69

4.3.3 Summary of statistical inference

4.3.3.1 No Tinnitus group and Distressing Tinnitus group

Frequentist and Bayesian GLMs showed that Distressing Tinnitus was associated with poorer task performance, and had the first or second largest parameter estimate for all tasks after controlling for *a priori* confounding variables, with the exception of Stroop response time, switch-cost, and rule-switching tasks. Sampling from the posterior distribution of Bayesian GLMs then demonstrated the probability that Distressing Tinnitus per se is associated with negative task performance: negative performance associations were extremely probable with matrix reasoning (1), dual task (.99), RIF effect (.97), and keep track task (.97); negative associations were less likely, but still highly probable for the number-letter sequencing task (.86) and Stroop errors (.75). Models for Stroop response time, switch-cost, and rule-switching task errors demonstrated near total uncertainty about the direction of parameter estimates.

4.3.3.2 No Tinnitus group and Non-Distressing Tinnitus group

There may be subtle decrements associated with Non-Distressing Tinnitus per se for some Updating and Shifting tasks when compared to the No Tinnitus group. However, the Non-Distressing Tinnitus group likely had improved task performance compared to the No Tinnitus group for Shifting and fluid ability tasks after accounting for the influence of other variables.

Frequentist and Bayesian GLMs revealed that Non-Distressing Tinnitus was associated with negative task performance for the keep track, and dual tasks after controlling for *a priori* confounding variables; however the effects were small (-.13 and -.16 standard deviation units respectively). Non-Distressing Tinnitus had the second largest association for the RIF effect and a parameter estimate of -.29 after controlling for *a priori* confounding variables. No association after controlling for other variables was seen in other objective cognitive tasks including switch-cost, Stroop response time, Stroop errors, or number-letter sequencing. However, being in the Non-Distressing Tinnitus group was associated with improved task performance compared to the Control

group on the rule-switching task (errors) and matrix reasoning task with parameter estimates of -.23 and .1 respectively.

Sampling from the posterior distribution of Bayesian GLMs then demonstrated the probability that Non-Distressing Tinnitus per se is associated with either positive or negative task performance: negative associations with the RIF effect were highly likely (.91), while the probability of negative associations were smaller but still likely with the keep track (.75) and dual task (.77); positive task performance associations were also likely with rule-switching errors (.84) and matrix reasoning tasks (.69). Models for Stroop response time, Stroop errors, and switch-cost demonstrated near total uncertainty about the direction of parameter estimates.

4.3.3.3 Relative importance of variables

Table 4—9. Table detailing the ratio of beta weight ratios for each predictor in each GLM relative to the distressing tinnitus predictor.

task	IV	distressing_vector	estimate	greater_beta_ratio	less_beta_ratio
matrix_reasoning	age_scaled	-0.61	-0.16	3.81	0.26
dual_task_words_correct	age_scaled	-0.50	-0.17	2.94	0.34
keep_track_correct	age_scaled	-0.41	-0.28	1.46	0.68
nls_correct	age_scaled	-0.25	-0.34	0.74	1.36
rule_errors	age_scaled	0.04	0.07	0.57	1.75
switch_cost	age_scaled	0.05	0.15	0.33	3.00
stroop_rt	age_scaled	-0.01	0.17	-0.06	-17.00
stroop_errors	age_scaled	0.17	-0.10	-1.70	-0.59
RIF_effect	age_scaled	-0.42	0.11	-3.82	-0.26
RIF_effect	DASS_A_scaled	-0.42	-0.11	3.82	0.26
matrix_reasoning	DASS_A_scaled	-0.61	-0.22	2.77	0.36
nls_correct	DASS_A_scaled	-0.25	-0.10	2.50	0.40
keep_track_correct	DASS_A_scaled	-0.41	-0.18	2.28	0.44
dual_task_words_correct	DASS_A_scaled	-0.50	-0.24	2.08	0.48
stroop_errors	DASS_A_scaled	0.17	0.11	1.55	0.65
switch_cost	DASS_A_scaled	0.05	0.06	0.83	1.20
rule_errors	DASS_A_scaled	0.04	0.10	0.40	2.50
stroop_rt	DASS_A_scaled	-0.01	0.02	-0.50	-2.00
keep_track_correct	DASS_D_scaled	-0.41	-0.03	13.67	0.07
RIF_effect	DASS_D_scaled	-0.42	-0.13	3.23	0.31
stroop_rt	DASS_D_scaled	-0.01	-0.05	0.20	5.00
rule_errors	DASS_D_scaled	0.04	-0.11	-0.36	-2.75

nls_correct	DASS_D_scaled	-0.25	0.09	-2.78	-0.36
stroop_errors	DASS_D_scaled	0.17	-0.05	-3.40	-0.29
dual_task_words_correct	DASS_D_scaled	-0.50	0.11	-4.55	-0.22
switch_cost	DASS_D_scaled	0.05	-0.01	-5.00	-0.20
matrix_reasoning	DASS_D_scaled	-0.61	0.04	-15.25	-0.07
rule_errors	DASS_S_scaled	0.04	0.11	0.36	2.75
switch_cost	DASS_S_scaled	0.05	0.19	0.26	3.80
stroop_rt	DASS_S_scaled	-0.01	0.19	-0.05	-19.00
keep_track_correct	DASS_S_scaled	-0.41	0.21	-1.95	-0.51
RIF_effect	DASS_S_scaled	-0.42	0.21	-2.00	-0.50
dual_task_words_correct	DASS_S_scaled	-0.50	0.13	-3.85	-0.26
matrix_reasoning	DASS_S_scaled	-0.61	0.13	-4.69	-0.21
nls_correct	DASS_S_scaled	-0.25	0.03	-8.33	-0.12
stroop_errors	DASS_S_scaled	0.17	-0.02	-8.50	-0.12
nls_correct	hearThresh_scaled	-0.25	-0.03	8.33	0.12
RIF_effect	hearThresh_scaled	-0.42	-0.06	7.00	0.14
matrix_reasoning	hearThresh_scaled	-0.61	-0.18	3.39	0.30
dual_task_words_correct	hearThresh_scaled	-0.50	-0.19	2.63	0.38
keep_track_correct	hearThresh_scaled	-0.41	-0.26	1.58	0.63
stroop_errors	hearThresh_scaled	0.17	0.13	1.31	0.76
stroop_rt	hearThresh_scaled	-0.01	0.28	-0.04	-28.00
rule_errors	hearThresh_scaled	0.04	-0.03	-1.33	-0.75
switch_cost	hearThresh_scaled	0.05	-0.01	-5.00	-0.20
keep_track_correct	tinnitus	-0.41	-0.13	3.15	0.32
dual_task_words_correct	tinnitus	-0.50	-0.16	3.12	0.32
RIF_effect	tinnitus	-0.42	-0.29	1.45	0.69
rule_errors	tinnitus	0.04	-0.23	-0.17	-5.75
stroop_rt	tinnitus	-0.01	0.01	-1.00	-1.00
switch_cost	tinnitus	0.05	-0.04	-1.25	-0.80
nls_correct	tinnitus	-0.25	0.05	-5.00	-0.20
matrix_reasoning	tinnitus	-0.61	0.10	-6.10	-0.16
stroop_errors	tinnitus	0.17	-0.02	-8.50	-0.12

Because variables within GLMs were standardised, comparisons can be made concerning the relative size of the variables within each model (Kline, 2016, p. 32). Therefore, beta weight ratios for Distressing Tinnitus and other theoretically important variables within

each model quantify its relative importance via standardised effect sizes (Table 4.9). Ratios of beta weights indicate that having distressing tinnitus per se is the most important predictor across a variety of cognitive tasks after controlling for other variables:

- Distressing Tinnitus was 5 to 8.5 times as important as Non-Distressing Tinnitus or stress for predicting poorer matrix reasoning performance, increased Stroop errors, and poorer number-letter sequencing performance.
- Distressing Tinnitus was 1.5 to 4 times as important as age for predicting worse matrix reasoning performance, increased Stroop errors, and poorer keep track and dual-task performance.
- Conversely, age was 3 times as important as Distressing Tinnitus for predicting increased switch-cost response times.

4.3.3.4 Notes on inferential GLM analyses

Challenges facing NHST inference were highlighted in frequentist GLMs, where few models were found to be statistically significant. Decisions regarding the importance of variables based on solely on statistical significance is not uncommon (McElreath, 2016), and it is possible that similar inferential decisions may have limited conclusions reached in previous research on this topic in the few studies that modelled confounding factors (Table 4.2).

The analysis presented in this chapter demonstrated that frequentist and Bayesian methods substantively agreed on parameter estimates, while advancing what can be correctly inferred about them (i.e. how likely it is that having Non-Distressing Tinnitus or Distressing Tinnitus is independently negatively associated with poorer objective cognitive task performance). However, it should be remembered that both methodologies are merely inferential tools. Decisions concerning evidential quality should be based on all available evidence and theoretical justifications. Hypotheses in this study were based on Clarke et al. (2020), which anticipated executive functioning differences

between No Tinnitus, Non-Distressing Tinnitus, and Distressing Tinnitus groups. Figure 3.12 provides a salient demonstration of this hypothesis.

This analysis also highlights technical issues concerning typical analysis of response time data. Among these are known challenges in analysing individual differences in tasks such as the Stroop due to low inter-individual variability in the effect, that may be exacerbated by group score aggregation (Hedge et al., 2018; Whelan, 2008). Whelan (2008) notes that common statistical tests that use aggregated group sample means may have a poor power to detect differences between experimental groups. Analysis of Stroop and rule-switching data indicate that this is likely to be the case, as Bayesian GLMs demonstrate that none-zero probability parameter estimates were highly uncertain. This suggests that it would be beneficial to undertake further modelling of the Stroop and rule-switching response time data using more powerful analytic strategies such as hierarchical linear models (Section 4.3.5).

4.3.4 Retrieval-induced forgetting (RIF) sub-analysis

The aforementioned analysis found that, as hypothesised, the RIF effect was present in the No Tinnitus group, but absent in the Non-Distressing Tinnitus and Distressing Tinnitus groups. It is reasonable to attribute the RIF effect as a measure of Inhibition (Storm and Levy, 2012), and therefore these results may indicate reduced inhibitory control in the Non-Distressing Tinnitus and Distressing Tinnitus groups. However, each category of word within the retrieval-practise paradigm can be scrutinised between the study groups to ascertain the relative categorical contribution to the RIF effect (Table 4.10). This provides further insight into the relationship between long-term memory and executive functioning that is driving the RIF effect or its absence.

Rp+ are categories and words that were practised, Rp- represent categories that were practised but words that were not, and Np represent both categories and words that were not practised. The mean proportions of word categories recalled correctly shows that the No Tinnitus group exhibited the RIF effect as expected through recalling more Rp+ items than Np items. However, the Non-Distressing Tinnitus group recalled an

equal proportion of Rp+ and Np, while the Distressing Tinnitus group recalled more Np than Rp+ items.

Table 4—10. Mean proportions of correct recalls on Rp+, Rp-, Np, total, and RIF effect score for each group.

THI_cat	Rp+ mean	Rp- mean	Np mean	RIF total mean	RIF Diff mean
control	87	51	55	62	3
tinnitus	88	56	56	64	0
distressing tinnitus	80	48	46	55	-2

To further investigate the association of RIF effects to group membership per se, GLMs were made for the various retrieval-practise categories with predictors used in GLMs for other objective cognitive measures including depression, anxiety, stress, hearing thresholds, and age. Additionally, RRS was also added to the models as rumination was hypothesised to have mechanistic significance in between-group differences in the RIF effect (Whitmer and Banich, 2010; Whitmer and Gotlib, 2013). A theoretical account that posits rumination to be linked to impaired inhibitory control. Investigation of rumination using retrieval-practise paradigms has been performed, with previous work having reported that high trait ruminators exhibit a greater recall of unpractised exemplars from practised categories (Rp-); this has been taken as evidence that ruminators have difficulties automatically inhibiting information that is irrelevant but strongly associated with the relevant information (Whitmer and Banich, 2010; Whitmer and Gotlib, 2013).

While no predictor was statistically significant according to frequentist GLMs, this could be due to the model support for H0 or data insensitivity (Table 4.11). Bayesian GLM estimates are provided in Table 4.12; these substantively agree with partial betas for frequentist GLMs, but also provide non-zero probability estimates to quantify uncertainty and support for observed differences.

Table 4—11. General linear models predicting Rp+, Rp-, and Np scores from tinnitus group category, age, hearing threshold, DASS_D, DASS_A, DASS_S, and RRS scores.

term	estimate	std.error	statistic	p.value	model
(Intercept)	102.05	9.39	10.87	< 0.01	RpPlus
THI_cat2	1.22	3.36	0.36	0.72	RpPlus
THI_cat3	-4.99	3.50	-1.43	0.16	RpPlus
age	-0.19	0.12	-1.58	0.12	RpPlus
hearThresh	0.04	0.14	0.26	0.80	RpPlus
DASS_A	-0.09	0.31	-0.29	0.77	RpPlus
DASS_D	0.07	0.23	0.30	0.77	RpPlus
DASS_S	-0.16	0.22	-0.72	0.48	RpPlus
RRS	-0.09	0.14	-0.65	0.52	RpPlus
(Intercept)	68.23	10.22	6.68	< 0.01	RpMinus
THI_cat2	7.28	3.65	1.99	0.05	RpMinus
THI_cat3	1.54	3.81	0.40	0.69	RpMinus
age	-0.24	0.13	-1.88	0.06	RpMinus
hearThresh	-0.13	0.15	-0.88	0.38	RpMinus
DASS_A	-0.28	0.34	-0.84	0.40	RpMinus
DASS_D	0.08	0.25	0.31	0.75	RpMinus
DASS_S	-0.45	0.24	-1.89	0.06	RpMinus
RRS	0.08	0.15	0.53	0.59	RpMinus
(Intercept)	71.87	10.12	7.10	< 0.01	Np
THI_cat2	2.87	3.62	0.79	0.43	Np
THI_cat3	-4.66	3.77	-1.24	0.22	Np
age	-0.16	0.13	-1.26	0.21	Np
hearThresh	-0.23	0.15	-1.55	0.12	Np
DASS_A	-0.58	0.33	-1.75	0.08	Np
DASS_D	-0.13	0.25	-0.50	0.62	Np
DASS_S	-0.05	0.24	-0.19	0.85	Np
RRS	-0.02	0.15	-0.16	0.87	Np

For Rp+, being in the Distressing Tinnitus group has the largest effect and negative association on practised items per se with -.33 of a standard deviation unit, and age had the second largest effect on the observed difference with -.18. Being in the Non-Distressing Tinnitus group has a negligible effect. For Rp-, being in the Distressing Tinnitus group had a negligible effect, while being in the Non-Distressing Tinnitus group had the largest association with the number of words recalled with an increase of .43 standard deviation units. Age again had a negative association, along with stress at -.23

and -.21 respectively. For the Np items, Distressing Tinnitus and Non-Distressing Tinnitus had the largest associations within the model, however, they were in opposite directions; being in the Distressing Tinnitus group had a negative association with -.27 standard deviation units, while being in the Non-Distressing Tinnitus group was associated with a better score than the Control group, with an average of .17 standard deviation units improved recall. Age and hearing threshold were also negatively associated with performance with partial betas of -.14 and -.16 respectively.

Table 4—12. Table showing partial beta weights and accompanying rank within each model for each retrieval-practise category.

task	IV	mean	partial_beta_rank
RIF_RpPlus	THI_cat2	0.08	7
RIF_RpPlus	THI_cat3	-0.33	1
RIF_RpPlus	age_scaled	-0.18	2
RIF_RpPlus	hearThresh_scaled	0.03	5
RIF_RpPlus	DASS_D_scaled	0.03	6
RIF_RpPlus	DASS_S_scaled	-0.09	3
RIF_RpPlus	RRS_scaled	-0.08	4
RIF_RpMinus	THI_cat2	0.43	7
RIF_RpMinus	THI_cat3	0.08	6
RIF_RpMinus	age_scaled	-0.21	2
RIF_RpMinus	hearThresh_scaled	-0.09	3
RIF_RpMinus	DASS_D_scaled	0.04	4
RIF_RpMinus	DASS_S_scaled	-0.23	1
RIF_RpMinus	RRS_scaled	0.06	5
RIF_Np	THI_cat2	0.17	7
RIF_Np	THI_cat3	-0.27	1
RIF_Np	age_scaled	-0.14	3
RIF_Np	hearThresh_scaled	-0.16	2
RIF_Np	DASS_D_scaled	-0.06	4
RIF_Np	DASS_S_scaled	-0.02	5
RIF_Np	RRS_scaled	-0.02	6

Table 4—13. Bayesian GLM with all predictors for RRS_scaled partial beta scores and none-zero probability estimates.

task	mean	partial_beta_rank	less_non_zero_prob	more_non_zero_prob
RIF_RpPlus	-0.08	4	0.74	0.26
RIF_RpMinus	0.06	5	0.30	0.70
RIF_Np	-0.02	6	0.57	0.43

Table 4—14. Bayesian GLM with all predictors for tinnitus and distressing tinnitus for each retrieval-practise category with accompanying beta weights and none-zero probabilities.

task	IV	mean	partial_beta_rank	less_non_zero_prob	more_non_zero_prob
RIF_RpPlus	tinnitus	1.18	8	0.36	0.64
RIF_RpPlus	distressing tinnitus	-5.01	1	0.93	0.07
RIF_RpMinus	tinnitus	7.26	8	0.02	0.98
RIF_RpMinus	distressing tinnitus	1.57	7	0.34	0.66
RIF_Np	tinnitus	2.79	8	0.23	0.77
RIF_Np	distressing tinnitus	-4.67	1	0.89	0.11

Partial beta weights for the RRS variable indicated that they had a negligible independent association with outcomes in Rp+, Rp-, and Np GLMs (Table 4.13). Regarding the proportion of scores that are attributable to being in the Non-Distressing Tinnitus and Distressing Tinnitus groups, frequentist and Bayesian partial regression coefficients agree that for Rp+, Non-Distressing Tinnitus and Distressing Tinnitus were associated with a 1.2% improvements and -5% decrement respectively compared to the No Tinnitus group; for Rp-, Non-Distressing Tinnitus and Distressing Tinnitus were associated with 7.3% and 1.5% improvements respectively compared to the No Tinnitus group; and for Np items, Non-Distressing Tinnitus and Distressing Tinnitus were associated with a 2.8% improvement and -4.7% decrement respectively relative to the No Tinnitus group.

As with analysis of other objective cognitive measures, the probability of none-zero parameter estimates conditional on each model were obtained for Non-Distressing Tinnitus and Distressing Tinnitus group membership (Table 4.14). Distressing Tinnitus per se was highly probably associated with poorer recall for Rp+ items (.93), while Non-

Distressing Tinnitus was more likely positively associated (.64). Non-Distressing Tinnitus group membership per se was highly probably associated with improved recall for Rp-items (.98), while Distressing Tinnitus was also associated with improved recall (.66). Non-Distressing Tinnitus group membership per se was probably associated with improved Np item recall (.77), while Distressing Tinnitus was probably associated with worse Np item recall (.89) than the No Tinnitus group.

4.3.4.1 RIF Summary

Taken together, analysis of recall within different categories of items revealed an interesting pattern of results, supporting different conclusions concerning the association between EF and Glr with people with Non-Distressing Tinnitus and Distressing Tinnitus compared to No Tinnitus participants. For the practised items (Rp+), No Tinnitus and Non-Distressing Tinnitus group participants scored similarly, while participants with Distressing Tinnitus had scores 7-8% lower. For unpractised items in practised categories items (Rp-), participants with Non-Distressing Tinnitus recalled approximately 5-8% more than the other two groups, and for unpractised items from unpractised categories (Np), No Tinnitus and Non-Distressing Tinnitus again scored similarly, while participants with Distressing Tinnitus scored 10-11% worse. This suggests that participants with Non-Distressing Tinnitus and Distressing Tinnitus have impaired inhibitory control, as they recall more Rp- items. The results also suggest that participants with Distressing Tinnitus have poorer long-term memory performance for both learning and retrieval, as they scored worse than No Tinnitus and those with Non-Distressing Tinnitus for recall of Rp+ and Np items. An alternative interpretation is that the pattern of results may reflect superior cognitive performance in the Non-Distressing Tinnitus group compared to No Tinnitus, as they scored highest for all categories of item.

As shown in the previous analysis, the primary outcome of interest for the EFIT study was the RIF effect. The SMDs between the No Tinnitus and Non-Distressing Tinnitus group are similar in size to those found in the meta-analysis by Murayama et al. (2014), which provides confidence in the RIF effect. Furthermore, this study controlled

for output interference, allowing the results to be attributable to an inhibitory-mechanistic account (Murayama et al., 2014). This breakdown shows what differential cognitive processes may be driving the absence of RIF effect in the Non-Distressing Tinnitus and Distressing Tinnitus group (i.e. poorer lower-level inhibition and Glr in Distressing Tinnitus, Poorer lower-level inhibition and similar/superior Glr in the Non-Distressing Tinnitus group).

4.3.5 Hierarchical linear models for response time tasks

Many studies investigating associations between tinnitus and cognitive processes use participant response time as an outcome measure (Amaral and Langers, 2015; Araneda et al., 2015, 2015; Das et al., 2012; Jackson et al., 2014; Li et al., 2018; Rossiter et al., 2006; Stevens et al., 2007; Trevis et al., 2016; Vanneste et al., 2016; Waechter and Brännström, 2015). Section 4.3.3 highlighted that tinnitus parameter estimates for tasks using response time as an outcome measure in the EFiT study were uncertain, suggesting that more powerful modelling methodologies are required to attribute between-group differences to having tinnitus per se. Various complications should be considered when accurately modelling response time data, these include: dependence and clustering of observations because many responses are obtained from the same individual, which can be addressed with hierarchical linear models; the presence of outliers within the data, and appropriate methods for their removal; and the non-normal distribution of data points from participant responses.

4.3.5.1 Hierarchical linear models

Hierarchical linear models (HLMs), also known as multilevel models, extend single-level linear regression by allowing the model coefficients to vary by grouping variables. This allows for observations to be nested within structures where dependency is likely to arise (e.g. pupils within schools). When modelling response time data, responses across trials are nested within participants (i.e. different responses from the same participant are recognised to be more similar, and not treated as arising independently). Generally, HLMs are more powerful than single-level models, and provide several advantages that

include: adjusting estimates for repeated sampling of units, adjusting estimates for sampling imbalances, incorporating more information than single-level models by explicitly accounting for both between- and within-group structure and variation, and improved population parameter estimates through partial-pooling and robust estimates through parameter shrinkage. In doing so, uncertainty of estimates is propagated through the model, avoiding the false certainty of pre-averaged values, while still allowing the average to be used to make predictions (McElreath, 2016; Muth et al., 2018).

4.3.5.2 Models of response time data

Methods for modelling response time data may broadly be split into mechanistic models and descriptive models (Lindeløv, 2019). Mechanistic models feature parameters that have theoretically-driven interpretations, and include sequential accumulator models such as diffusion and linear ballistic accumulator models (Brown and Heathcote, 2008; Ratcliff et al., 2016; Wagenmakers et al., 2007). Descriptive models simply approximate the skewed, long, right-tailed shape of typical response time distributions (Figures 4.4 & 4.5), and may feature parameters with varying degrees of interpretability (Lindeløv, 2019).

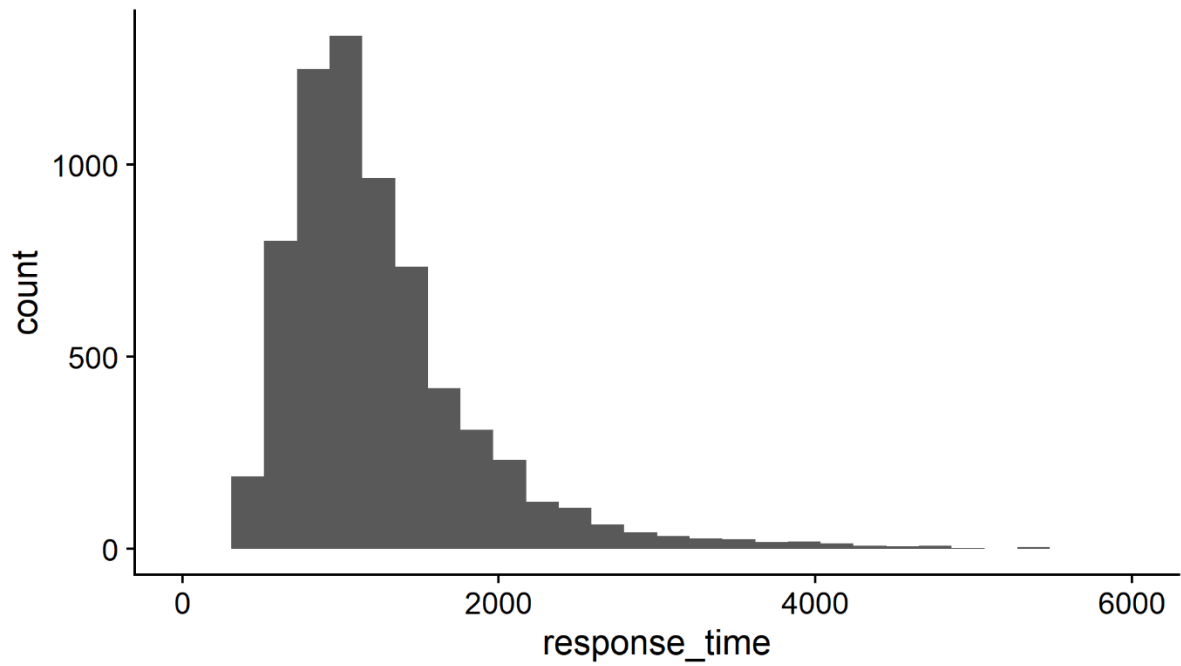


Figure 4—4. Histogram of all incongruent responses in Stroop task.

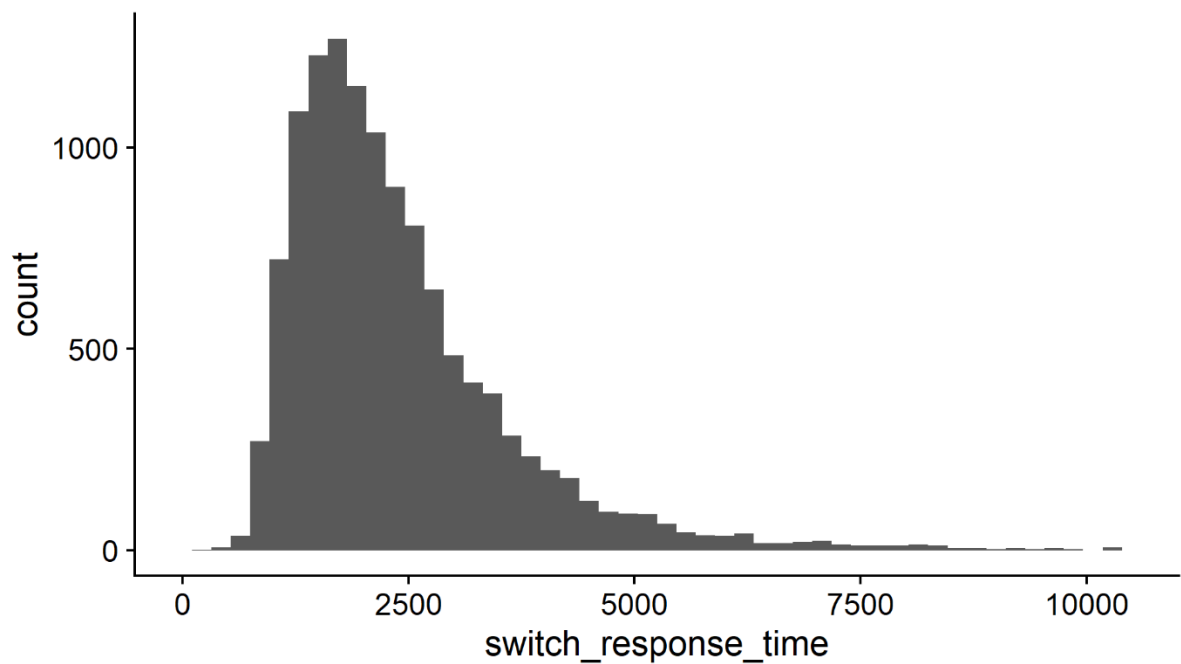


Figure 4—5. Histogram of all switch responses in rule-switching task.

Although there is no universally accepted optimal descriptive distribution for modelling response times, the normal distribution is a common option. Indeed, all studies to date that have investigated response times in participants with tinnitus have used models

implicitly incorporating the normal distribution. However, this can be problematic because symmetrical dispersion around the mean forces such models to permit approximately 5% of response times that are faster than the quickest response ever recorded (Jain et al., 2015; Lindeløv, 2019). However, various options for descriptive models of response time are available that improve on the normal distribution, including (shifted) log-normal, ex-Gaussian, and inverse-Gaussian or Wald distribution (Ratcliff, 1993).

Although there is no “best”, a key consideration of distribution is the interpretability of the parameter estimates; in this regards the shifted-log normal is advantageous. It is simply the log of a normal distribution with an extra shift parameter that identifies the time of the earliest response; and as it is simply the logarithm of a normal distribution, the parameter estimates that it provides can be interpreted as percentage increase or decreases in difference of response times (e.g. group 1 responded 5% faster than group 2). Lindeløv (2019) suggests that because working on this scale emphasises multiplicative rather than additive relationships, it better captures the intuition that an absolute value of 40 milliseconds can be insignificant over the course of 10 seconds, but crucial in shorter timescales such as those provided by simple reaction tasks.

4.3.5.3 *Outlier management*

Ratcliff (1993) describes outliers as “response times generated by processes that are not the ones being studied”. Removal of short outliers (i.e. participants pressing impossibly quickly) and long outliers (e.g. failing to press in a reasonable amount of time through inattentiveness), may be detected relatively easily. However, spurious outliers may be embedded within the distribution and impossible to detect, reducing the power of an analysis of the underlying psychological process. Various methods for outlier removal are available to increase the power of an analysis, including using the median value, trimming responses via cut-off values, or trimming response by standard deviations. Ratcliff (1993) recommends proportional trimming of response to improve analytic

power, where cut-off values are used to remove a proportion of responses in a given data set. This can range from 5-15%, which can be done over a range to check the consistency/robustness of result.

4.3.5.4 Results

All HLMs were fit using the *brms* package in R statistical programming environment (Bürkner, 2017). HLMs were fit in accordance with recommendations for proportional response time outlier removal (Ratcliff, 1993): HLMs for the Stroop task were fit using correct responses for the incongruent condition; two models were fit with cut-points at 200-2500 ms and 200-1800 ms, removing 4.8% and 14.9% of the observations respectively. The parameter estimates and conditional effects that were calculated for each model were not substantively different; therefore the 200-2500ms model results are reported as this excluded fewest observations. HLMs for the rule-switching task were fit using correct, switch-only trials (i.e. trials where a letter rule followed a number rule or vice versa); models were fit with cut-points at 200 – 5000 ms, 200-4000 ms, and 200-3500 ms, removing 4.5%, 9.6%, and 14.7% of the observations respectively. The parameter estimates and conditional effects that were calculated for each model were not substantively different, therefore the 200-5000 ms model results are reported as this excluded fewest observations. All HLMs converged successfully, as assessed by trace plots and potential scale reduction factor statistics (i.e. $\hat{R}$ values < 1.1 for all model parameters). Effective posterior sample size (ESS) was also sufficient for all HLMs with the number of effective samples achieving > 1000 for all parameters (Muth et al., 2018).

4.3.5.4.1 Stroop

To assess whether a model achieved an acceptable fit to the observed data, posterior predictive checks (PPCs) should be used within the Bayesian framework (Kruschke, 2014; McElreath, 2016). PPCs generate simulations of the data implied by the specified model. Simulations are generated by sampling from the posterior predictive distribution, which is the distribution of the outcome variable implied by the posterior distribution of model parameters; a new dataset is created with every MCMC draw, and the results of

these simulations can be compared with the observed data to indicate the fit, and thereby check that key features of the observed data have been replicated (Muth et al., 2018).

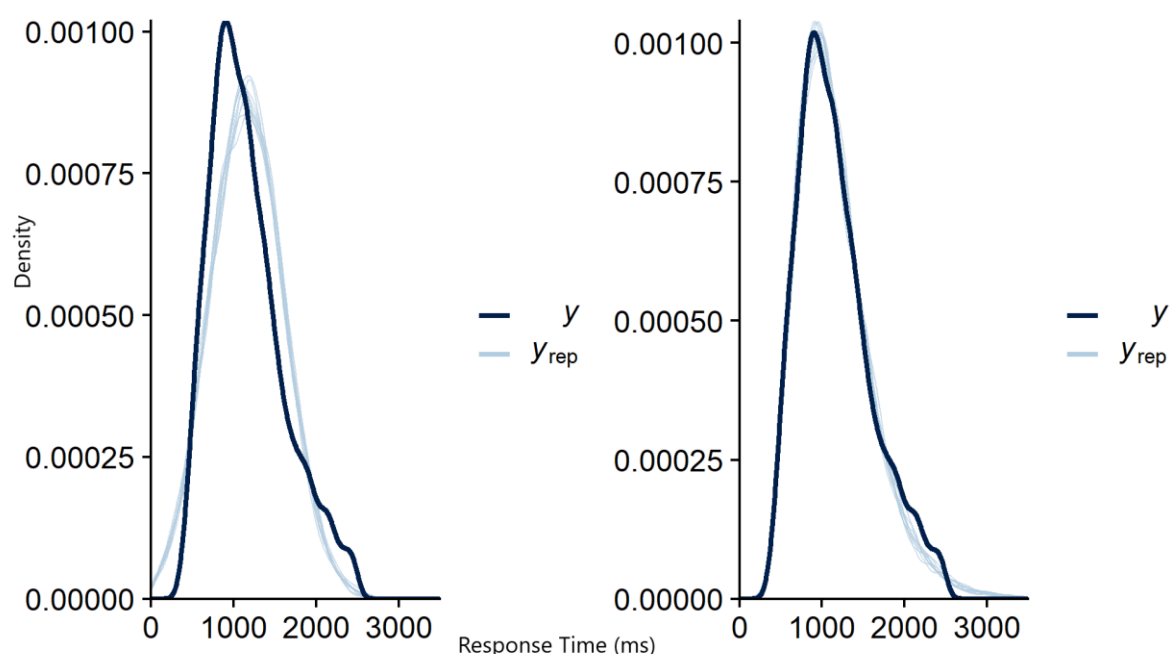


Figure 4—6. Posterior predictive checks for Stroop incongruent normal and shifted log-normal models

HLMs for the Stroop task were fit using the same parameters, but specifying both normal and shifted log-normal distributions. PPCs demonstrate that the shifted log-normal distribution in the right-sided panel provides a superior fit to the observed data compared to the normal distribution (Figure 4.6), as the simulated draws (light lines) tightly follow the observed data (dark lines).

Table 4—15. Hierarchical linear model parameter estimates for Stroop task (incongruent) response time.

term	estimate	std.error	lower	upper
b_Intercept	6.96	0.04	6.9	7.03
b_THI_cat2	-0.01	0.05	-0.09	0.07
b_THI_cat3	0.04	0.05	-0.05	0.12
b_scaleage	0.08	0.02	0.04	0.12
b_scalehearThresh	0.08	0.02	0.04	0.12
b_scaleDASS_D	-0.01	0.03	-0.05	0.03
b_scaleDASS_A	-0.01	0.03	-0.05	0.04

b_scaleDASS_S	0.03	0.03	-0.01	0.08
sd_participant__Intercept	0.21	0.01	0.19	0.24
sigma	0.31	0.01	0.3	0.32
ndt	21.49	15.29	1.87	50.55

Group-level parameter estimates for the incongruent Stroop model are displayed in Table 4.15. They show that having distressing tinnitus was most probably associated with 4% slow response times than the No-Tinnitus group, while having non-distressing tinnitus was most probably associated with 1% faster responses than the No Tinnitus group after controlling for other variables. The conditional effects for various groups are shown in Figure 4.7.

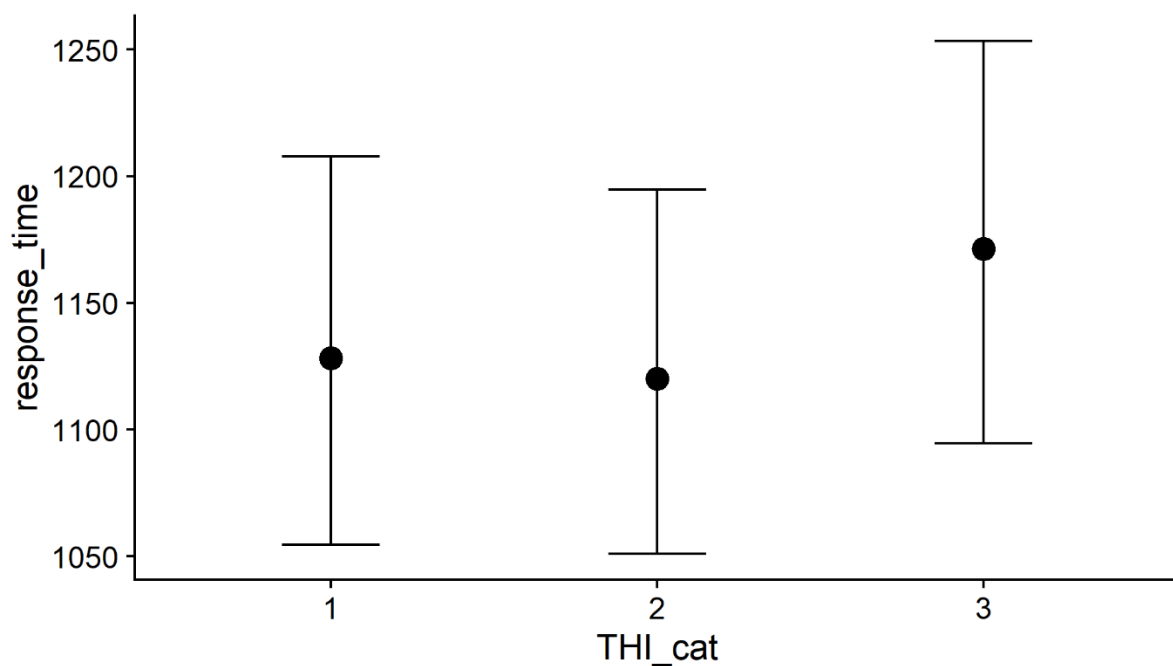


Figure 4—7. HLM parameter estimates for each tinnitus groups predicting incongruent Stroop task response times (milliseconds). 1 = No Tinnitus, 2 = Non-Distressing Tinnitus, 3 = Distressing Tinnitus. THI = Tinnitus Handicap Inventory.

Participant age, and hearing thresholds (Figures 4.8 & 4.9) were associated with increased response time (most probably 8%); stress (Figure 4.12) scores were also associated with an increase in response times (most probably 3%). However, higher depression, and anxiety (Figures 4.10 & 4.11) scores were most likely associated with a decrease in response times (most probably -1%).

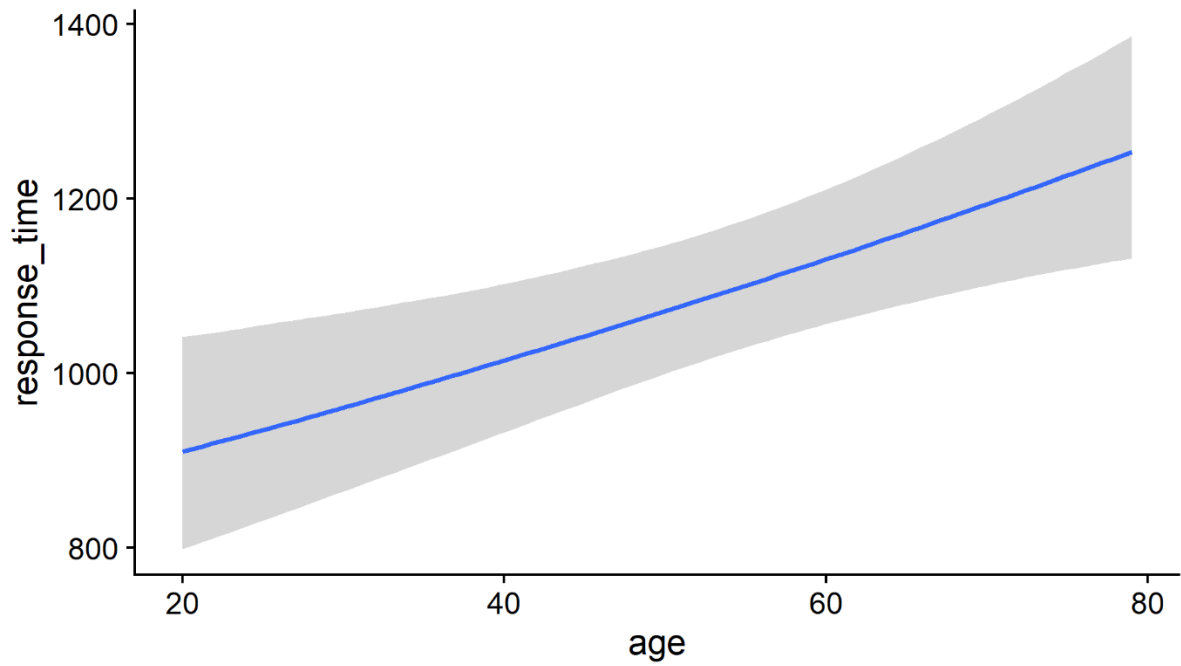


Figure 4—8. HLM parameter estimates for age predicting incongruent Stroop task response times (milliseconds) with accompany uncertainty boundaries.

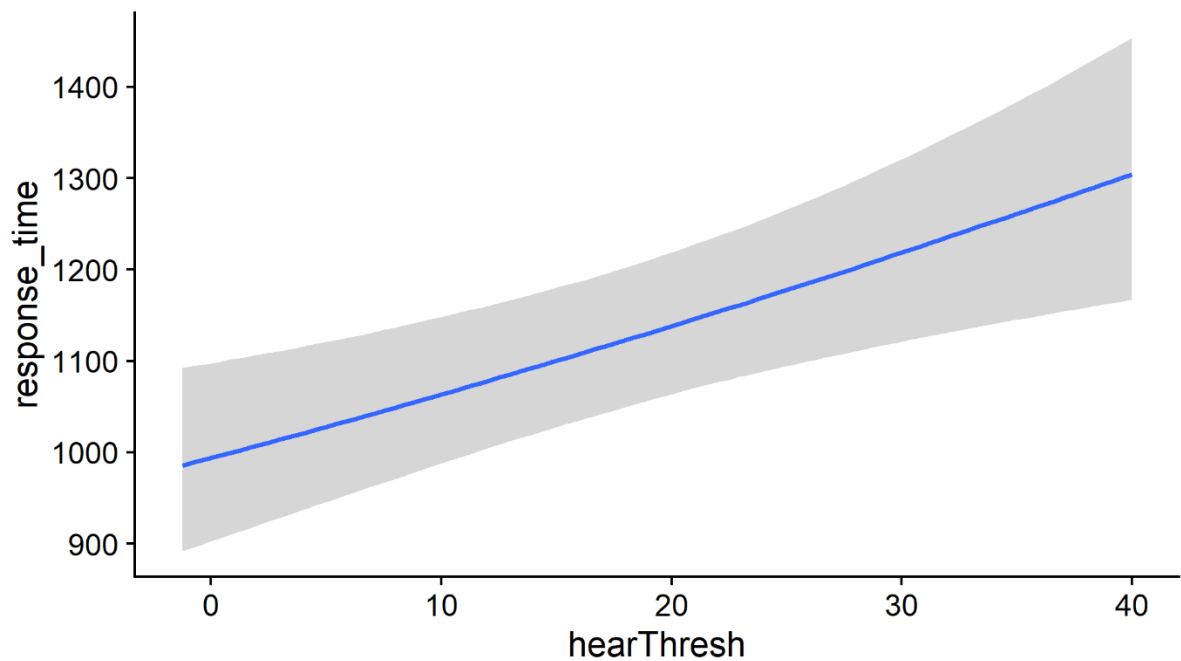


Figure 4—9. HLM parameter estimates for participant hearing thresholds predicting incongruent Stroop task response times (milliseconds) with accompanying uncertainty boundaries.

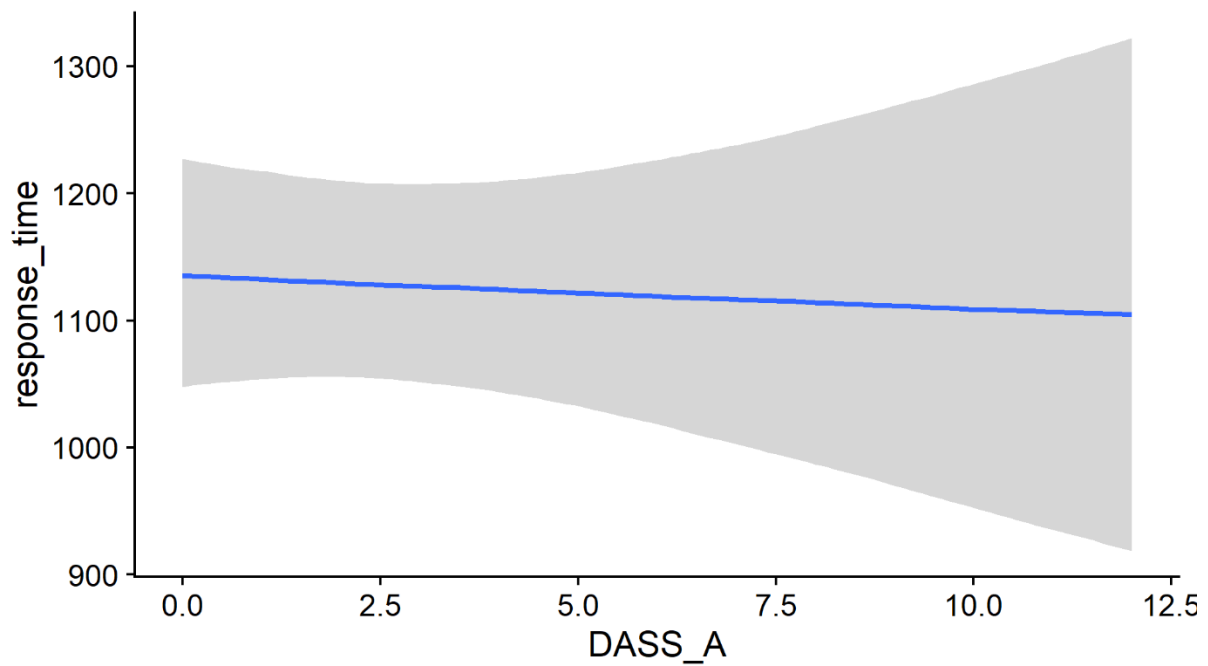


Figure 4—10. HLM parameter estimates for DASS_A predicting incongruent Stroop task response times (milliseconds) with accompanying uncertainty boundaries. DASS_A = Depression, Anxiety, and Stress 21 (Anxiety).

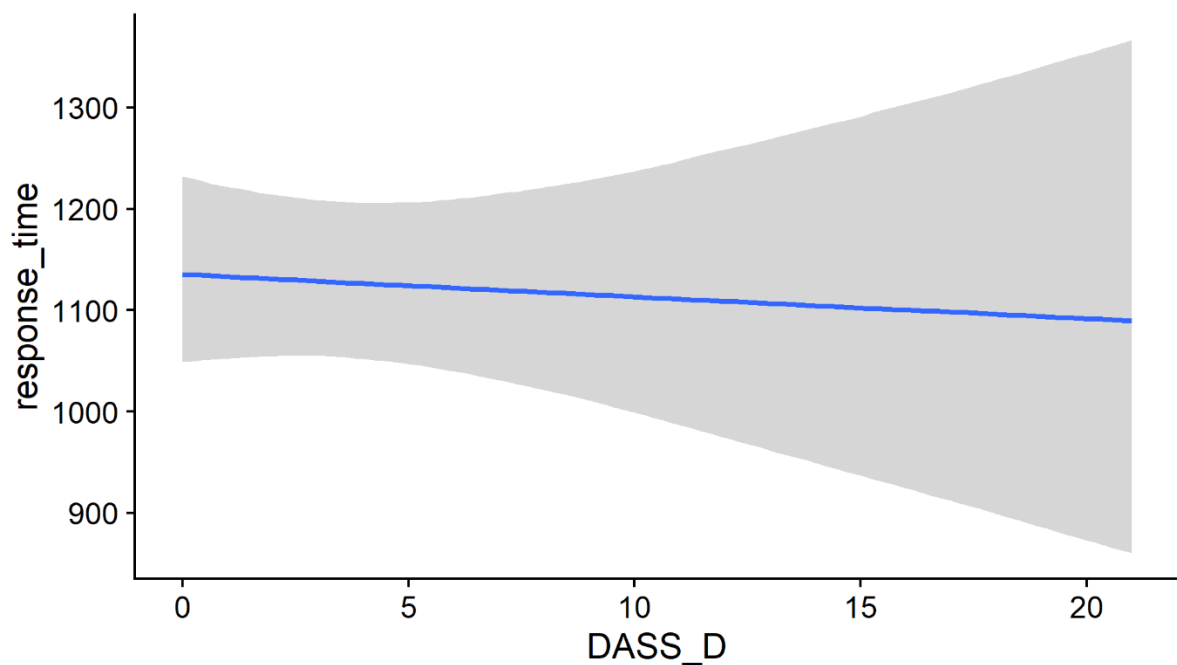


Figure 4—11. HLM parameter estimates for DASS_D predicting incongruent Stroop task response times (milliseconds) with accompanying uncertainty boundaries. DASS_D = Depression, Anxiety, and Stress 21 (Depression).

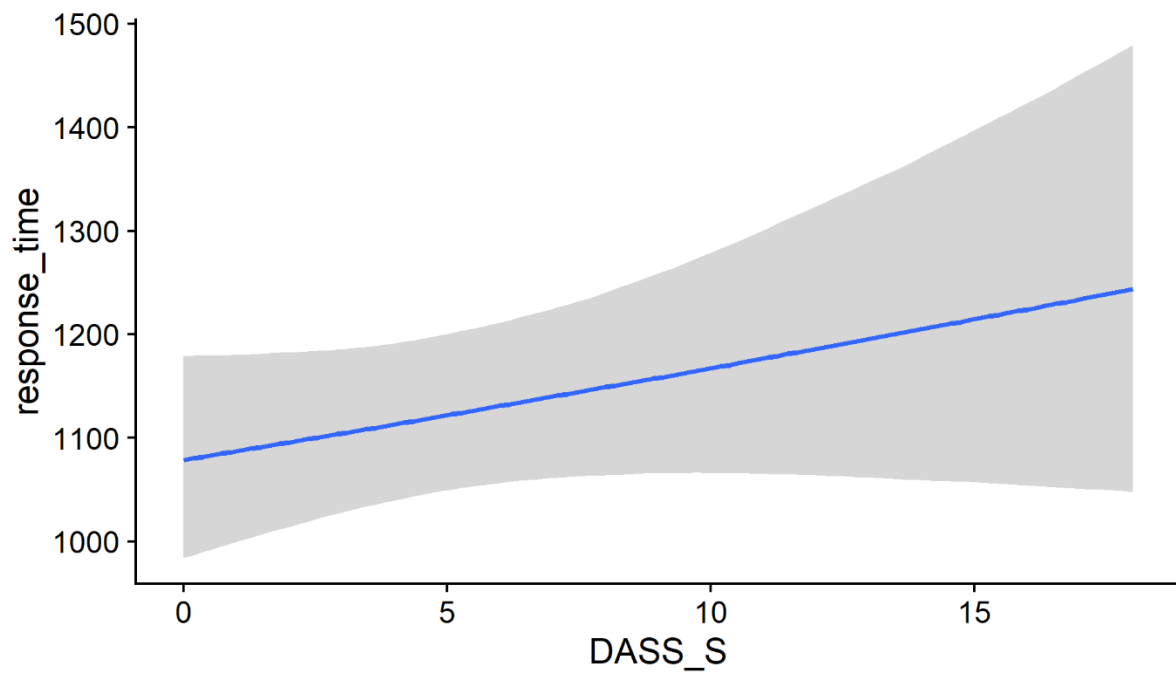


Figure 4—12. HLM parameter estimates for DASS_S predicting incongruent Stroop task response times (milliseconds) with accompanying uncertainty boundaries. DASS_S = Depression, Anxiety, and Stress 21 (Stress).

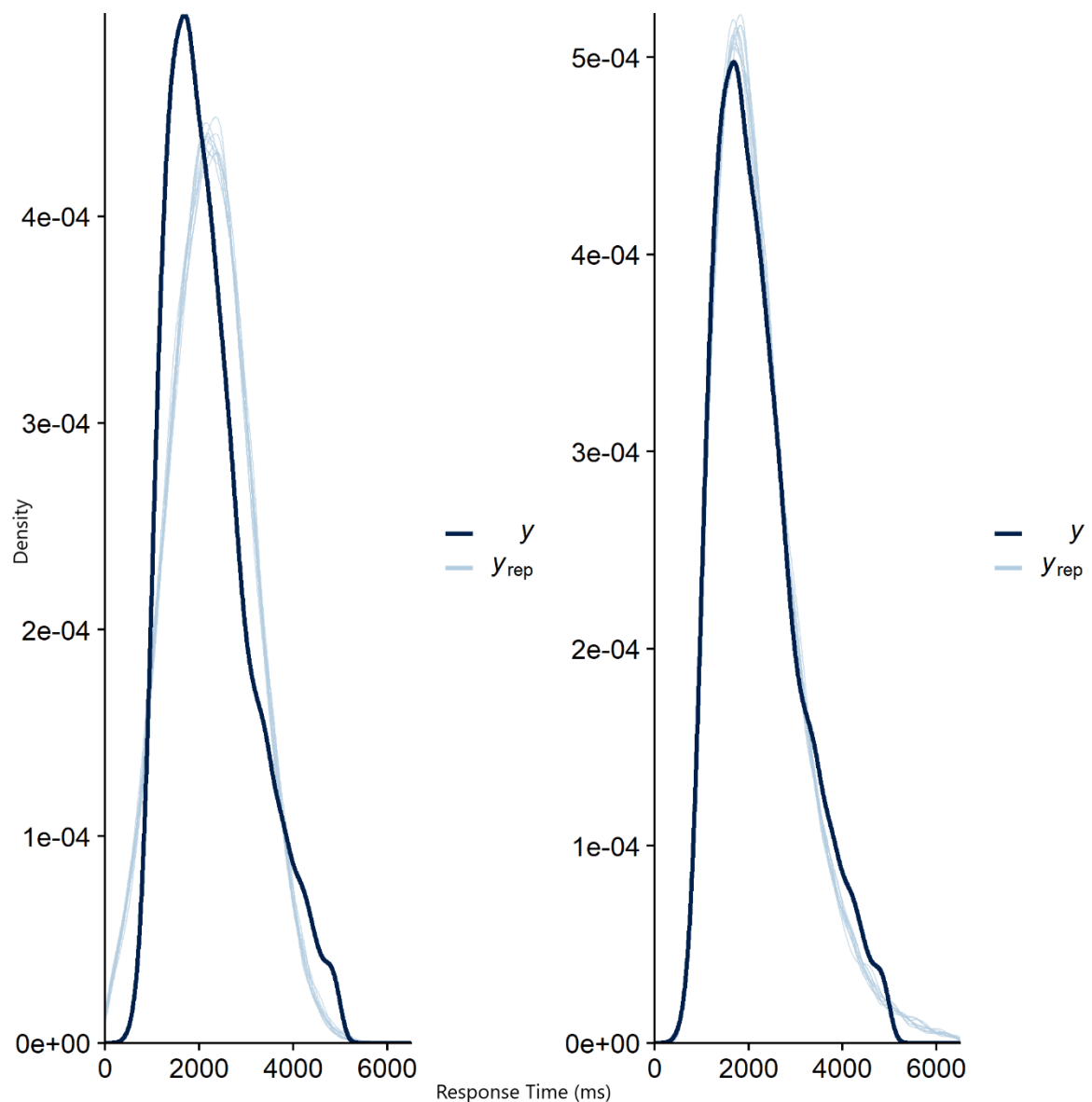


Figure 4—13. Posterior predictive checks for rule-switch trials normal and shifted log-normal models.

HLMs for the rule-switching task were fit using the same parameters, but specifying both normal and shifted log-normal distributions. PPCs demonstrate that the shifted log-normal distribution in the right-sided panel provides a superior fit to the observed data compared to the normal distribution (Figure 4.13), as the draws (light lines) tightly follow the observed data (dark lines).

Table 4—166. Hierarchical linear model parameter estimates for rule-switching task (switch trials) response time.

term	estimate	std.error	lower	upper
b_Intercept	7.53	0.04	7.46	7.6
b_THI_cat2	0.01	0.06	-0.08	0.1
b_THI_cat3	0	0.06	-0.09	0.1
b_scaleage	0.12	0.03	0.07	0.16
b_scalehearThresh	0.04	0.03	0	0.08
b_scaleDASS_D	0.01	0.03	-0.04	0.06
b_scaleDASS_A	0.03	0.03	-0.02	0.08
b_scaleDASS_S	0.04	0.03	-0.01	0.09
sd_participant__Intercept	0.25	0.02	0.22	0.27
sigma	0.36	0.01	0.35	0.37
ndt	177.33	25.63	134.3	218.06

Group-level parameter estimates for the rule-switch model are shown in Table 4.16.

They show that having distressing tinnitus was most probably not associated with any difference in response times compared the No Tinnitus group, while having non-distressing tinnitus was most probably associated with 1% slower responses than the No Tinnitus group after controlling for other variables. The conditional effects for the various groups are shown in Figure 4.14.

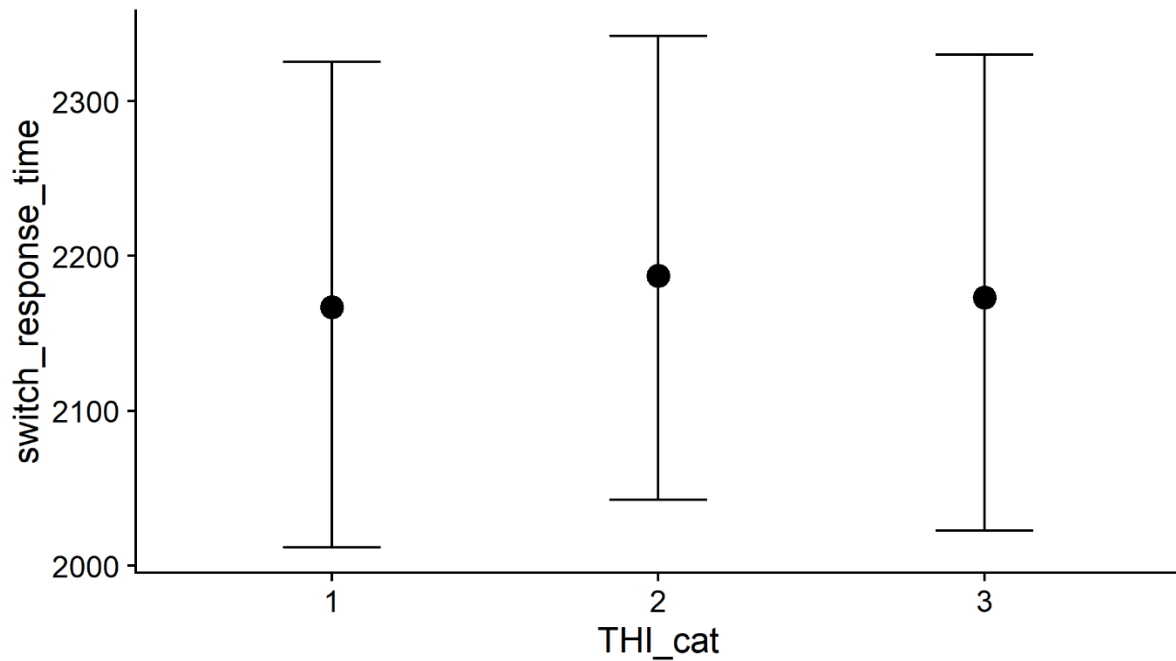


Figure 4—14. HLM parameter estimates for each tinnitus groups predicting incongruent switch trials task response times (milliseconds). 1 = No Tinnitus, 2 = Non-Distressing Tinnitus, 3 = Distressing Tinnitus. THI = Tinnitus Handicap Inventory.

Participant age, and hearing thresholds (Figures 4.15 & 4.16) were associated with increased response time (most probably 12% and 4% respectively). Anxiety and stress (Figures 4.17 & 4.19) scores were also associated with an increase in response times (most probably 4%, 3%, and 1% respectively).

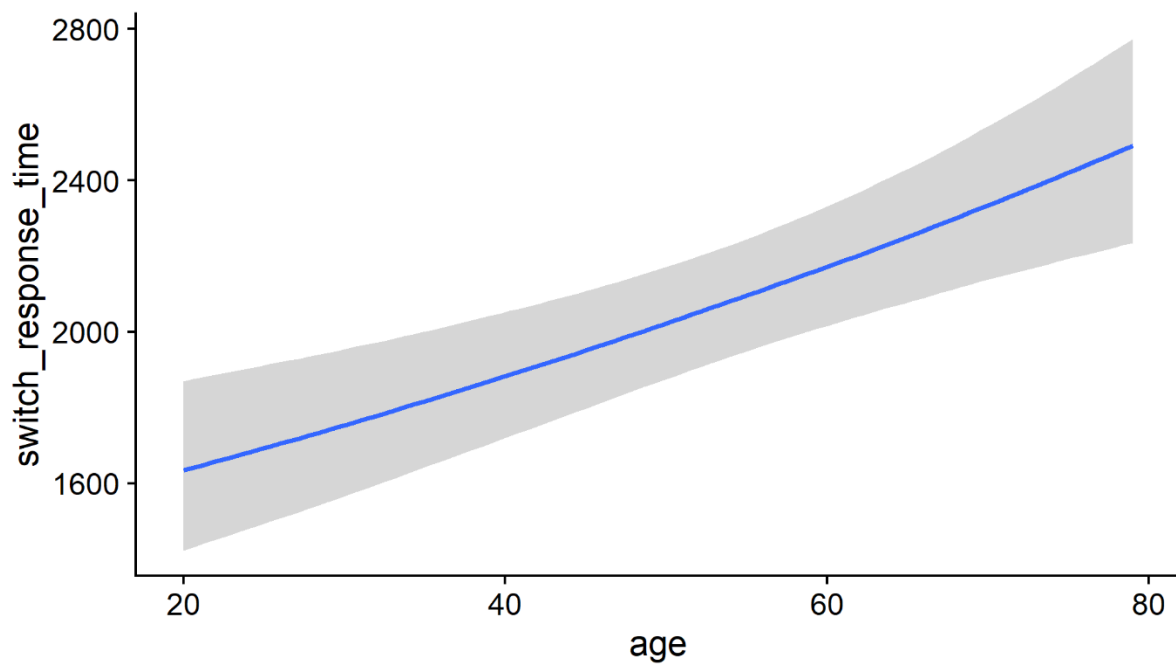


Figure 4—15. HLM parameter estimates for age predicting incongruent switch trials task response times (milliseconds) with accompanying uncertainty boundaries.

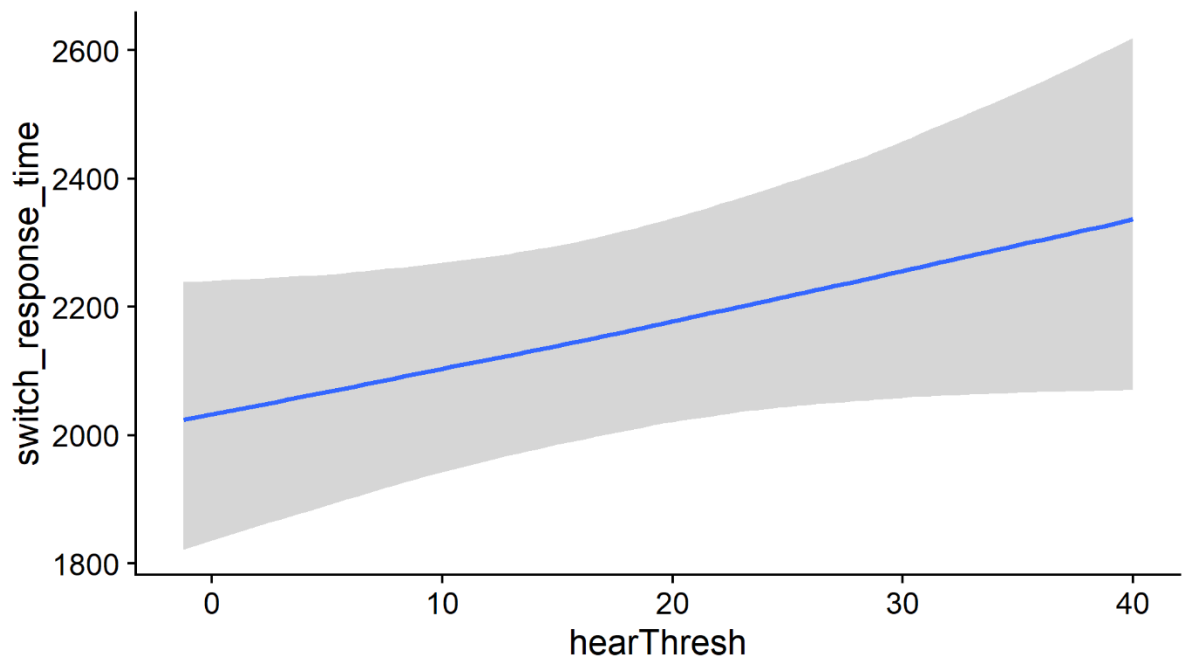


Figure 4—16. HLM parameter estimates for participant hearing thresholds predicting incongruent switch trials task response times (milliseconds) with accompanying uncertainty boundaries.

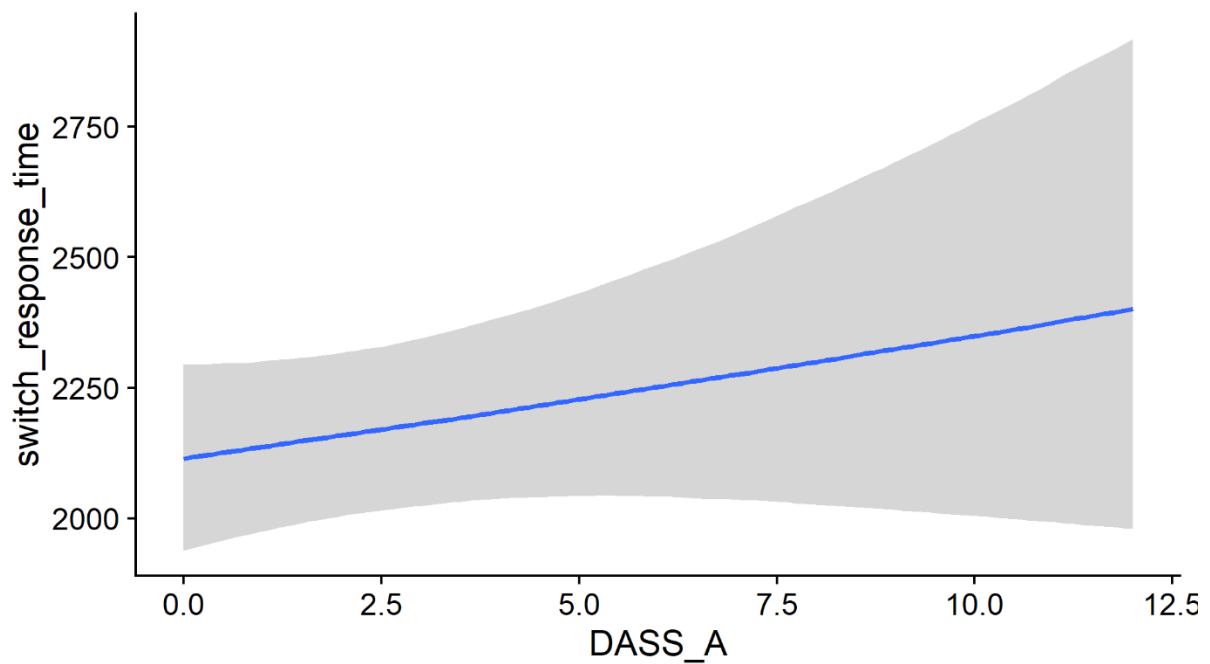


Figure 4—17. HLM parameter estimates for DASS_A predicting incongruent switch trials task response times (milliseconds). DASS_A = Depression, Anxiety, and Stress 21 (Anxiety) with accompanying uncertainty boundaries.

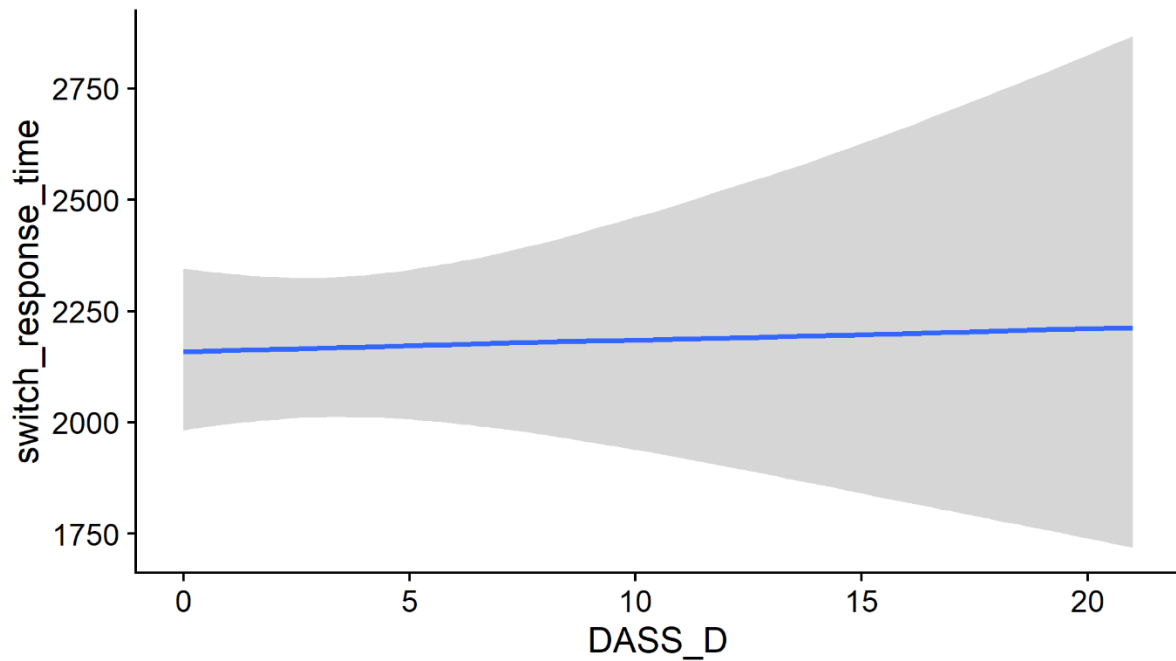


Figure 4—18. HLM parameter estimates for DASS_D predicting incongruent switch trials task response times (milliseconds) with accompanying uncertainty boundaries. DASS_D = Depression, Anxiety, and Stress 21 (Depression).

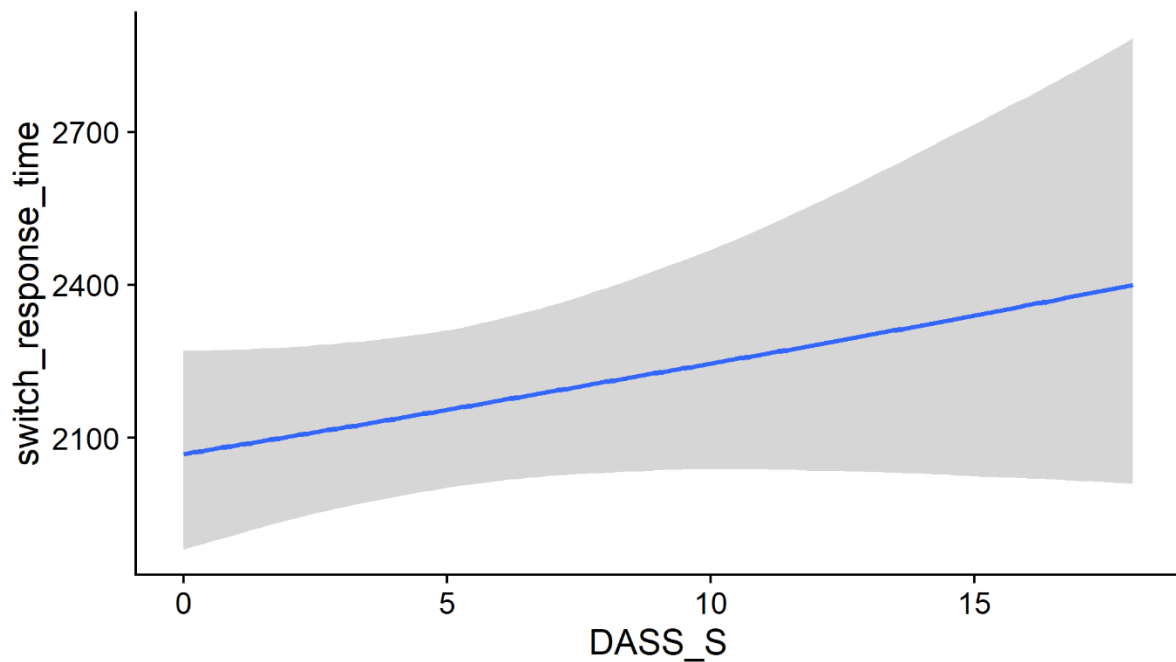


Figure 4—19. HLM parameter estimates for DASS_S predicting incongruent switch trials task response times (milliseconds) with accompanying uncertainty boundaries. DASS_S = Depression, Anxiety, and Stress 21 (Stress).

4.3.5.5 Summary of response time models

Various hypotheses concerning the HLM parameter values can be directly tested with Bayes factors (Bürkner, 2017). For the Stroop task, the hypothesis that distressing

tinnitus per se was associated with increased response times provided moderate evidence for H1 ($BF = 3.3$), while the Tinnitus group being associated with quicker response times provided slight anecdotal evidence for H1 ($BF = 1.3$). In the Stroop task, the Distressing Tinnitus group responded 4-5% slower than the other groups. On average, the Distressing Tinnitus group responded 43-53.5 ms slower than either group after controlling for other variables. Given that the interquartile-range of responses on this task was 643ms, the relative effect of 53.5 ms poorer performance associated with having distressing tinnitus might be considered truly impactful.

For the rule-switching task, there was no difference in response times between the Distressing Tinnitus and No Tinnitus groups. The Non-Distressing Tinnitus group was associated with 1% slower response (18.7 ms), which implied slight anecdotal evidence for H1 ($BF = 1.3$). The median response time was 2100 ms. Given the interquartile range of 1333 ms for this task, the effect of an 18.7 ms increase in response time for the Non-Distressing Tinnitus group appears non-impactful.

Although the parameter values represent the most probable values for the each model, uncertainty is inherent in inference, and it is therefore important to note that the credible intervals for parameter values are still broad. However, this was not the case for age and hearing thresholds for both tasks. In the Stroop task, these provide strong evidence for H1 with BFs of 2749 and 1999 respectively. For the rule-switching, the evidence that hearing thresholds had an influence was more modest, but still strong ($BF = 18.9$). Age was strongly associated with increased response time ($BF = 21999$).

In the Stroop task, both anxiety and depression were weakly associated with improved response times, providing anecdotal evidence in support of H1 with BFs of 1.5 and 1.6 respectively. In the rule-switching task, depression was appeared to have no association, while anxiety was associated with 3% worse response times, providing moderate evidence ($BF = 6.6$). Stress was associated with increased response times in both models, providing moderate-to-strong evidence for H1 in both tasks ($BF = 9.2$ both models).

The results of this analysis should be considered in relation the speed-accuracy trade off that is inherent in response time data; therefore errors need to also be considered (Heitz, 2014). In light of this, there was moderate evidence that distressing tinnitus is associated with increased response times (probably 4-5% slower), which is significant in the context of a relatively simple task like the colour-word Stroop paradigm employed in the EFiT study. Additionally, being in the Distressing Tinnitus group was associated with making more errors after controlling for other variables ($\beta = .17$), and Distressing Tinnitus was the first ranked explanatory variable for these errors. This demonstrates that the Distressing Tinnitus group had poorer Inhibition performance compared to the No Tinnitus group; while membership of the Non-Distressing Tinnitus group per was not associated with poorer inhibitory performance compared to the No Tinnitus group after controlling for confounding variables.

There was no difference between response times attributable to distressing tinnitus per se, or in the number of errors produced compared to the No Tinnitus group ($\beta = 0.04$). As the rule-switching is a more complex task, this analysis therefore supports the conclusion that there is no difference in Shifting ability in those who have distressing tinnitus. The Non-Distressing Tinnitus group performed negligibly worse than the No Tinnitus group after controlling for confounding variables; however, they made noteworthy fewer errors than the No Tinnitus group ($\beta = -.23$), and supports the concept that participants with tinnitus who are not distressed by it demonstrate improved Shifting performance.

4.4 General discussion of advanced analyses and modelling

4.4.1 Objective measures

This study aimed to identify which core executive functions exhibited poorer performance for participants with Non-Distressing Tinnitus and Distressing Tinnitus, and if differences were likely attributable to having Non-Distressing Tinnitus or Distressing Tinnitus per se.

Taken together, the results demonstrate clear support that Distressing Tinnitus is independently associated with poorer executive functioning across all tasks. However, having Non-Distressing Tinnitus presents a more complicated picture, as results suggest that this group had slightly poorer functioning for some executive domains but improved performance in others domains.

When solely considering between-group effect sizes, a clear pattern of poorer performance appears to emerge across all tasks, worsening when comparing No Tinnitus to Non-Distressing Tinnitus, and Non-Distressing Tinnitus to Distressing Tinnitus groups. Attribution of poorer performance to group membership generally remained consistent for the Distressing Tinnitus group following advanced analyses that accounted for known confounding factors. The Distressing Tinnitus group displayed poorer performance across Updating, Inhibition, Shifting, and fluid ability. This was demonstrated by between-group effect sizes ranging from 0.13 to .68. Bayesian GLMs then demonstrated that it was highly probable that these effects were attributable to having distressing tinnitus *per se*. This study is therefore consistent with previous research concluding that having tinnitus is associated with poorer executive functioning (Clarke et al., 2020; Mohamad et al., 2016; Tegg-Quinn et al., 2016; Trevis et al., 2017a). This analysis also revealed that performance varied across core domains of executive function.

However, advanced analyses for the Non-Distressing Tinnitus group revealed a pattern of results where initial poorer performance (as indicated by between-group effect sizes), are in fact attributable to confounds rather than having non-distressing tinnitus; for example, although number-letter sequencing performance was poorer in the Non-Distressing Tinnitus compared to Control Group (and closer to the performance of the Distressing Tinnitus group), GLMs demonstrated that participant age was the most powerful explanatory variable, leaving essentially no effect associated with Non-Distressing Tinnitus group membership. This study also builds on previous research through demonstrating that non-distressing tinnitus is likely associated with improved abilities in specific executive domains compared to people without tinnitus (i.e. Shifting

and fluid ability). The Non-Distressing Tinnitus group made less errors than the No Tinnitus group on the rule-switching task; and although the Non-Distressing Tinnitus group exhibited slightly poorer Matrix reasoning performance than the No Tinnitus group, GLMs demonstrated that being in the Non-Distressing Tinnitus group was, in fact, probably associated with improved Matrix reasoning performance when covariates were considered.

GLMs demonstrated that the Non-Distressing Tinnitus group performed slightly worse than the No Tinnitus group in the Keep Track, but no differently in the NLS task, providing mixed evidence for poorer Updating performance overall. The Non-Distressing Tinnitus group also performed slightly worse in the dual-task, however the associated effects were larger for anxiety, hearing thresholds, and age. Given the uncertainty of response times and decreased error rates for the rule-switching task, the results for poorer Shifting being associated with Non-Distressing Tinnitus are also mixed. The results are also consistent with the interpretation that having Non-Distressing Tinnitus is actually associated with improved Shifting and fluid ability compared with the No Tinnitus group.

This study supports previous research that has concluded that having tinnitus is associated with poorer executive functioning compared to not having tinnitus (Araneda et al., 2015); however, it shows that having Distressing Tinnitus is independently associated with poorer cognitive control, whereas having Non-Distressing Tinnitus is generally not; this is contrary to previous reports that have suggested that deficits in cognitive control may be generalisable to low tinnitus severity (Jackson et al., 2014). In fact, those who have Non-Distressing Tinnitus appear to perform better in various cognitive domains once confounding factors have been accounted for. This suggests a more fundamental, and perhaps causal role in the relationship between cognitive control the tinnitus percept in determining whether an individual is distressed by it.

Moreover, the Distressing Tinnitus group made more Stroop errors suggesting response inhibitory errors. This is consistent with Clarke et al. (2020) and a hypothesis

of Trevis et al. (2016), which postulates that people with tinnitus may fail to inhibit the percept. Conversely, the Non-Distressing Tinnitus group made less rule-switching errors than both Distressing Tinnitus and No Tinnitus groups, suggesting that there is an enhanced Switching ability in the Non-Distressing Tinnitus group. T

The Stroop task was chosen and suggested to be an appropriate measure to emphasise the core executive function of Inhibition (Diamond, 2013). A limitation to the interpretation of the experimental results using the Stroop task concerns the attribution of slower response time to inhibition per se. This is because the incongruent condition of the Stroop task was the sole condition used in analysis. This obfuscates interpretation of the results as a measure of specific interference because the incongruent condition measures both general slowness (i.e., processing speed) and interference. It has been suggested that using the difference between the neutral and incongruent condition should remove factors common to both conditions such as general slowing (Wright, 2017). Therefore it may be argued that general slowing could be responsible instead of an inhibitory effect.

However, the interpretation of the results as an inhibitory effect is supported by results from Clarke et al. (2020), which demonstrate that EF-inhibition was poorer across various inhibitory paradigms and analytic methods, including various Stroop paradigms and analyses, with the meta-analysis providing no evidence of a single result driving this finding. Furthermore, the additional measure of interference controlled for output interference within the paradigm administered to all participants, which increases the likely attribution of the differences between groups to an inhibitory mechanism (Murayama et al., 2014). Therefore, the union of results from this measure and previous synthesis and inhibitory paradigms should increase confidence in attributing Stroop findings to an inhibitory mechanism.

Absence of the RIF effect in the Non-Distressing Tinnitus group and a reversal of the effect in the Distressing Tinnitus group provides evidence for Inhibition being poorer in both groups, particularly as output interference was controlled for, suggesting the mechanism is more strongly attributable to an inhibitory explanation (Murayama et al., 2014; Storm and Levy, 2012). This result is consistent with previous findings that reported an absence of the RIF effect in clinical populations; for example, Groome and Sterkaj (2010) reported an absence of the RIF effect in subjects with clinical depression. GLMs demonstrated that while depression did account for a small negative association in recall of Np items, this association was smaller than with anxiety and tinnitus. As shown in the descriptive statistics, the RIF effect has a large variability and therefore provides weak evidence for Inhibition being poorer per se and required further analysis (Section 4.3.4). Detailed analysis of the RIF effect demonstrates that participants in the Distressing Tinnitus group scored 7-11% worse in both practised and unpractised items. This supports Clarke et al. (2020) through demonstrating that Distressing Tinnitus is associated with poorer encoding and retrieval in Glr. However, for unpractised items, the Non-Distressing Tinnitus group performed 5-8% better than either the Distressing Tinnitus and No Tinnitus groups. Although this findings could be viewed as providing evidence for poorer inhibition (of non-practised words), it is more consistent with a generally superior Glr performance by the Non-Distressing Tinnitus group.

Previous research has suggested that tinnitus affects cognitive performance through disruption of auditory attention and cross-modal interference (Cuny et al., 2004; Li et al., 2018). As the objective cognitive measures used in the EFiT study were all visual (with the exception of the dual-task which had an auditory component), the reported results provide support to the idea that domain general executive processes are typically poorer in persons with distressing tinnitus.

While all of these tasks are presumed to measure core executive functions, they vary in their complexity and difficulty, thereby applying different amounts of cognitive load (Lavie, 2010). For instance, the Keep track is more complex example of updating

than NLS, and the dual task more complex than the rule shifting. Therefore the general increase in task complexity (coupled with latent load from varying tinnitus severity) could be argued to account for between-group differences in effect sizes. However, viewed in the Diamond (2013) framework, the performance of the Non-Distressing Tinnitus group in the complex matrix reasoning task would conflict with this interpretation. However, this presumes the statistical model allowed for correct inference, whereas in fact, the observed effect size for the Non-Distressing Tinnitus group does show a small poorer performance relative to the No Tinnitus group. Another factor could be that the matrix reasoning was the only task not performed under a time constraint, thereby mitigating the effect of increasing load. This indicates that some cognitive load could be required to deal with a constant tinnitus percept. Load could be increased, or cognitive executive functioning capacity could be decreased. Both would result in poorer cognitive performance.

A further interpretation for the observed pattern of results suggests that people who have either distressing or non-distressing tinnitus may in fact have diminished updating of working memory performance relative to people without tinnitus in some domains, such as updating of working memory (Nagaraj et al., 2020), but improved and therefore compensatory performance in other domains (i.e. fluid ability and Shifting) that facilitates adequate functioning in daily situations despite potential intrusion of the tinnitus percept.

4.4.2 Self-report measures

The Distressing Tinnitus group had higher CFQ scores than the No Tinnitus Group, indicating that they had more perceived cognitive failures within the last six months. This size of this effect is large ($d = .59$; 95% CI[.17 , 1.02]), which is consistent with the full cognitive self-report model estimated in Clarke et al. (2020), and suggests that meta-analysis results were being driven by participants with Distressing Tinnitus. The Non-Distressing Tinnitus group actually reported less cognitive failures than the No

Tinnitus group ($d = -.18$; 95% CI $[-.56, .23]$, which suggests that they perceive improved performance in certain cognitive tasks.

Taken together, these results contribute strong evidence to the question posed by Mohamad et al. (2016) of whether cognitive performance relates to self-reported concentration difficulties for people with tinnitus, by showing that cognitive constructs and evaluation of concentration difficulties are associated in people with tinnitus. However, further work is required to ascertain whether these results reflect that people with non-distressing tinnitus/distressing tinnitus are simply aware that they perform better/worse, or whether there is a causal element to this (i.e. expecting poorer cognitive performance is self-fulfilling). For instance, previous research has indicated that self-reported cognitive impairment and related anxiety can be alleviated by clinical contact with health professionals (Hill et al., 2016).

Rumination scores for both the Non-Distressing and Distressing Tinnitus groups ran in opposite directions relative to the No Tinnitus group; they were larger in the Distressing Tinnitus group ($d = .49$; 95% CI $[.06, .91]$), and smaller in the Non-Distressing Tinnitus group ($d = -.41$; 95% CI $[-.83, .02]$). This indicates that people with distressing tinnitus engage in more ruminative behaviour than those with non-distressing tinnitus. These results support previous research that has suggested increased rumination may contribute to negative thinking about the tinnitus percept, as well as emotion dysregulation (Trevis et al., 2016).

The CRIq initially showed that the Non-Distressing Tinnitus group had the highest scores and therefore the most cognitive reserve; however, as this group was the oldest on average, their score is confounded by the age of the participants, because cognitive reserve is thought to grow through engaging with cognitively enriching activities across the lifetime (Stern, 2013, 2009), and improved cognitive function is positively associated with greater cognitive reserve (Opdebeeck et al., 2016). However a linear model with age as a covariate shows that having Non-Distressing Tinnitus and Distressing Tinnitus is negatively associated with cognitive reserve. Cognitive reserve has been linked to

memory and executive function performance, as well as neuroplasticity (Barulli and Stern, 2013; Nyberg et al., 2012; Roldán-Tapia et al., 2012); it has also been studied in relation to cognitive aging and in clinical populations such as having multiple sclerosis (Cabeza et al., 2018; Nunnari et al., 2016; Sumowski and Leavitt, 2013). To the best of my knowledge, no previous study has investigated the association between tinnitus and cognitive reserve; however, the results presented here are suggestive of a causal relationship and therefore warrant further investigation.

4.4.3 Methods and analysis

Although previous research has implicated cognitive control as being associated with tinnitus, the exact nature of this relationship has remained unclear, with studies produced inconsistent results across various cognitive domains (Clarke et al., 2020; Mohamad et al., 2016; Tegg-Quinn et al., 2016). The analyses presented here suggest that the observed inconsistencies in results may be due to inferential methods typically employed in the analysis of cognitive data. NHST can lead to results being subjected to binary classification of significance in a general, not merely statistical sense (McElreath, 2016). Table 4.2 shows that when considered with other variables, GLMs only provide statistically significant p-values for a small set of variables, with Distressing Tinnitus appearing only twice, despite the clear evidence of non-zero effects provided by equivalent Bayesian GLMs. Bayesian methods are becoming increasingly used and accessible within the wider scientific literature, particularly with psychological fields (Etz et al., 2018; Muth et al., 2018; Quintana and Williams, 2018). Although there are some examples of Bayesian methods being applied to tinnitus and hearing sciences (McMillan et al., 2014; McMillan and Cannon, 2019), future research into associations between tinnitus and cognitive performance would likely benefit from their inclusion.

Although effect sizes found in this study are generally smaller than those found in Clarke et al. (2020), this was likely due to confounding from variables highlighted in this study. However, this is arguably consistent with broader expectations of effect replications. Camerer et al. (2018) demonstrated that when replicating effect sizes, they

were generally smaller (typically .75 of original effect) than the study they intended to replicate.

Bayesian GLMs revealed that having Non-Distressing Tinnitus or Distressing Tinnitus provided uncertain (essentially 50/50) parameter estimates of response times in the Stroop and rule-switching tasks. Modelling of response times is a complex topic due to issues such as speed-accuracy trade-offs, distribution of responses, and outliers (Baayen and Milin, 2010; Heitz, 2014; Whelan, 2008). The insights afforded by non-Gaussian HLMs in this study suggest that more advanced modelling strategies should be routinely employed to provide better parameter estimation for the observed effects when using response times as a measure in a tinnitus population (Muth et al., 2018; Whelan, 2008).

Although the effects are generally small, the poorer executive functioning performance elicited in a laboratory setting should be considered as a demonstrable example of the executive abilities that are required in a person's day-to-day functioning. These findings are consistent with broader empirical findings concerning expected effect sizes in individual difference research (see Section 4.5), but should be afforded particular importance given their likely cumulative nature (Funder and Ozer, 2019; Gignac and Szodorai, 2016).

4.4.4 Other variables

As anticipated, participant age was found to be an important variable associated with poorer performance in many GLMs within this study. This is consistent a large cognitive gerontological literature that has detailed age-related cognitive changes across the life span, and suggests that fluid ability typically declines during aging (Hartshorne and Germine, 2015; Perfect and Maylor, 2000; Verhaeghen, 2013). While the fluid ability of the Distressing Tinnitus group was poorer, this was not the case for the Non-Distressing Tinnitus group. Interestingly, GLMs suggest that the Non-Distressing Tinnitus group exhibited improved fluid ability compared to the No Tinnitus group. Furthermore, the Non-Distressing Tinnitus group had the oldest average age, arguably making this finding

more striking and suggesting a key role for fluid ability in relation to Distressing Tinnitus. One interpretation of this might be that people who have tinnitus but do not find it distressing are afforded a compensatory effect through improved fluid ability.

Stress levels were found to be substantially different across groups with over half of the Distressing Tinnitus group reporting at least mild levels of recent stress (Figure 3.10). Furthermore, stress was an important variable in GLMs, repeatedly accounting for variance in tasks (Stroop response, rule errors and costs). This is consistent with research that has found that stress can result in impairment in memory and reasoning tasks and emerging literature highlighting its importance to tinnitus distress (Mazurek et al., 2019; Sandi, 2013). Anxiety was found to be one of the two highest ranked variables in GLMs, including dual task, matrix reasoning, and rule-switching errors. Previous research has shown an association between poorer cognitive performance and anxiety, which has been theorised to result from poorer attentional control (Eysenck and Derakshan, 2011). Despite a quarter of the Distressing Tinnitus group having moderate or greater levels of depression, this variable was not found to be strongly associated most GLMs objective cognitive measures; effects were small and for RIF effect, rule-switching errors, and dual task.

Taken together, these observations highlight that successful cognitive performance is associated with many variables, with participant age, anxiety, stress, and hearing levels all contributing to varying extents towards successful performance in cognitive tasks. The results of this study show that having Distressing Tinnitus also makes an important contribution.

Despite the importance of stress and anxiety in many models, the Non-Distressing Tinnitus group clearly exhibited lower DASS scores than either group (Figure 3.11). Approximately one-third of the No Tinnitus group and over half of the Distressing Tinnitus group had moderate scores for either depression, anxiety, or stress, while less than one-tenth of the Distressing Tinnitus group did. Taken together with GLM results, these observations highlight the importance of a profile of effects that contribute to

cognitive performance individuals have varying age, hearing, DASS levels and all of these factors will *combine* to interact with cognitive performance, with small effects adding up.

Hearing thresholds were found to be an important variable in various models, particularly the dual task, which indicates that performance on this task was confounded by the participants co-varying hearing thresholds, which are typically expected to be poorer in participant with tinnitus (Rauschecker et al., 2010; Sedley et al., 2016). Although participants in the study all had a better-ear-average of no greater than a mild hearing impairment, the audiometric configuration of participants appears to have affected speech intelligibility and subsequently performance on the secondary cognitive task (i.e. digit recall) in the dual task paradigm. Hearing thresholds were also notably independently associated with poorer performance on the Stroop task for error rates and response times (Table 4.6). This could be interpreted as consistent with a literature that details poorer cognitive performance in persons with hearing impairment (Wayne and Johnsrude, 2015). Another interpretation is that participants with a higher hearing thresholds did not hear the instruction, but this is unlikely as participants all had a mild hearing impairment, were given consistent instructions in a quiet room, and the investigator ensured that the participant performed the practise trials to an adequate standard.

4.5 Consideration of observed effect sizes: small differences add up

To effectively appreciate the relationship between tinnitus and cognitive performance, it is useful to 1) understand the size of the observed between-group differences in more detail, and 2) to consider what these might practically mean for people who have Non-Distressing and Distressing tinnitus.

To achieve the first of these aims and effectively interpret the differences found between the groups in the EFiT study, implied distributional differences between groups

can be considered. This can be achieved using graphical means through overlapping distributions arising from the calculated SMDs (Figure 4.20). These distributions can be summarised with statistics that provide a measure of their characteristics (i.e. overlap, non-overlap, and the probability of superiority) to aid in further interpreting the standardised effect sizes (Table 4.17):

- **Overlap (OVL):** generally called the overlapping coefficient (OVL), it is a measure of agreement between two distributions (Reiser and Faraggi, 1999)
- **Non-overlap (U3):** Cohen (1988) defined U3 as a measure of non-overlap, where "we take the percentage of the A population which the upper half of the cases of the B population exceeds".
- **Probability of superiority (PoS):** is effect size with many names, including common language effect size and area under the receiver operating characteristics (Ruscio and Mullen, 2012). PoS is thought to be more intuitive for persons without any training in statistics. The effect size gives the probability that a person picked at random from the treatment group will have a higher score than a person picked at random from the control group.

Table 4—177. Effect size distribution summary statistics for Non-Distressing Tinnitus and Distressing Tinnitus groups compared to the No Tinnitus Group, showing standardised mean differences, overlap (OVL), non-overlap (U3), and probability of superiority (PoS) for objective cognitive measures.

Groups	Task	d	OVL	U3	PoS
Control and Distressing Tinnitus	Matrix Reasoning	.69	73	75.5	68.7
	Keep Track	.53	79.1	70.2	64.6
	Dual Task	.41	83.8	65.9	61.4
	NLS	.37	85.3	64.4	60.3
	Switch-cost	.27	89.3	60.6	57.6
	Stroop	.28	88.9	61	57.8
Control and Non-Distressing Tinnitus	Matrix Reasoning	.12	95.2	54.8	53.4
	Keep Track	.52	79.5	69.8	64.3
	Dual Task	.41	83.8	65.9	61.4
	NLS	.25	90.1	59.9	57
	Switch-cost	0	0	0	0
	Stroop	.28	88.9	61	57.8

While GLMs demonstrated that the association between the Non-Distressing Tinnitus group and various executive functioning measures was likely small, the SMDs between No Tinnitus and Distressing Tinnitus groups were similar to simple between-groups models. Figures 4.20 illustrates the size of these differences for the Distressing Tinnitus group in relation to the No tinnitus group for the matrix reasoning task. The smallest of these differences was for switch-cost, where implied distributions overlap entirely. However, for the Matrix reasoning task, approximately two-thirds of the No Tinnitus group are expected to be above the mean of the Distressing Tinnitus group, and approximately two-thirds of the distributions for the No Tinnitus and Distressing Tinnitus groups overlap; the PoS highlights that there is a nearly 70% chance that a person picked at random from the No Tinnitus group will have a higher score than a person picked at random from the Distressing Tinnitus group.

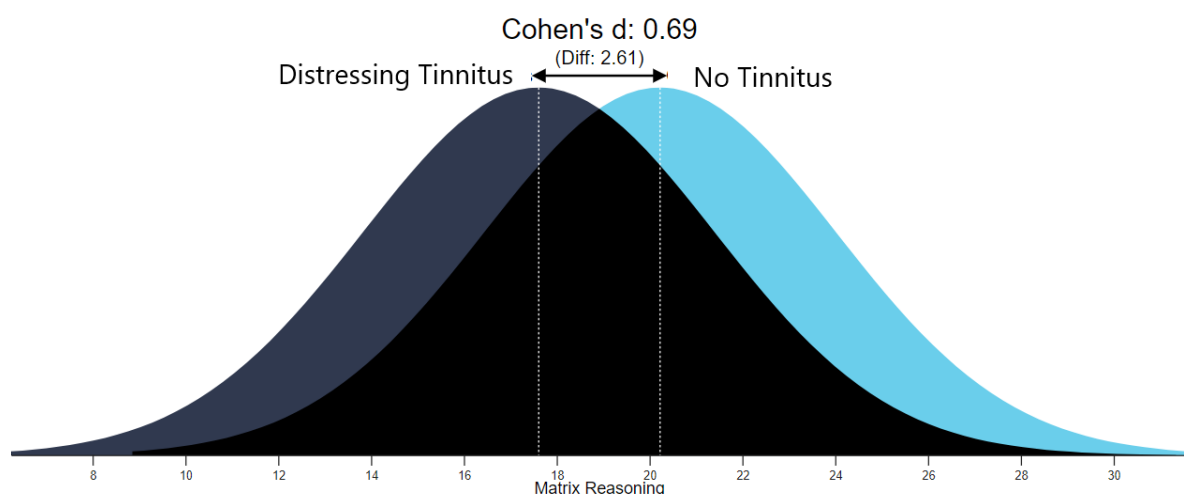


Figure 4—20. For the Matrix reasoning task, a Cohen's d of 0.69 implies that 75.5% of the No Tinnitus group will be above the mean of the Distressing Tinnitus group (Cohen's U3), 73% of the two groups will overlap, and there is a 68.7% chance that a person picked at random from the No Tinnitus group will have a higher score than a person picked at random from the Distressing Tinnitus group (probability of superiority)

Having considered the effect sizes between groups, it useful to turn to the second aim to consider if these differences might be considered important for people with non-distressing tinnitus and distressing tinnitus.

When considering the size of the effects demonstrated in the EFiT study, a common heuristic is to invoke descriptive labels developed by Cohen (1988), which were (reluctantly) introduced within the context of power analysis; for correlations, these are small (.1), medium (.3), and large (.5). Therefore, the between-group effects in this study might be labelled non-existent-to-small for the Non-Distressing Tinnitus group, and small-to-medium for the Distressing Tinnitus group. Despite their pervasive use, Cohen's labels were only guidelines, recommended as a last resort only when no empirical information was available (Correll et al., 2020; Gignac and Szodorai, 2016). Furthermore, the broad applicability of these labels to psychological and cognitive research is being reconsidered (Correll et al., 2020; Funder and Ozer, 2019). Funder and Ozer (2019) suggest that such labels without context are nonsensical, and simply prompt the questions:

- a) 'small', 'medium', or 'large' compared to what?
- b) 'small', 'medium', or 'large' for what purpose?

Regarding the first question, various references and alternative ways to characterise the size of an effect exist, such as intuitively understood comparisons (e.g. $r = .44$ for the correlation between height and weight in the US adult population). However, a preferable method for researchers is through empirically benchmarking effects sizes in their specific area of research using an effect size distribution analysis (Gaeta and Brydges, 2020; Quintana, 2017). Benchmarking through effect size distribution is best employed across psychological sub disciplines and within specific literatures (Schäfer and Schwarz, 2019). Therefore an effect size distribution analysis was undertaken on the eighty effect sizes detailed between tinnitus and objective cognitive measures in Clarke et al. (2020).

Table 4—188. Quantile effect size distribution analysis from Clarke et al. (2020).

Effect size from distribution analysis	Quantile	Correlation
Small	25%	.11
Medium	50%	.26
Large	75%	.45

Correlation benchmarks obtained from this analysis of the effect size distribution (Table 4.18) suggests that within this literature, Cohen's general guidelines are appropriate for small effects (i.e. 25% quantile), but tend to slightly overestimate what would be classified as medium and large effects. This is also important when considering future research in this field, because benchmark effect sizes will affect the number of participants required when estimating sample sizes in power calculations, where relatively small differences could translate to non-trivial amounts of unnecessary work for researchers through over-recruitment of participants.

Ozer's (2019) first question ('compared to what?') concerns setting of effect size thresholds for descriptive adjectives, whereas the second ('for what purpose?') concerns the purpose of these labels, and therefore requires attention to be appropriately afforded to the context in which an effect might be considered small, medium, or large. The effects demonstrated in the EFiT data arise from laboratory-based demonstrations of executive performance that can be conceived as providing insight into reduced efficiency expected in a wide variety of day-to-day tasks in a participant's lived experience. Therefore, the poorer performance demonstrated across tasks in this study provides objective evidence for the common complaint of concentration difficulties experienced by those with distressing tinnitus.

While many of these effects are small, and therefore potentially tempting to dismiss as inconsequential, small effects can be important in the long run (Funder and Ozer, 2019). In a classic analysis, Abelson (1985) examined the role of batting skill variance in success for major league baseball, and demonstrated that an individual player's skill could only explain 1/3 of 1% (i.e., .003 proportion) of the variance within various models. This finding was surprising and counterintuitive, as no theory of baseball success would deny the crucial contribution of the player's skill (particularly their employing teams, who set their wages accordingly). A crucial insight into the nature of this importance is that small effects can be important when considered cumulatively.

The importance of cumulative effects over the long-run has already been recognised in some psychological disciplines such as psychotherapy. For example, Imel et al. (2015) observed that small differences in performance across therapists can have important impacts on treatment outcome, while large differences are likely to have dramatic effects. Given the importance of such cumulative effects in human cognition, Funder and Ozer (2019) even suggest revised effect size guidelines that explicitly take them into account: “an effect-size r of .05 indicates an effect that is very small for the explanation of single events but potentially consequential in the not-very-long run, an effect-size r of .10 indicates an effect that is still small at the level of single events but potentially more ultimately consequential, an effect-size r of .20 indicates a medium effect that is of some explanatory and practical use even in the short run and therefore even more important, and an effect-size r of .30 indicates a large effect that is potentially powerful in both the short and the long run. A very large effect size ($r = .40$ or greater) in the context of psychological research is likely to be a gross overestimate that will rarely be found in a large sample or in a replication.”

Therefore, the between-group effects for the Non-Distressing Tinnitus group might be labelled very small for the explanation of single events but potentially consequential over a short period of time (i.e. .05-.2). Whereas the effect sizes for the Distressing Tinnitus Group range from .2 to .33, and should therefore be considered as medium to large effects that are potentially powerful in the short-term, but almost certainly important when considered cumulatively over a long period of time.

4.6 Conclusions

Advanced analyses and modelling of the experimental data presented in this chapter have established independent associations between executive functioning and Distressing Tinnitus. Although participants with Non-Distressing Tinnitus also showed subtle deficits in some tasks compared to the No Tinnitus group, this was generally

better accounted for by other variables (i.e. stress, anxiety, and hearing thresholds). Furthermore, fluid and switching ability were likely associated with improved performance for the Non-Distressing Tinnitus group after accounting for covariates. Analysis of subjective measures also demonstrated opposing patterns of results between Distressing and Non-Distressing Tinnitus groups for rumination and everyday cognitive failures, while both groups had poorer cognitive reserve compared to those with No Tinnitus (after age was account for). This chapter also considered the expected size of observed effects relative to those observed in the literature, and to daily concentration difficulties reported by people with distressing tinnitus. Taken together, the patterns of differences between Non-Distressing and Distressing Tinnitus groups are suggestive of a causal role for executive functioning in tinnitus distress that should be investigated in future research.

5 Chapter 5 – Preliminary validation of a causal model of tinnitus distress

5.1 Introduction

The purpose of this chapter is to expand upon the causal relationships suggested between tinnitus distress and cognitive performance by presenting a novel causal model of this relationship: the Executive Disruption Model (EDM). A conceptual overview of the EDM will be provided that incorporates the experimental findings described in previous chapters, as well as findings from the tinnitus and relevant wider literatures. I will then present a rigorous methodological framework (Structural Causal Modelling) and use it to obtain testable implications of the EDM. Finally, an analysis of these testable implications will be performed using data from the previous chapter and an independent dataset, to serve as a preliminary validation of the conceptual model. Presentation of the experimental data will follow brief reviews and overviews of the relevant conceptual and causal modelling elements that underpin the subsequent analyses.

5.2 From conceptual to causal cognitive models of tinnitus distress

Section 1.5.1 introduced various models pertaining to tinnitus and cognitive performance (Andersson and McKenna, 2006; McKenna et al., 2014; Mohamad et al., 2016; Trevis et al., 2016). At this stage it is worth noting that these models can be viewed as purely conceptual models and/or scientifically testable models. Conceptual models are a composition of concepts used to help people know or understand the subject that the model represents. They are essentially graphical expositions (usually a limited number) of the authors' assumptions regarding the data generating process. A conceptual model is typically imprecisely defined and may not be scientifically testable (Figure 5.1).

McKenna et al. (2014) suggest that 'for a model to be regarded as scientific, there must

be somewhat of testing whether it is false and so a model should at least provide testable hypotheses about a particular process or mechanism'. This viewpoint of science is associated with the falsificationist philosophical view of science popularised by Karl Popper, though others exist (Dienes et al., 2008; McElreath, 2016).

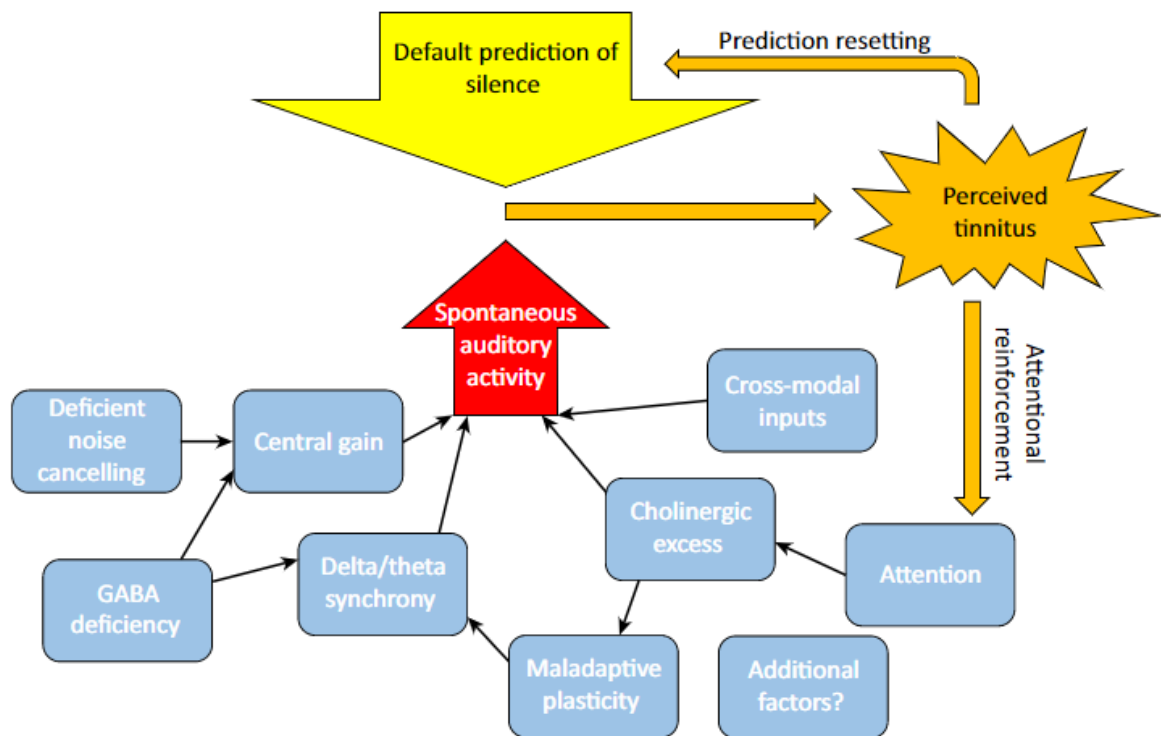


Figure 5—1. Conceptual pathophysiological model of tinnitus generation and maintenance from Sedley et al. (2016).

When considering models of tinnitus and cognitive performance, there has generally been a progression within the field from purely conceptual to scientifically testable models (e.g. Figure 5.2). Andersson and McKenna (2006) published the first discussion of tinnitus and elements of cognitive performance. Due to the early stage of research in this field, their discussion was based on the limited behavioural evidence that was available at the time and their clinical observation/intuitions about tinnitus and cognition. They provided an accompanying conceptual model that is not obviously scientifically testable. However, in the intervening years, several potentially scientifically testable models have emerged.

Trevis et al. (2016) presented a 'neurocognitive' conceptual model that also outlines scrutable causal claims. The model shows decreased cognitive control, which is associated with experiencing negative emotions, causing attentional failure, which in turn, is the cause of chronic tinnitus. However, this model has questionable validity. First, chronic tinnitus is not necessarily distressing tinnitus, which is evidenced by the majority of people with chronic tinnitus do not report being distressed or having significant concentration difficulties. Second, results from the EFiT study suggested that people without distressing tinnitus may have improved executive functioning compared to people with Distressing Tinnitus and No Tinnitus, suggesting that chronic tinnitus may be associated with enhanced cognitive control in some individuals. Trevis and colleagues model also suggests that cognitive components act as a mechanism for the perpetuation (chronification) of tinnitus, and it cannot therefore be considered a cognitive model of tinnitus distress per se.

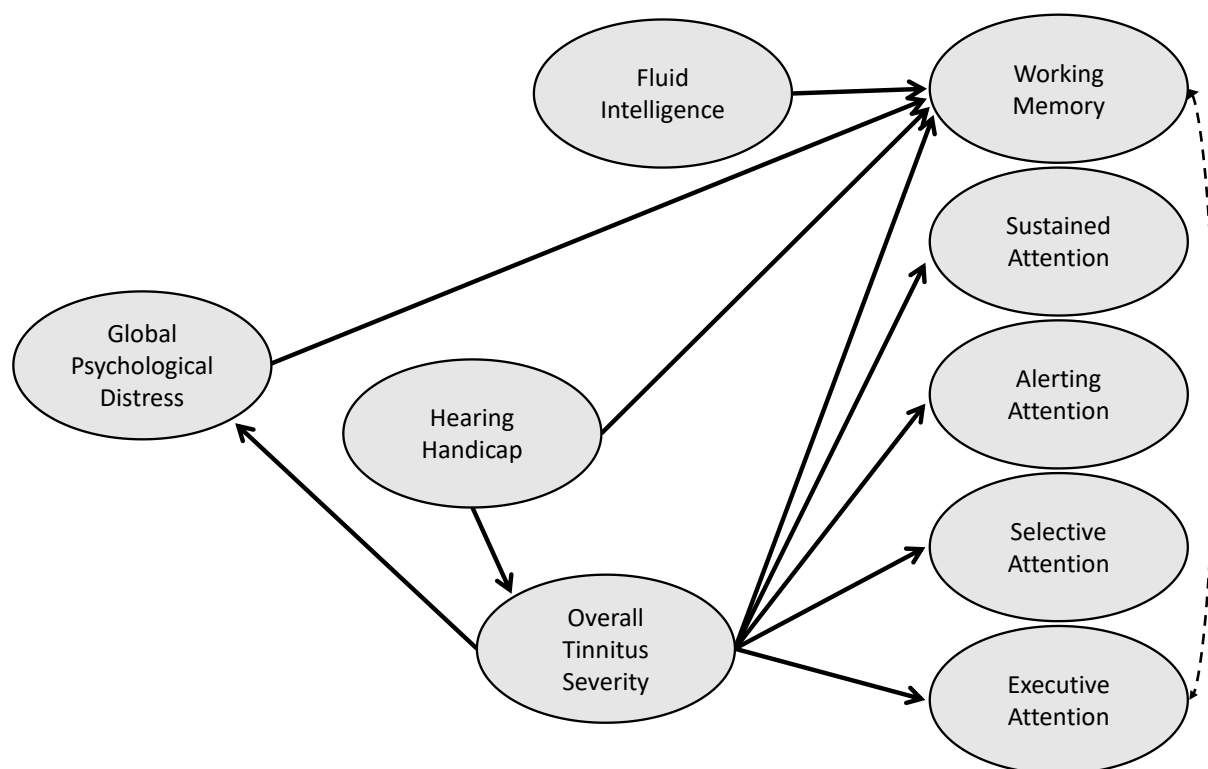


Figure 5—2. Causal model of the effect of tinnitus on cognitive performance adapted from Mohamad et al. (2016).

McKenna et al. (2014) and Mohamad et al. (2016) have provided cognitive models of tinnitus distress that were designed to be testable through structural equation modelling (SEM). When discussing the importance of scientifically testable models, McKenna et al. (2014) suggested that SEM can be used 'to test both goodness-of-fit of a theoretical model to observed data and to test the validity of causal structures'. However, there are technical limitations to SEM that limit its ability to scientifically validate a conceptual model. For example, SEM provides a test of global model fit and only allow evaluation of entire models through provision of a fit statistic, which is then compared to known statistical distributions. Furthermore, McElreath (2016) notes that an unknown number of process (i.e. causal) models could generate the observed data, so although a causal structure may be validated in the sense of being a consistent fit with the data, there are an unknown number of such models (Raykov and Penev, 1999).

This point is illustrated in Handscomb et al. (2019), who performed a confirmatory path analysis study to assess how well different configurations of McKenna et al's. (2014) 'Cognitive' model fit questionnaire data from people with tinnitus. The study found that five out of six path models reached the threshold for adequate fit. Testing causal validity of a proposed process model cannot therefore be indicated through fit indices alone. Moreover, McElreath (2016) notes that scientific models are not hypotheses per se. Testable hypotheses follow from competing proposed process models. Additional methods of testing and consensus are therefore required to test the scientific validity of a proposed conceptual model.

Structural causal models (SCMs) offer an alternative framework for creating models with scientific testability. The SCM framework was conceived following work on Bayesian networks in the field of artificial intelligence, designed from the perspective of how an intelligent agent may conduct inferential analysis (Pearl and Mackenzie, 2018). SCMs use domain specific knowledge to specify assumptions (typically in graphical form), and build a causal model with testable implications (Figure 5.3).

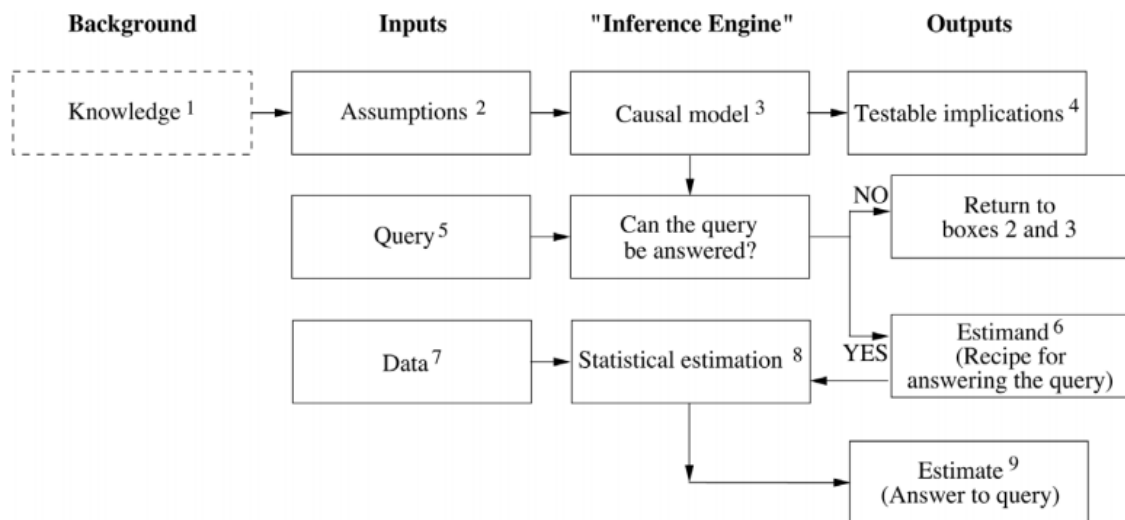


Figure 5—3. Illustration of how an 'inference engine' combines data with causal knowledge to answer queries of interest (Pearl and Mackenzie, 2018).

Analytic principals underpinning SCMs are discussed further in Section 5.4, but the SCM framework is generally considered to be advantageous to SEMs for several technical reasons. First, SEMs inherently make strict linear assumptions about data. Due to their non-parametric nature, no assumptions regarding distributional or functional form need to be made for SCMs. Furthermore, if any parameter within an SEM cannot be estimated (e.g. if error terms are correlated, or some variables are unobserved), then the joint distribution of the model cannot be estimated, and the model cannot be tested at all. However, for SCMs, if a coefficient cannot be identified in one area of the model, incomplete information about the rest of the model can still be obtained. This is beneficial because when a model is tested globally, measurement of a large number of variables with accompanying noise and/or sampling variation can make subsequent tests unreliable.

These technical advantages provide several attractive features for modelling the relationship between cognitive performance and tinnitus distress. As conceptual models are typically poorly defined, SCMs provide a systematic and rigorous means of mapping them to a model with testable implications through directed acyclic graphs (DAGs). Because SEMs and DAGs are both path models, competing process models such as those

suggested by McKenna et al. (2014) and Mohamad et al. (2016) could be tested in an SCM framework. Furthermore, cognitive models of tinnitus distress are generally complex, which is problematic for replication and comparisons across studies because SEMs require an accurate specification and measurement of every variable within the model. In contrast, SCMs can facilitate replication and comparisons across studies with incomplete information from a model. Therefore, SCMs increase the opportunity for local elements of a model to be tested in independent datasets. Furthermore, validated SCMs can be expanded using do-calculus (Pearl and Mackenzie, 2018), to test hypotheses that are not possible from observational data without a causal model (i.e. interventional and counterfactual queries).

Moving towards causal models of tinnitus distress and cognitive performance is crucial because understanding which variables “listen” to one another is essential for successful intervention development. While various interventional options are available (see Section 6.4.3), the causal structure of a model has implications regarding the likely success of a given intervention. For example cognitive training may do little to affect tinnitus distress if the cause of the problem is solely a loud tinnitus percept, however, if tinnitus distress is caused through inherent and/or temporary disruption of executive abilities, then such interventions would be more likely to reduce an individual’s tinnitus distress.

The process of advancing from a conceptual to causal model is undertaken in the remainder of this chapter. First, a conceptual model is outlined that suggests that tinnitus distress is caused by executive ability. The theorised conceptual model is then mapped to a DAG to be analysed in an SCM framework. Finally, the causal model is validated by obtaining a causal estimate in two independent datasets, to provide a theoretically-derived and rigorous quantification of the effect of executive abilities on tinnitus distress.

5.3 Executive Disruption Model of Tinnitus Distress (EDM)

The Executive Disruption Model (EDM) is a conceptual model of tinnitus distress based on findings from this thesis and the wider tinnitus literature. Hallam et al. (1984) make a useful distinction regarding “loudness” of the tinnitus percept and subsequent “distress”. The causal interrelations between tinnitus distress, percept loudness, and cognitive abilities (specifically executive functions) underpin the EDM.

The EDM is depicted in Figure 5.4. The bottom triangle depicts a tinnitus percept of variable intensity across individuals. The percept is subsequently blocked by a person’s latent EF abilities shown by the middle triangle. An individual’s tinnitus distress is depicted by the top triangle and is posited to arise from the combination of these two elements. For example, at far left side, a quiet tinnitus percept is combined with a large latent executive ability that will impede the tinnitus percept from intruding, and subsequently cause minimum tinnitus distress. At the opposite ends of these spectrums, a loud tinnitus percept combined with reduced executive ability and result in maximum tinnitus distress.

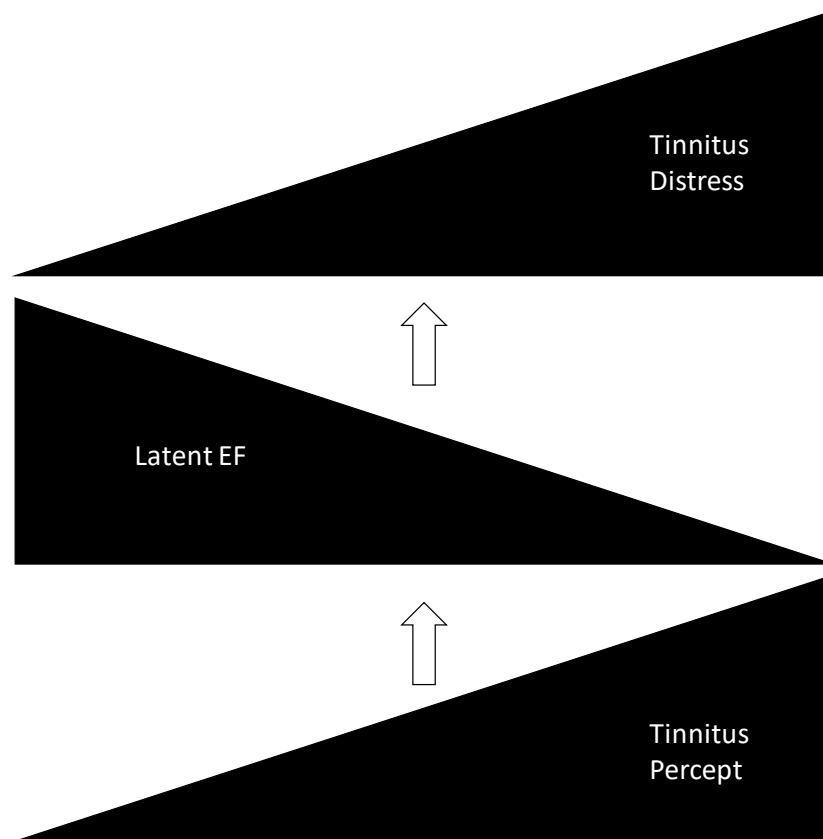


Figure 5—4. Illustration of Executive Disruption Model. A variable tinnitus percept is combined with variable latent EF and results in tinnitus distress.

The EDM proposes that tinnitus distress is caused by inherent variability and/or temporary disruption of latent executive abilities, and not solely by the percept per se. Execution functions can be compromised by various disruptive factors. Latent executive abilities are typically well established in the majority of people by the time they develop chronic tinnitus. They are known to develop throughout the life course, and by 9 months of age, there is evidence of attention control in infants. Development of executive abilities appears to be non-linear, with different subskills developing at different ages and having different development trajectories (Best and Miller, 2010; Gilmore and Cragg, 2018). By adulthood, individual differences are established in executive abilities (Friedman and Miyake, 2017), potentially predisposing some individuals to find a tinnitus percept intrusive.

Diamond notes (2013) that executive abilities are the first cognitive abilities to be compromised by stress and life events. Many of these disruptive factors are established

correlates of tinnitus distress, such as poor sleep and stress. Andersson and McKenna (2006) first suggested that cognitive factors, such as 'working memory deficits might worsen the *effects* of tinnitus, while still not being *caused* by tinnitus'. The idea that tinnitus distress may be worsened through other causes is a crucial insight. Figure 5.5 depicts how this disruptive effect might reduce the protective effect of latent executive abilities, resulting in relatively increased tinnitus distress.

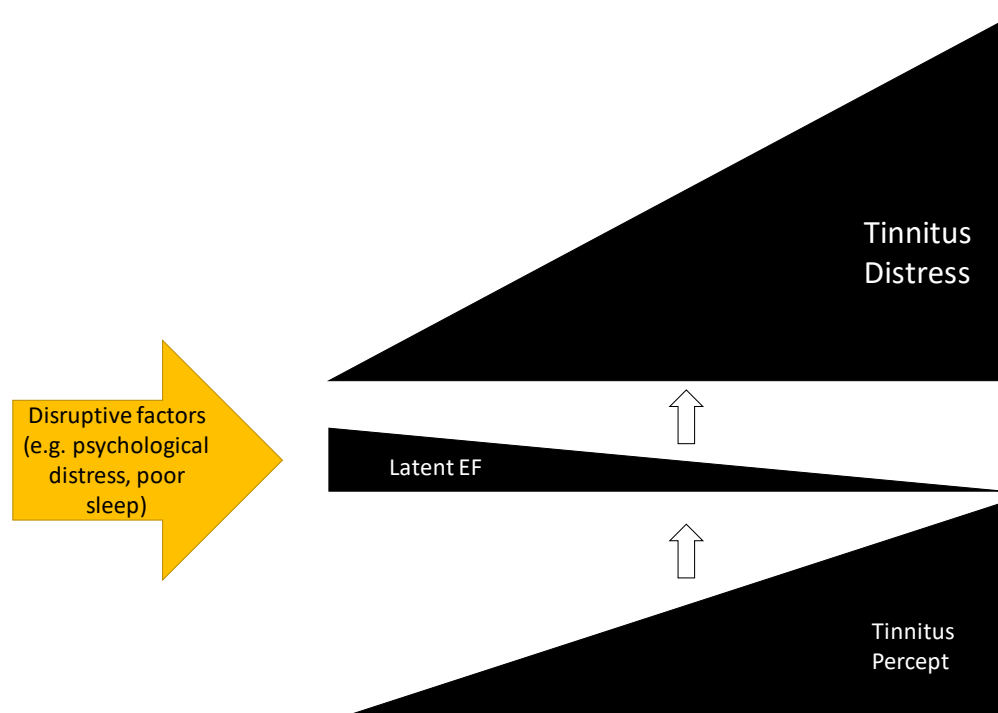


Figure 5—5. Illustration of compromised latent EF in Executive Disruption Model, depicting relatively more tinnitus distress for an unchanging percept following disrupted latent EF.

Tolerance of the percept (i.e. tinnitus distress) may alter depending on stress, depression, anxiety, or cognitive changes. This would lead to increased perception of tinnitus intrusiveness. By means of analogy with vision: if an individual had an entirely benign disturbance in their visual field (e.g. a simple floating shape). Would the disturbance be distracting if told not to pay attention to it? Would everyone be equally successful at ignoring it? Intuition suggests that some people would be more adept at ignoring this hypothetical visual intrusion. And even among individuals who typically

tolerated it, tolerance might fluctuate according to disruptive influences to their cognition, including stress, disturbed sleep, and cognitive changes (e.g. age-related).

5.3.1 Causal assumptions

Schmidt et al. (2014) note that ‘the development and evaluation of explanatory or “casual” models of psychological distress amongst persons with tinnitus has been limited’. In the EDM, the presumed causal nature of tinnitus distress is paramount, as many associations are known, their interrelations need to be specified and adequately accounted for to obtain causal estimates, hopefully enabling subsequent successful intervention. As the EDM is only a conceptual model, in the Section 5.5 it is mapped to a structural causal model (SCM), which provides a rigorous framework for addressing causality in complex models.

Hallam et al. (1984) observed that tinnitus ‘constitutes a complaint for only a proportion of people who experience it’, and note a ‘common assumption’ in tinnitus literature, which is that ‘the most annoying aspect of tinnitus is its loudness’. I argue that this is a causal assumption that generally persists within the tinnitus literature, and tacitly drives an assumption that a loud (annoying) sound *causes* concentration difficulties (Mohamad et al., 2016; Tegg-Quinn et al., 2016).

Tinnitus is typically conceptualised as a physical sound source, such as a tone or noise-like stimuli, which can presumably vary along the same dimensions as its physical analogue (i.e. frequency, amplitude, and timbre). In this conception, percept loudness would intuitively appear to be the most important variable with regards to causing concentration difficulties and decrements in cognitive performance (Hallam et al., 1984). There is often an unstated ‘loud sounds are distracting’ assumption (Figure 5.6), which often forms the basis of subsequent claims that tinnitus disrupts executive functioning (Heeren et al., 2014; Tegg-Quinn et al., 2016).

There is an intuitive appeal to such conceptualisations because it is trivial to demonstrate that certain sounds are distracting. Furthermore, the most common

measures of tinnitus distress are self-reported, and it is important to note that these measures also tacitly assume that the percept is *causing* the subsequent rating. For example, every question on the THI contains this assumption: 'Because of your tinnitus is it difficult for you to concentrate?', 'Does your tinnitus *make* it difficult to enjoy life?'

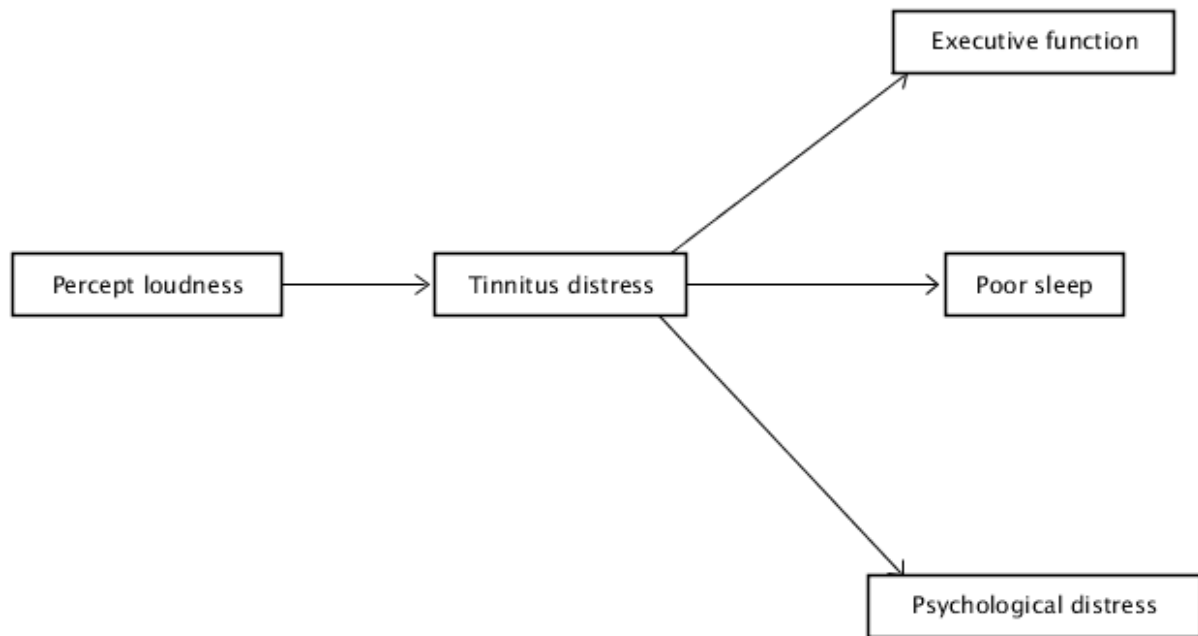


Figure 5—6. Causal model implicit in the 'loud sounds are distracting' conceptualisation.

The EDM suggests a different causal mechanism, where inherent and/or disrupted executive abilities facilitate intrusion from the tinnitus percept. The fundamental difference between these causal conceptions is illustrated in figures 5.6 and 5.7. Established associations within the tinnitus literature (e.g. psychological distress and poorer sleep) may be causal agents contributing to tinnitus distress through the general mechanism proposed in the EDM.

There is preliminary evidence for a causal role of executive ability causing tinnitus distress. This comes from literature concerning chronic pain, with which tinnitus is often compared (Peker and Sirin, 2016; Tonndorf, 1987). For example, Attal et al. (2014) prospectively sampled participants who were due to undergo knee surgery. The participants underwent cognitive performance testing prior to surgery. Performance on executive functioning tasks was predictive of which participants went on to develop

chronic pain after undergoing the surgery, suggesting that this factor had a causal role in subsequent pain perception.

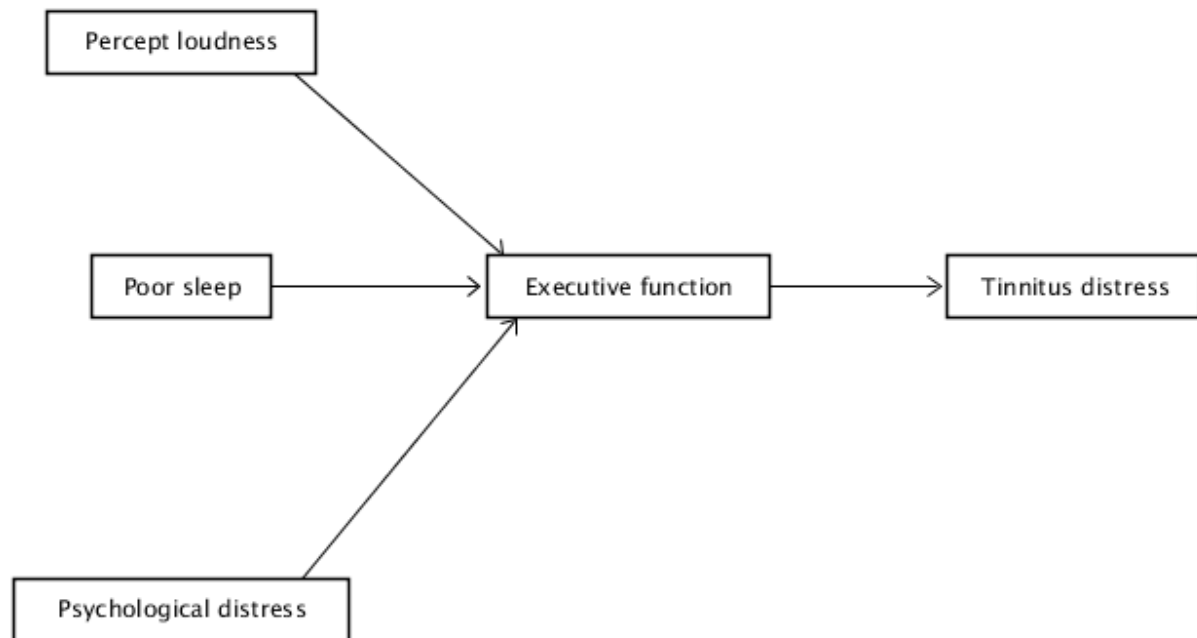


Figure 5—7. General causal model proposed by Executive Disruption Model.

Despite this, it is worth noting that, fundamentally, the EDM does not preclude a loud tinnitus percept having a disruptive influence on a person’s executive functioning (subsequently causing tinnitus distress). The model posits that inherent latent executive abilities may be particularly protective, therefore mitigating its potential for distress. However, the EDM accommodates an alternative, somewhat counterintuitive possibility: loudness of a tinnitus percept may be uniform across individuals (Figure 5.8).

5.3.2 Percept loudness disruption

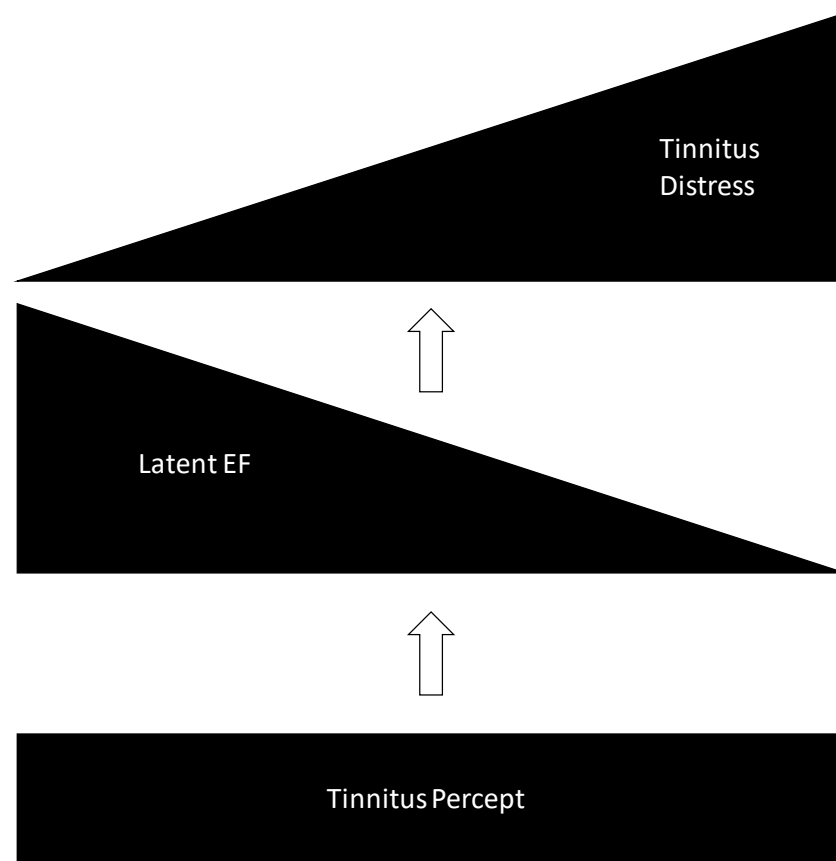


Figure 5—8. Illustration of variant Executive Disruption Model. A tinnitus percept of uniform loudness is combined with variable latent EF and results in variable tinnitus distress.

Most people with tinnitus do not find it distressing, and among those who do, loudness is not a good predictor of self-reported severity (Manning et al., 2019; Meikle et al., 1984; Moore, 2012). When typical loudness matching methods are used, tinnitus is recorded as having a surprisingly quiet sensation level (Moore, 2012). This has been viewed as a paradox for researchers: if the tinnitus is not that loud, why is it so distressing? When posing this question, tinnitus is framed as the causal agent, leading to a focus on the properties that might endow it with this ability (i.e. the LSD assumption). This causal assumption has led to many researchers concluding that lack of correspondence between percept loudness and tinnitus distress must be a technical issue or measurement error (Henry, 2016; Moore, 2012). Procedural difficulties with loudness matching methods and abnormal loudness growth (or recruitment) that may accompany hearing impairment

(and therefore often accompany tinnitus) are the presumed causes (Goodwin and Johnson, 1980; Moore, 2012).

In light of this, various alterations have been proposed to the scale on which psychoacoustic measurements of tinnitus loudness are made (Henry and Meikle, 2000). Moore (2012) has suggested that a loudness model should be invoked to correct the perceived tinnitus to a loudness level that would presumably be bothersome or distressing. However, the possibility remains that intensity measurements of the tinnitus percept have been reasonably accurate, and inherent and/or disrupted executive functioning results in increased tinnitus distress.

The relatively small number of people with severely distressing tinnitus may be considered evidence of the typical inability a loud percept to cause significant disruption to cognitive performance, though the technical problems of measuring tinnitus loudness may confound this assumption. In spite of this, evidence from the 'irrelevant sound effect' literature suggests that tinnitus-like sounds should not causally interfere with cognitive performance per se, (Banbury et al., 2001; Ellermeier and Zimmer, 2014).

In this literature, stimulus loudness has been demonstrated as being generally a non-disruptive variable to cognitive performance (though boundary conditions presumably apply to very intense sounds), and acoustic change has been shown to have the main causal effect on cognitive performance. Furthermore, this line of research shows that stimuli sounds which are steady-state and lack informational content are not typically disruptive to cognitive performance. This appears problematic for a conception of tinnitus as a disruptive agent to cognitive performance if tinnitus is generally a constant, relatively simple sound such as pure tones or narrowband noise. The possibility of a tinnitus percept of uniform intensity has been proposed previously, with Searchfield et al. (2012) suggesting that 'the intensity of the signal representing tinnitus may be constant but its contribution to tinnitus audibility and perceived magnitude will be determined by attention'. This puts focus on exactly how attention might factor into perceived tinnitus magnitude and/or distress (see Section 6.3).

At its core, the EDM accounts for the empirical evidence that loudness does not predict tinnitus distress very well, and is parsimonious in not requiring a corrective loudness model to adjust the tinnitus percept loudness to a presumably intrusive level. However a uniform percept intensity is not a fundamental element of the EDM, and the model is compatible with a tinnitus percept of variable intensity, whatever the “true” loudness of the tinnitus percept may be.

5.4 Structural Causal Modelling (SCM)

SCMs use domain specific knowledge to specify assumptions (typically in graphical form), to build causal models with testable implications. Data are “blind” to the underlying process that produced them (Imbens and Rubin, 2015; Pearl and Mackenzie, 2018). Understanding data generating, or causal processes, is arguably the primary goal of many sciences because scientists generally want to intervene on the system they are studying (Pearl et al., 2016).

Causal diagrams for regression analyses can be justified from a purely technical perspective. A known issue inherent in applied multiple regression analysis is the issue of multicollinearity between predictor variables. Causal diagrams consequently inform which variables should be specified within a model (Cohen et al., 2003; McElreath, 2016).

Cohen et al. (2003) is a widely-cited text on regression analysis within medical, social, and psychological sciences. An infrequently cited passage of this text is its endorsement of causal considerations to “make the underlying theoretical rationale and goals of the analysis as explicit as possible” (Cohen et al., 2003, p. 64). Furthermore, Cohen (2003, p. 64) also advocated the use of diagrammatic representation of causal models, a technique popularised in structural equation modelling (SEM).

Kline’s (2016) popular reference text on SEMs now introduces and advocates the SCM framework, arguing that it supersedes alternatives causal frameworks such as the

Potential Outcomes framework (Imbens and Rubin, 2015). Kline (2016, p. 165) also notes that its use in psychological research is desirable because:

- causal hypotheses are represented both graphically and simultaneously as a mathematical language that is subject to theorems, lemmas, and proofs
- assumptions in causal models are communicated in a precise language
- use of SCM distinguishes between questions that can be answered empirically and those that cannot
- SCMs can determine what new measurements would be required to address an “unanswerable question”

Despite the many benefits of the SCM framework, limitations of this framework should also be considered relative to other modelling options. For example, although SCMs do not require data from every variable within the entire model to be identified and estimate the path of interest (unlike SEMs), SCMs do not provide an overall test of model fit through chi-square statistics and therefore do not provide a test of data fit relative to a null model (Kline, 2016). Though it should be noted that even within an SEM framework, a model exhibiting good fit still has many mathematically equivalent models that support different theoretical accounts (Hershberger and Marcoulides, 2013; Raykov and Penev, 1999).

An additional consideration regarding limitations of the SCM framework concerns the mapping of various competing theoretical accounts, or process models, to the statistical model of the SCM (McElreath, 2016). For example, although a process model may be specified and evidence derived from its usage in support of a particular theory, there are various other theories and process models that could be correct, specified, and then tested that imply a similar result. However, McElreath (2016) argues that this should be considered an epistemological concern of the scientific method more generally, and the SCM framework and DAGs are at least advantageous in providing a clear, readily-communicable, and explicit means of sharing theoretical assumptions with other researchers.

Some authors therefore argue that causation should be a central feature of scientific analyses within complex multivariate disciplines, such as epidemiology and psychology (Grosz et al., 2020; Hernán, 2018). They argue that this will improve the quality of observational research, and allow researchers to capitalise on recent technical advances such as causal path analysis with DAGs (Pearl, 2009; Pearl et al., 2016). DAGs provide a clear means of expressing SCM assumptions, and provide corollary rules for entering covariates within statistical models, which is important because unprincipled entry of covariates can lead to incorrect inference (see Section 5.4.2).

For this reason, DAGs are already commonplace in fields such as epidemiology and increasingly in economics (Pearce and Lawlor, 2016; Tennant et al., 2019), but there are now increasing examples of DAGs being used and promoted in various areas of healthcare and psychology, including respiratory, sleep, and critical care medicine, heart and renal medicine, and substance abuse (Lederer et al., 2019; Lewis and Kuerbis, 2016; Rohrer, 2018; Staplin et al., 2017; Tennant et al., 2019). The remainder of this section provides a brief non-technical primer in DAG terminology, and related fundamental concepts that underpin the causal path analysis presented in Section 5.5.

5.4.1 Directed acyclic graphs

A DAG is a way of describing qualitative causal relationships among variables and contains information that a purely statistical model does not. Unlike a statistical model, a DAG, encodes the consequences of intervening to change a variable (McElreath, 2016).

A DAG is an abstract structure that connects nodes with lines (or edges) which are directed arrows. In these structures a variable is not permitted to cause itself, hence it is acyclic (Rohrer, 2018).

5.4.1.1 Basic terminology

DAGs represent effects through sets of nodes (or vertices) and causal arrows (or directed edges/arcs). Arrows represent causal relations and an exposure and outcome variable are typically the focus of consideration. Nodes with an edge pointing to another

node are said to be parents of that node, while nodes that receive an edge are children of that node (Figure 5.9).

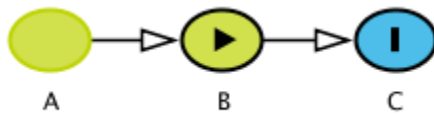


Figure 5—9. DAG where B is the exposure and C is the outcome. An ancestral relationship is also depicted, where B is a child of A and A is an ancestor of B.

5.4.1.2 Elementary causal data structures

All DAGs are comprised of only three structures, which are primitives for modelling increasingly complex relationships.

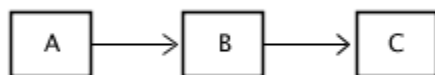


Figure 5—10. The elemental Chain DAG.

Chains have the structure $A \rightarrow B \rightarrow C$, and can transmit an association between the node at the beginning and a node at the end. In Figure 5.10 there is an open path between A and C through B, where A causes B, and B causes C.

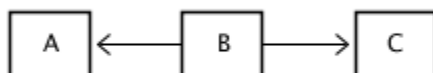


Figure 5—11. The elemental Fork DAG.

Forks have the structure $A \leftarrow B \rightarrow C$. Figure 5.11 depicts a forking non-causal open path between A and C through B, where B causes both A and C.

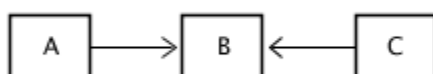


Figure 5—12. The elemental Collider DAG.

Colliders have the structure $A \rightarrow B \leftarrow C$ and are so-called because of both arrows pointing or 'colliding' on the same node. This closes the path through the collider. In Figure 5.12, the path between A and C is closed by the collider node B.

5.4.1.3 D-separation

When two variables in a DAG are independent conditional on a third, it is known as D-separation. Figure 5.9 shows the most basic example of D-separation, where knowledge about A gives you no extra knowledge about C once B is known. An illustrative example would be a causal path between fire, smoke, and a smoke alarm (Pearl and Mackenzie, 2018). If we know that a smoke alarm is active because there is smoke in the room (i.e. conditional on this effect), then knowing if there is also a fire will provide no extra information on whether the smoke alarm will be active.

DAGs can therefore be viewed as encoding a "flow" of information from node to node through arrows. The elemental structures describe open and closed paths through which information can flow. An open path allows for informational flow, while closed paths block the flow. When a path is open, variables are said to be D-connected. Information can flow along a path regardless of which direction the arrows are facing. The only exception are collider nodes, which are naturally closed paths and stop informational flow. When all paths between two nodes are closed, the nodes are referred to as D-separated.

5.4.1.4 Confounding

Rohrer (2018) makes the useful distinction between causal and non-causal paths (or spurious correlations); a path that contains at least one fork or collider can transmit a non-causal association. Confounding therefore occurs when open paths bias the estimate of the relationship between two nodes

5.4.1.5 Backdoor paths

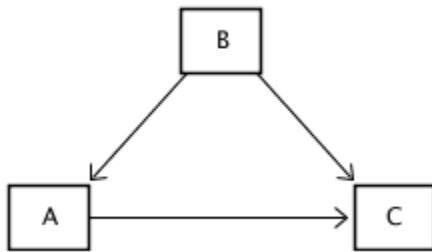


Figure 5—13. A DAG showing an open non-causal backdoor path from A to C through B.

Confounding commonly occurs when a backdoor path is left open, enabling a flow of spurious associational information. Figure 5.13 shows a backdoor path that biases the estimate of A on C because B is a parent of A and C. B is therefore a confounder of the relationship between A and C in this example.

5.4.1.6 Blocking by conditioning

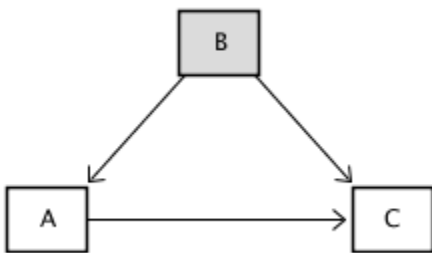


Figure 5—14. A DAG showing a closed non-causal backdoor path from A to C through B.

In Figure 5.14, the backdoor path between A and C has been blocked by conditioning on B. This could be achieved, for example, by including B as a statistical control within a regression model.

5.4.1.7 Collider bias

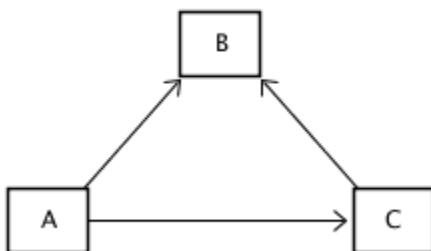


Figure 5—15. A DAG with a closed backdoor path between A and C as B is a collider.

However, conditioning on an inappropriate variable can actually introduce confounding into a relationship. Collider nodes are naturally closed paths. Conditioning on a collider opens the path, and therefore lead to biased and inaccurate conclusions (see Section 5.4.2). Figure 5.15 shows an example where the backdoor path is naturally closed because B is a collider. Unnecessarily conditioning on B would open a non-causal backdoor path, biasing our estimate of the effect of A on C by introducing information from B (Figure 5.16).

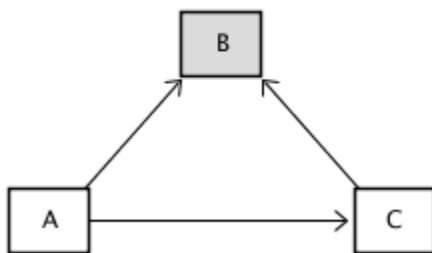


Figure 5—16. . A DAG with an open non-causal backdoor path between A and C because of conditioning on the collider B.

Collider bias can even occur when conditioning on a descendant of a collider. Figure 5.17 shows an example where conditioning on D has opened a non-causal path that confounds the path between A and C because D is a descendant of the collider B.

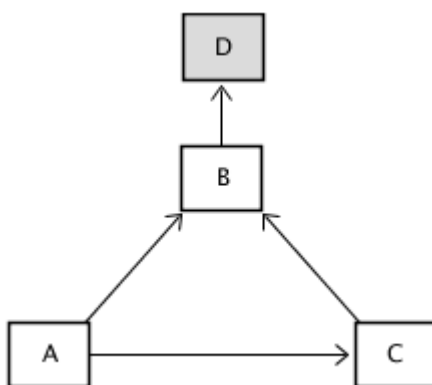


Figure 5—17. A DAG with an open non-causal backdoor path between A and C because of conditioning on D, which is a descendant of the collider B.

5.4.2 Statistical “control” can be misleading

Concerning statistical control of observational data, Rohrer (2018) notes that ‘it has become ‘common practise to include as many covariates as possible’ and that ‘it is not true that simply adding more covariates will improve the estimate of a causal effect’. Inference cannot be made purely with data (McElreath, 2016; Pearl and Mackenzie, 2018); therefore, “control” variables should not be unnecessarily added to statistical models. Doing so can provide inaccurate estimates and conclusions, and can even appear paradoxical.

Simpson’s paradox (or Simpson’s reversal) is an example of when unnecessarily conditioning on a variable can lead to the direction of an association switching based on level of analysis. (Pearl and Mackenzie, 2018). While Berkson’s paradox is an example of when conditioning on a variable can lead to apparent correlations that are not present in population in which we are actually interested (McElreath, 2016). Figure 5.18 shows results of simulation study, where selecting the top 10% of studies from a research proposal process based on trustworthiness and newsworthiness can introduce a spurious association between these variables.

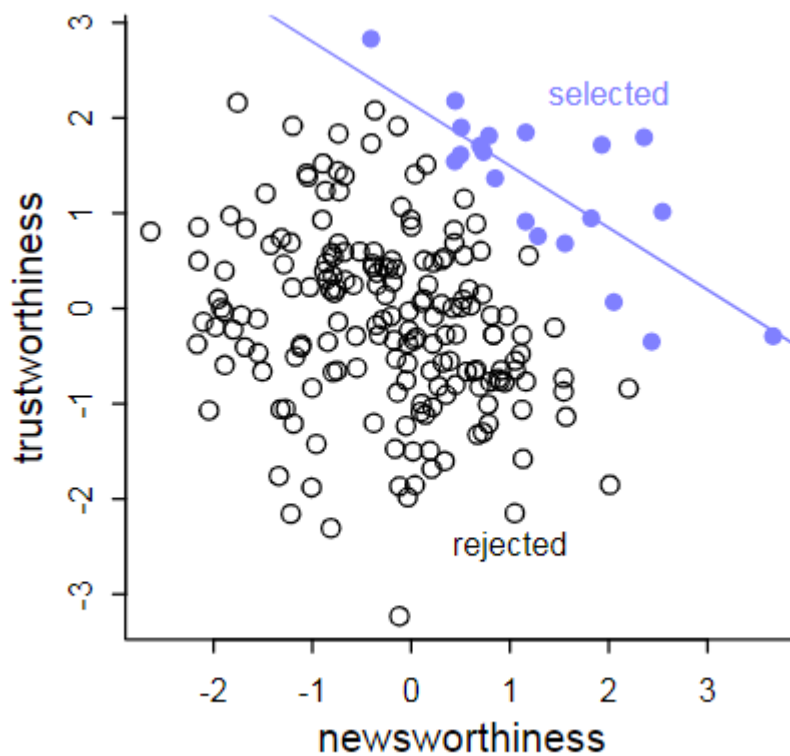


Figure 5—18. Why the most newsworthy studies might be the least trustworthy. 200 research proposals are ranked by combined trustworthiness and newsworthiness. The top 10% are selected for funding. While there is no correlation before selection, the two criteria are strongly negatively correlated after selection (McElreath, 2016).

5.4.3 Minimum adjustment sets

The key point to take away from the previous section is that it is not advisable to include every measured covariate within a statistical model as it may actively hinder inferences concerning a researcher's question of interest (McElreath, 2016). A powerful feature of DAGs is that they inform researchers about which variables do or – equally importantly – do not require inclusion in proposed models to obtain a causal estimate of interest. These are known as minimum adjustment sets. In Figure 5.13, the minimum adjustment set to estimate the causal effect of $A \rightarrow C$ would be $\{B\}$.

5.5 Mapping the EDM to a DAG

The EDM was placed in the SCM framework by using Dagitty (Textor et al., 2016) to create a DAG of variables relevant to the EDM that were contained in both the EFiT dataset and an independent dataset. The aim of performing causal path analysis to compute a minimum adjustment set to estimate the primary causal effect of interest, where latent executive ability has a causal impact on tinnitus distress through having a protective effect against a potentially intrusive tinnitus percept.

5.5.1 Complete Variable DAG

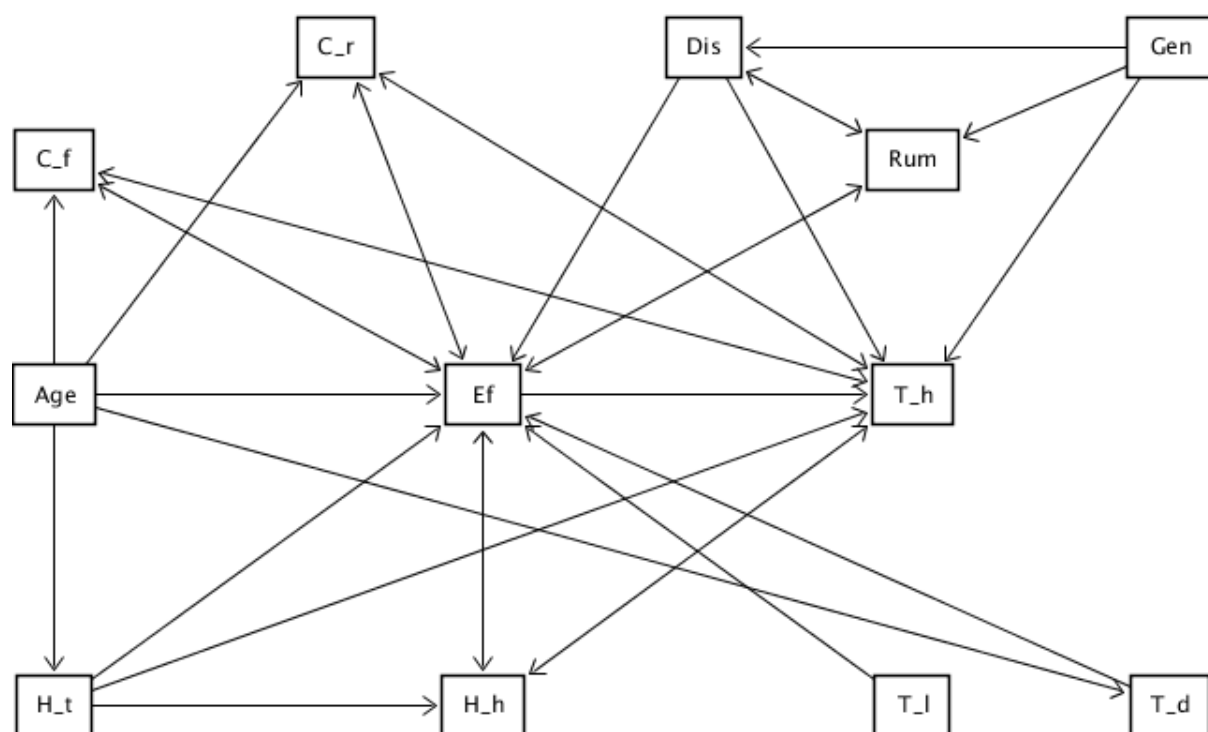


Figure 5—19. Full DAG of Executive Disruption Model of tinnitus distress. Age = age, C_f = self-perception of cognitive failure, C_r = cognitive reserve, Dis = psychological distress, Ef = executive function, Gen = gender, H_h = hearing handicap, H_t = hearing thresholds, T_d = tinnitus duration, T_h = tinnitus handicap, T_l = tinnitus laterality, T_p = tinnitus presence, Rum = rumination.

The DAG presented in Figure 5.19 represents my assumptions regarding the relationship between cognitive performance and tinnitus distress. The primary theoretical claim is that tinnitus distress is partly caused by latent executive abilities ($Ef \rightarrow T_h$). The other nodes within the model are specified based on logical causal relationships, insights from research in this thesis, and theoretical interrelations informed by the wider literature.

Logical causal relationships are primarily nodes on the periphery of the model and include relationships between age, cognition, and hearing thresholds. For example, while there is a well-known association between age and hearing thresholds, the causal relationship can be confidently specified with a unidirectional arrow ($\text{Age} \rightarrow \text{H}_t$) because hearing thresholds can never be the cause of a person's chronological age; neither can their self-perceived cognitive failures or executive abilities. Similarly, an individual's level of tinnitus distress cannot cause their gender, duration of tinnitus, or perceived tinnitus laterality. Hearing handicap cannot cause hearing thresholds. Global psychological distress or ruminative tendencies cannot cause gender.

Relevant nodes have been included in Figure 5.19 due to their demonstrated association with tinnitus distress in previous sections of this thesis. Clarke et al. (2020) demonstrated an association between executive functioning and tinnitus distress through meta-analysis. The EDM suggests a causal relationship, with executive abilities causing tinnitus distress (see Section 5.3). Clarke et al. (2020) also demonstrated an association between tinnitus distress and self-perceived cognitive failures. The EFiT study results presented in Chapter 3 also verified this association. However, the causal relationships between cognitive reserve, cognitive failures, rumination, and EF are not readily inferred and cannot be confidently specified in the model (e.g. people could experience more cognitive failures because of poorer EF, or an expectation of failures may impact EF performance). In these instances, the model is agnostic about plausible but uncertain causal relationships, and simply specifies an association with a bidirectional arrow. For example, tinnitus distress shares an association with global psychological distress ($\text{T}_h \leftrightarrow \text{Dis}$)

Theoretical interrelations for causal and associative relationships between various nodes are also drawn from the wider literature. Associations have been reported between rumination and executive function (Yang et al., 2017), as well as rumination and general psychological distress (Michl et al., 2013; Wilkinson et al., 2013). Gender differences in rumination have been reported (Johnson and Whisman, 2013), and a

causal relationship could be specified because rumination cannot cause gender. Associations have been reported between Hearing handicap and tinnitus distress (Henry, 2016; Ivansic et al., 2017); as the causal relationship between them is unknown, an association is specified. Executive function and hearing handicap share an association (Lee et al., 2019). Cuny et al. (2004) have suggested that the laterality of a tinnitus percept may direct attention towards to the tinnitus sound; however, EF cannot cause tinnitus laterality, a causal relationship could be specified. Diamond (2013) notes that executive functions are the first cognitive functions to 'suffer and suffer disproportionately' due to global psychological distress; a causal relationship is therefore specified. Finally, an individual's tinnitus duration may feasibly impact both their tinnitus distress and executive ability. Habituation to stimulus may lessen tinnitus distress over time, therefore, a causal relationship between these variables can be specified.

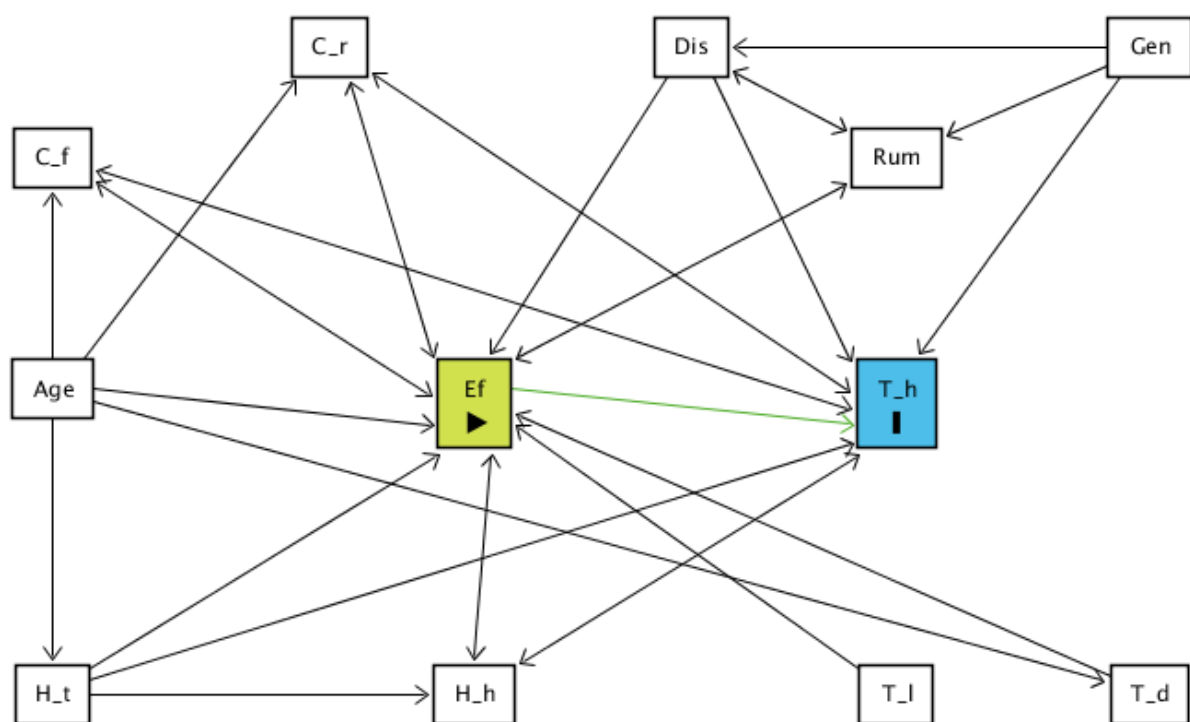


Figure 5—20. Full DAG of Executive Disruption Model of tinnitus distress showing causal analysis path of interest in green. Age = age, C_f = self-perception of cognitive failure, C_r = cognitive reserve, Dis = psychological distress, Ef = executive function, Gen = gender, H_h = hearing handicap, H_t = hearing thresholds, T_d = tinnitus duration, T_h = tinnitus handicap, T_l = tinnitus laterality, T_p = tinnitus presence, Rum = rumination.

Having specified the SCM, the analytic concepts outlined in Section 5.4 can be algorithmically applied to the DAG to provide the minimum adjustment set for the designated path to provide a causal estimate. Failing this, the causal path analysis also importantly indicates where no estimate is possible based on causal structure specified in the DAG. The direct causal path of interest (i.e. the effect of executive ability on tinnitus distress) is shown in Figure 5.20. Causal path analysis of the DAG shows that there is only one minimum adjustment set based on this model is where the effect of executive ability on tinnitus distress can be estimated after correcting for global psychological distress, and hearing thresholds (i.e. minimum adjustment set = {Dis, H_t}). It is noteworthy that age is absent from the minimum adjustment set in this model, and unnecessary inclusion of this covariate within a regression model could impact the accuracy of parameter estimates due to multicollinearity.

5.5.2 Unobserved Variable DAG

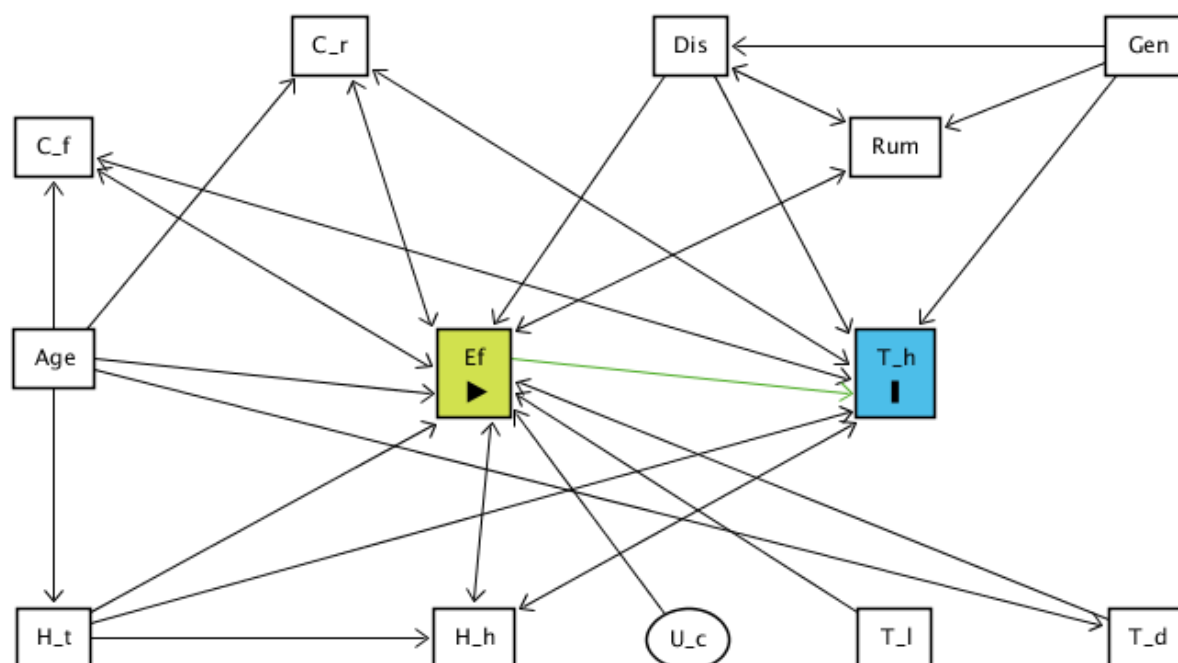


Figure 5—21. Full SCM of Executive Disruption Model of tinnitus distress showing causal analysis path of interest in green with added unmeasured confounds. Age = age, C_f = self-perception of cognitive failure, C_r = cognitive reserve, Dis = psychological distress, Ef = executive function, Gen = gender, H_h = hearing handicap, H_t = hearing thresholds, T_d = tinnitus duration, T_h = tinnitus handicap, T_l = tinnitus laterality, T_p = tinnitus presence, Rum = rumination, U_c = unmeasured confounds of tinnitus loudness and disrupted sleep.

An important feature of casual path analysis with DAGs is that they allow sensitivity and robustness checks to be performed on modelling assumptions; this may be done through adding or remove nodes and paths to demonstrate how the resulting model would impact the ability to estimate a causal effect. For example, in the context of the EDM, tinnitus loudness and disrupted sleep could both be considered important variables that could both affect cognitive performance, but were not measured in the EFiT study.

To understand the impact of variables within the EDM SCM, such variables were considered unmeasured confounds (U_c) and added to a DAG (Figure 5.21). The minimum adjustment set analysed for the DAG was then obtained. Causal path analysis of Figure 5.21 shows that, even with the addition of these variables, the estimate of the causal effect of E_f on T_h can still be performed with the same minimum adjustment set (i.e. minimum adjustment set = $\{Dis, H_t\}$).

5.6 Independent validation analysis

In this section, a preliminary validation of the EDM will be undertaken in two independent datasets by using latent measures of executive ability to estimate the effect of executive abilities on tinnitus handicap. Figure 5.22 shows a portion of the Complete Variable DAG (Figure 5.19), highlighting the variables that appear in both datasets. The minimum adjustment set for the causal path analysis in Figure 5.22 is equivalent to that presented in Figure 5.19 (i.e. minimum adjustment set = $\{Dis, H_t\}$).

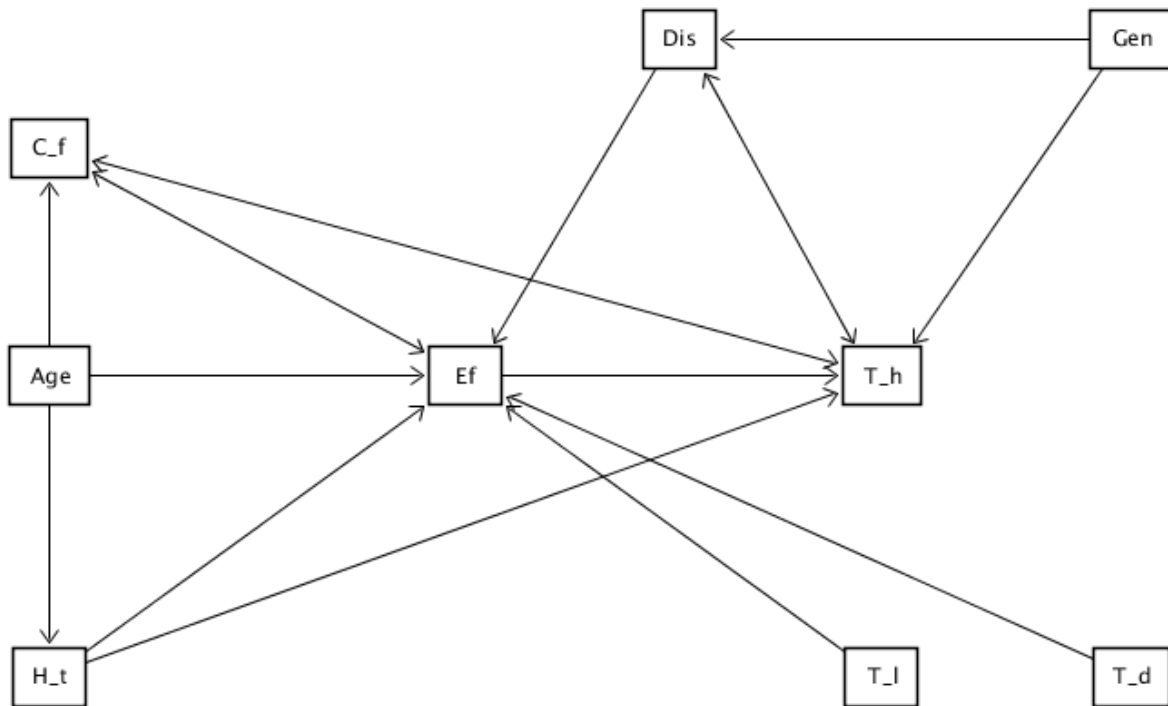


Figure 5—22. Overlap DAG highlighting the variables that appear in both Dataset 1 and Dataset 2.

5.6.1 Overview of datasets

Dataset 1 was the EFiT dataset presented in Chapters 3 & 4. Because the EDM is model of tinnitus distress, only participants with tinnitus were suitable for inclusion in the data presented in Chapters 3 and 4; therefore, participants from the No Tinnitus group were removed, leaving a total of 96 participants.

Dataset 2 was generated from experimental work for a PhD thesis from the Tinnitus and Hyperacusis laboratory at University of Nottingham (Mohamad, 2015). According to Mohamad, the study design was ‘a cross-sectional study of a tinnitus population recruited from the UK general public’; the primary study aim was ‘to define the relationships between tinnitus, working memory, selective, sustained and switching attention’. Mohamad subsequently tested relationships that they aimed to define using structural equation modelling. Mohamad recruited their study sample from clinical sites in the Derby and Nottingham area from 2014-2015. The study aimed to achieve a balanced sample of participants across tinnitus distress categories defined in the Tinnitus

Handicap Inventory (i.e. no/slight, mild, moderate, and severe). The final study sample featured 200 participants with a mean age of 59.5-years-old. Full descriptive characteristics of the sample are provided in Section 3.6.6 of Mohamad (2015, p. 166).

5.6.2 Principal axis factoring

Both datasets were checked for outliers and missing data prior to analysis. In each dataset, common factor analysis was performed in the psych package (Revelle, 2017) to extract shared variance between measures and obtain an estimate of the latent EF construct. Principal axis factoring (PAF) is a data reduction technique, similar to principal components analysis that extracts common variance from variables to obtain a measure of latent EF. PAF was performed on the most theoretically comparable EF measures from each dataset. In Dataset 1, these were the matrix reasoning, letter-number sequencing, and keep track tasks (Section 3.3.2.4.2). In Dataset 2, the measures were the raw correct score on matrix reasoning taken from the Weschler Abbreviated Scale of Intelligence – Second Edition (WASI-II), number of categories completed in Wisconsin Card Sorting Task, and the digit span backwards subset of the WAIS-IV (Milner, 1963; Weschler, 2011, 2010).

Table 5—1. Cognitive tasks and ascribed cognitive domains detailed in Mohamad (2015). WASI = Weschler Abbreviated Scale of Intelligence (2nd ed.), WAIS-IV = Weschler Adult Intelligence Scale (4th ed.), TEA = Test of Everyday Attention.

Task	Cognitive Domain
WASI-II (Matrix Reasoning)	Fluid Intelligence
Wisconsin Card Sorting Task	Switching Attention
WAIS-IV (Digit Span Backwards)	Working Memory
Corsi Block Tapping Task	Working Memory
TEA Subtest 1	Selective Attention
TEA Subtest 2	Sustained Attention
TEA Subtest 4	Switching Attention
TEA5 Subtest 5	Working Memory
TEA Subtest 6	Selective Attention
TEA Subtest 7	Sustained Attention

Table 5.1 outlines cognitive measures available in Dataset 2, with full details of the testing procedures and rationale for inclusion being provided in Mohamad (2015, pp. 116–124). These measures were identified as most appropriate for cross-validation due

to their theoretical similarity and overlap for representing EF across the datasets. A single factor PAF was specified (Figure 5.23), with factor loadings for all measures exceeding a minimum cut-off value of .4 (Matsunaga, 2010).

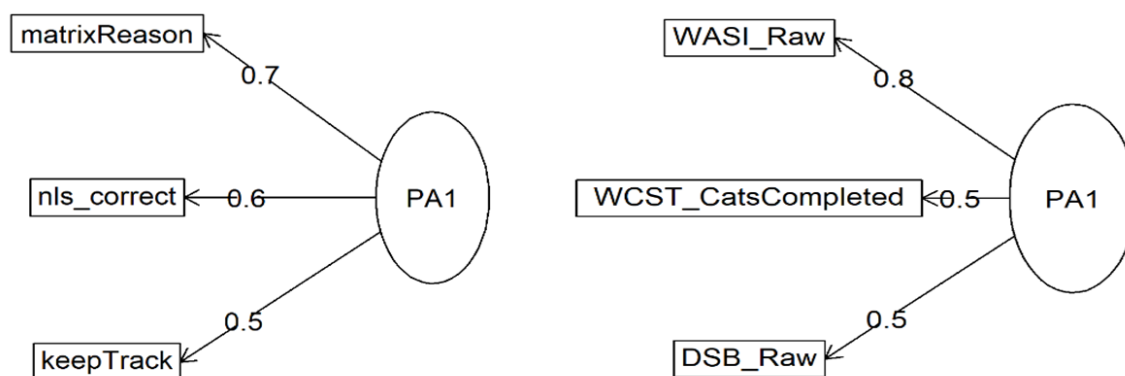


Figure 5—23. Principal axis factor from Dataset 1 (left) and Dataset 2 (right). Ovals depict latent executive factor, arrows show factor loading onto each task.

The Kaiser-Meyer-Olkin index was calculated to assess the appropriateness of applying PAF to both datasets, providing a measure of sampling adequacy (MSA) ranging from 0-1 where values <.5 are considered unacceptable (Kaiser and Rice, 1974). MSAs were acceptable in both datasets (>.6). PAF appropriateness was also assessed with Bartlett's test of sphericity, which calculates the statistical probability that the correlation matrix has significant correlations amongst the variables, and was statistically significant in both datasets ($p < .001$).

Matsunaga (2010) notes that 'latent factors should be determined primarily on the grounds of theoretical expectations and conceptualizations of the target construct'. In this instance, a single factor relating to EF was conceptualised and expected. However, many procedures have been created to provide indication of adequacy of a factor solution. Several of these were implemented to provide support for the appropriateness of a single factor solution; these included eigenvalues, scree tests, and parallel analysis (Revelle, 2013). Eigenvalues measure the amount of variation in the total sample accounted for by each factor. The Kaiser-Guttman criteria suggests that

eigenvalues exceeding 1 are viable factors and should be retained, as they explain more variance within the data than a single variable. Scree tests plot successive eigenvalues, and look for a sudden drop in eigenvalues or 'break' in the plot, which indicates the appropriate number of factors. Parallel analysis extracts factors until the eigenvalues of the real data are less than those of a random dataset of the same size. Eigenvalues for both datasets exceeded the Kaiser criteria, and both scree plots and parallel analysis supported the theoretical expectation and confirming of the appropriateness of a single factor solution.

5.6.3 Factor score regression (FSR)

Factor scores are the latent (composite) scores for each subject on each factor, which can then be used for further analysis in place of measured variables. This is known as factor score regression or FSR (Odum, 2011). Factor scores were computed using refined and non-refined methods. Non-refined methods use a simple, cumulative procedure to provide information about an individual's placement on a factor distribution, while refined methods are more exact and provide estimates that are standardised scores (Distefano et al., 2008). Various refined factor score methods exist, including regression or 'Thurstone' scores, 'Bartlett' scores, and Anderson-Rubin scores, each providing slightly different estimates of factor scores (Distefano et al., 2008; Revelle, 2013). When used in FSR, different factor score methods may produce biased parameter estimates when used in further regression modelling, however, unbiased/corrected estimates can be obtained through various approaches, including by using the Thurstone scoring method when using factor scores as independent variables in regression models (Devlieger and Rosseel, 2019; Skrondal and Laake, 2001).

As a sensitivity check, factor scores were computed with both Thurstone and Bartlett scoring methods, which providing no substantive difference; therefore, FSRs with Thurstone factor scores are reported in line with recommendations in Skrondal and Laake (2001). FSR was performed using the minimum adjustment set obtained from

path analysis of the SCM of tinnitus distress (i.e. {H_t, Dis}) to estimate the effect of executive functioning on tinnitus handicap in both datasets.

5.6.4 Results

The frequentist FSR GLM results for datasets 1 and 2 are shown in Table 5.2. The estimates for the factor scores in the two datasets were highly congruent. Dataset 1 had a partial beta coefficient of -3.56 ($\beta = -.015$, $p = 0.12$), while dataset 2 had a partial beta coefficient of -3.74 ($\beta = -.016$, $p = 0.02$).

Table 5—2. Frequentist factor score regression parameter estimates for Datasets 1 & 2.

term	estimate	std.error	statistic	p.value	dataset
(Intercept)	0.00	0.08	0.00	1.00	1
efit_factor_scores	-0.15	0.10	-1.55	0.12	1
scale(hearThresh)	-0.11	0.08	-1.37	0.17	1
scale(glo_psych)	0.63	0.08	8.02	0.00	1
(Intercept)	35.74	1.79	19.91	0.00	1
efit_factor_scores	-3.56	2.29	-1.55	0.12	1
scale(hearThresh)	-2.57	1.88	-1.37	0.17	1
scale(glo_psych)	14.95	1.86	8.02	0.00	1
(Intercept)	0.00	0.06	0.00	1.00	2
nm_factor_score	-0.16	0.07	-2.35	0.02	2
scale(PTA_Better)	0.02	0.06	0.33	0.74	2
scale(core_total)	0.59	0.06	10.34	0.00	2
(Intercept)	35.78	1.30	27.61	0.00	2
nm_factor_score	-3.74	1.59	-2.35	0.02	2
scale(PTA_Better)	0.43	1.32	0.33	0.74	2
scale(core_total)	13.59	1.31	10.34	0.00	2

The Bayesian GLM results for datasets 1 and 2 are shown in Table 5.3. The estimates for the factor scores in both datasets are in agreement with the frequentist GLMs. However, Bayesian GLMs allow the calculation of the probability that the factor score has non-zero effect (i.e. does the latent factor have a negative effect on tinnitus handicap). The probability of a non-zero effect in Dataset 1 was very high (.93), and virtually certain in Dataset 2 (.99).

Table 5—3. Bayesian factor score regression parameter estimates for Datasets 1 & 2

term	estimate	std.error	dataset	EF_less_than_0
(Intercept)	0.00	0.07	1	0.93
scale(glo_psych)	0.63	0.08	1	0.93
efit_factor_scores	-0.15	0.10	1	0.93
scale(hearThresh)	-0.11	0.08	1	0.93
(Intercept)	35.74	1.86	1	0.93
scale(glo_psych)	14.88	1.79	1	0.93
efit_factor_scores	-3.58	2.30	1	0.93
scale(hearThresh)	-2.61	1.89	1	0.93
(Intercept)	0.00	0.06	2	0.99
core_total_scaled	0.58	0.06	2	0.99
nm_factor_score	-0.16	0.07	2	0.99
PTA_Better_scaled	0.02	0.06	2	0.99
(Intercept)	17.77	2.54	2	0.99
core_total	0.75	0.07	2	0.99
nm_factor_score	-3.71	1.64	2	0.99
PTA_Better	0.02	0.07	2	0.99

This analysis provided initial validation of the Executive Disruption Model of tinnitus distress by providing identical estimates of the causal effect of EF on THI. The EDM model was converted to an SCM, and path analysis was used to identify the minimum adjustment set to estimate the causal effect of EF on THI. In two independent datasets, a latent EF factor score was created for participants using theoretically comparable measures. FSR then was used to provide a casual estimate of EF on THI using the minimum adjustment set obtained from a SCM following causal path analysis. The causal estimates of the latent factor were essentially identical in both datasets, with a 93-99% certainty of a negative effect of EF on tinnitus distress.

5.7 Study strengths and limitations

This modelling work was robust in being able to provide an estimate of the latent executive ability in each dataset through the use of common variance across multiple cognitive measures. However, although these measures were adequate in each dataset, they were not theoretically optimal and could only test broad executive functioning. A priori selection of measures for this purpose could improve measurement further and enable testing of specific executive functions.

A further limitation stems from the partial overlap of variables that were available between the two datasets. This study was, therefore, only able to examine a limited part of the proposed conceptual model. However, by using a validated and replicable methodology, this study provided a consistent result across independent datasets. The flexibility and reproducibility of the methods will enable this causal model to be transferred to new datasets for future studies.

5.8 Summary

This chapter began by drawing a distinction between conceptual and causal models of tinnitus and cognitive performance. I argued that although the field has generally progressed towards understanding the need for causal models, there are limitations inherent in previous approaches that can be improved upon through the use of a rigorous methodological framework (i.e. Structural Causal Modelling).

I then explicated a conceptual model of tinnitus distress, the Executive Disruption Model (EDM), which is grounded in findings from previous chapters of this thesis, combined with established findings from the tinnitus and wider literatures. The EDM posits that inherent variability and/or temporary disruption of latent executive abilities causes tinnitus distress. To investigate this conceptual model, the EDM was mapped to a directed acyclic graph and subjected to causal path analysis to compute the necessary variables required to estimate the causal effect of executive ability on tinnitus distress (i.e. the minimum adjustment set).

Preliminary validation of the EDM was performed by using the minimum adjustment set derived from the SCM to estimate the size of the causal effect, in both the EFIT dataset and an unpublished independent dataset containing a large sample of people with varying levels of tinnitus distress. To perform this analysis, the three best available theoretical measures of general executive ability were selected in each dataset. A data reduction technique (principal axis factoring) was then used to extract only

common variance from these tasks to obtain a measure of the latent executive ability factor.

Factor scores were subsequently generated for all participants (i.e. a weighted measure of latent executive ability). The factor scores were then used in Factor Score Regression analyses, in combination with the minimum adjustment set, to obtain a causal estimate of the effect of executive ability on tinnitus distress. The results of the analyses were essentially identical across the datasets, and showed that executive ability has a significant causal effect on tinnitus distress (i.e. as executive ability increases, tinnitus distress decreases).

The results of this analysis have provided preliminary validation of the EDM. This is important because it has demonstrated that the EDM produces highly consistent estimates of the effect of latent executive ability across independent datasets and cognitive measures, and supports this result with a clear underpinning theoretical model. The theoretically-derived result represents a progression towards a clear theory of the role of cognitive performance in tinnitus distress.

This progression is highlighted by the juxtaposition between the results of this chapter and the start of my thesis. My thesis began by highlighting the inconsistency of implicated cognitive domains associated with tinnitus and the variability of effect size estimates across studies, while this chapter presented consistent results in independent datasets, both in terms of the implicated cognitive domain and size of effect. This validation analysis has therefore demonstrated the potential of the EDM to advance theoretical and analytical understanding of tinnitus distress, which is essential to developing interventions to ameliorate it. The implications of this are considered in the final chapter.

6 Chapter 6 – General Discussion

6.1 Introduction

The purpose of this chapter is to discuss key findings from my thesis that contribute to knowledge concerning the relationship between tinnitus distress, concentration difficulties, and cognitive performance. I aim to achieve this by presenting and discussing key findings from my thesis in relation to current knowledge within the literature. I will also consider the role of executive attention in tinnitus more widely, as well as insights provided by the Executive Disruption Model (EDM). Finally, I will build on these discussions to look towards the implications and future work that may follow from the analytic methodologies employed, extent of analysis, and subsequent findings of my thesis.

6.2 Discussion of key findings

The systematic review and meta-analyses presented in Chapter 2 was the first comprehensive quantification of the association between tinnitus and cognitive performance (Clarke et al., 2020). Previous reviews had suggested that executive attention was effected by tinnitus (Mohamad et al., 2016; Tegg-Quinn et al., 2016), but also noted that findings had been inconsistent across the literature (i.e. a genuine association between tinnitus and cognitive performance was uncertain). The methods applied in Chapter 2 addressed previous limitations and reliably demonstrated that tinnitus is associated with poorer performance across multiple cognitive domains. Furthermore, the analysis provided consistency in establishing that the majority of cognitive tasks investigating a relationship between tinnitus and cognitive performance were executive functioning tasks.

However, Chapter 2 also highlighted the importance of potential confounding factors in this literature, by demonstrating that some reported associations between

tinnitus and cognitive performance may be spurious. For example, depression was found to be a moderating factor in studies that had measured the relationship between tinnitus and general short-term memory. This analysis also demonstrated that confounding had generally not been adequately considered within the literature to date, and showed the importance of introducing a “language” for confounding if advancements are to be made in this topic (see Section 6.4.2). Furthermore, the analysis also showed that tinnitus distress had typically not been well considered in previous research designs on this topic, as well as indicating promising domains of cognition to investigate that had received little research interest (i.e. fluid ability).

The experimental data presented in Chapter 3 was from a novel study designed to address confounding factors that were present in previous research, through focusing on the established association with executive functioning and investigating its relationship to tinnitus distress. Fluid ability has considerable overlap with executive functioning (Diamond, 2013; Duncan, 2010) but had not previously been investigated in relation to tinnitus distress. The study results demonstrated poorer performance across all domains of executive functioning, which suggested that cognitive performance generally worsens as a function of tinnitus distress (i.e. as tinnitus distress increases, performance gets steadily worse).

However, several confounding variables were noted in Chapter 3 that precluded the conclusion of an independent association between tinnitus distress and cognitive performance, and suggested that advanced methodological analyses were required. This is important because the original between-group analysis of the EFiT study was inspired by, and remains typical of research on this topic (Andersson et al., 2005; Jackson et al., 2014; Waechter and Brännström, 2015), and many new studies (including my own) are variations of existing designs in the literature. This suggested that inferences obtained from the data generated in such designs could be improved through the use of advanced analytical and statistical methods. A key finding in this chapter was that confounding

variables should not be considered equivalent in regards to confounding variation based on p-values and subsequently excluded from data analysis.

The importance of analytic methodology was highlighted in Chapter 4, as advanced modelling techniques were found to afford more robust conclusions of an independent association between distressing tinnitus and poorer cognitive performance, but able to demonstrate that modelling of confounding variables altered the *prima facie* interpretation of results described in Chapter 3 (i.e. that cognitive performance varies as a function of tinnitus distress). General linear models, Bayesian methods, and hierarchical linear models were able to demonstrate that non-distressing tinnitus is not typically independently associated with poorer executive functioning. In fact, non-distressing tinnitus was suggested to be independently associated with improved executive abilities in some domains, which suggested a causal role for executive functioning in distressing tinnitus.

The key finding of Chapter 5 was a demonstration that consistent estimates of the contribution of cognitive abilities to tinnitus distress were obtained through the conjunction of various theoretical and methodological considerations. These included consideration of appropriate theory, causal models, latent variable techniques, and validation in an independent dataset. A new conceptual model of tinnitus distress was proposed, which postulates a core role for executive abilities (the Executive Disruption Model or EDM). The implications of this model allowed the effect of executive abilities on tinnitus distress to be estimated consistently across independent datasets using a structural causal model (SCM).

Causality is proposed to be a fundamental language of our minds (Pearl and Mackenzie, 2018). I argue that appropriate consideration of causality is crucial to consider in the relationship between tinnitus and cognitive performance, because whether researchers are aware of it or not, we tend to default into causal discourse. Understanding the relationship between two latent variables is incredibly complicated. I propose that there is a widespread tacit assumption (originating from early peripheral

models of tinnitus – see Section 1.3.4) that the tinnitus percept causes poorer cognitive performance. The tacit conceptualisation of the relationship between tinnitus and cognitive performance is clear through the causal language that researchers generally slip into. For example, descriptions such as, “the impact of tinnitus severity on cognition”, “tinnitus and its effect on working memory and attention”, and “the impact of tinnitus upon cognition”(Mohamad et al., 2016; Rossiter et al., 2006; Tegg-Quinn et al., 2016).

The role of causation in relation to cognition and tinnitus distress is becoming increasingly understood to be an important issue (Leong et al., 2020; Schmidt et al., 2014). However, I argue that a confusion persists regarding the causal relationship between executive abilities and tinnitus distress (see section 6.3). Leong et al. (2020) suggests that ‘distress may be an important factor in the causal mechanism between tinnitus and attention’, but also highlight the default conception of the perceptual cause with the title ‘interruptive effect of tinnitus-related distress on attention’. While Nagaraj et al. (2020) suggest that ‘tinnitus disrupts the functioning of auditory working memory’, data in this thesis suggest that distress is a resultant factor from inherent/disrupted executive functions, with the EDM providing a logical, parsimonious causal role for this link that has been preliminarily validated across independent datasets.

To conclusively understand the causal role of the tinnitus percept in associated distress, it is important that future studies incorporate psychoacoustic measures of tinnitus loudness into their experimental design and data analysis, as this is frequently assumed to be the main driver of disruption. This has not been typical in the literature to date. For example, a recent study by Cardon et al. (2019) did consider percept loudness and suggested that it ‘influenced both the overall cognitive performance and Attention scores’. However, as this study used subjective tinnitus loudness (i.e. visual analogue scale or similar), they are likely to be a variant measure of perceived tinnitus distress (Henry, 2016).

The consequence of a cause is its effect, and a key finding from this thesis is that researchers should give due consideration to expectation of what is considered “significant” when looking at objective measures related to tinnitus and cognitive performance, both in terms of effect size and their potential implications outside of a laboratory setting (see Section 4.5). Between-group differences in cognitive task effect sizes presented in Chapter 3 are routinely considered small to medium if appraised through typical guidelines (Cohen, 1988; Correll et al., 2020); however, such guidelines are devoid of context, with their author noting that literature specific standards are preferable. My thesis presented an effect size distribution analysis that demonstrates that these guidelines effect sizes generally overestimate expectations. Clarke et al. (2020) noted the methodological limitations in the studies that were synthesised, which suggest that the actual effect sizes are directly attributable to having distressing tinnitus could well be smaller still.

Nonetheless, Funder and Ozer (2019) note that effect sizes should be considered within the context of both their research and real world implications. I propose that the laboratory-based differences that I have demonstrated for executive functioning should be viewed as a window into the day-to-day experience of people with distressing tinnitus. I suggest that it is reasonable to assume that many of these smaller effects can compound over the daily lives of people with distressing tinnitus, with disrupted and/or inherently poorer executive functioning, thereby causing an increased perception of concentration and attention difficulties, and the subsequent appraisal of a tinnitus percept that is more distressing/intrusive than is typical.

Additional insights from my thesis result from highlighting the limitations of inference from certain methods previously employed to investigate cognitive performance in people with tinnitus. Chapter 3 showed that p-value based group matching has been the predominant paradigm within the literature to date (Andersson et al., 2005, 2000; Jackson et al., 2014), and that this has resulted in inconsistent use of statistical control in subsequent analyses. Furthermore, this thesis also demonstrated

that more consistent estimates may be acquired through the use of latent variable techniques to measure executive functioning. Latent variable methods have been used previously (Miyake et al., 2000), and Mohamad et al. (2016) suggested that multiple cognitive measures should be employed.

Many previous studies have also attempted to experimentally dissociate tinnitus from typical covariates such as age, hearing loss, depression, anxiety, and stress by only recruiting those with subclinical hearing levels or young individuals (Degeest et al., 2017; Jackson et al., 2014; Nagaraj et al., 2020). However, it may be that these variables are generally not dissociable from distressing tinnitus. Therefore, in the relatively rare instances where researchers succeed in sampling this dissociation, a question would remain as to whether such a sample would be a valid representation of the general population with distressing tinnitus. Perhaps studying such dissociable samples with qualitative and psychoacoustic methods would provide insights as to what is unique about their tinnitus that it does not relate to hearing thresholds, generalised psychological distress, or age.

6.2.1 Support for established models

6.2.1.1 *Tinnitus cognitive models*

The results of this thesis and the wider literature may be evaluated and compared to various existing theoretical accounts that consider the relationship between tinnitus and cognitive performance (presented in Sections 1.5.1 and 1.5.2). The theoretical account provided by Mohamad et al. (2016) highlights on the Baddeley model of working memory (Baddeley, 2012). They suggest that cognitive performance detriments associated with having tinnitus represent a failure to stop attending to an 'essentially irrelevant signal'; this is compatible with McKenna's (1997) speculation that tinnitus may 'interfere with the capacity of the phonological loop to store auditory verbal information, in the same way as does for hearing loss'. It has therefore been suggested that having both tinnitus and hearing loss could separately 'add load to the central executive'.

The results of the experimental data analysis presented in Chapters 3 and 4 provide some support for this theoretical perspective. For example, the keep track task was chosen as a measure that was aligned with working memory, and participants in the Non-Distressing Tinnitus group had slightly lower average scores on this compared to the No Tinnitus group, even after adjusting for potential confounds, which could be interpreted as evidence for additional load resulting from the tinnitus signal. However, the updating focus of the keep track task means that a capacity measure of working memory may be a more theoretically relevant measure to provide evidence for this interpretation (Oberauer, 2019). Nonetheless, other authors have reached similar conclusions concerning the potential for additional cognitive load caused by the presence of the tinnitus percept (Khan and Husain, 2020), suggesting that various studies support this result (Rossiter et al., 2006; Stevens et al., 2007).

Similarly, the theoretical perspective of tinnitus causing disruption to the auditory verbal store could also be viewed as supported by some experimental findings from this thesis; for example, participants from the Non-Distressing Tinnitus Group had lower scores on the dual task measure, which was the only task that featured auditory stimuli. However, the absence of a difference between the Non-Distressing and No Tinnitus groups on the other related working memory task (letter-number sequencing) suggests that no additional load is caused by the presence of a tinnitus signal per se, and supports a theoretical account such as the EDM (Section 5.3), where poorer performance in the associated cognitive system would be a predisposing and contributory aspect in the experience of tinnitus distress.

McKenna et al. (2014) have proposed a theoretical perspective suggesting that tinnitus distress is caused by a conscious process of monitoring, where sustained attention is central to this account and failures of selective attention may be a contributing factor. Various studies have also implicated sustained monitoring and attention (Leong et al., 2020; Lima et al., 2019). Some aspects of the experimental data in this thesis may be interpreted as being consistent with this theoretical account, such

as the poorer performance in tasks by the Distressing Tinnitus group; however, it is challenging to explain the improved performance associated with the Non-Distressing Tinnitus group membership from this theoretical perspective (these challenges are discussed in Section 6.3).

The aforementioned observations also present challenges the theoretical account provided by Trevis et al. (2016) in their 'neurocognitive' model. This account suggests that chronic tinnitus is caused by a failure of the Cognitive Control Network (CCN). Various other theoretical perspectives have also implicated attention and related cognitive concepts as a contributory factor to the chronification of tinnitus (Roberts et al., 2013; Sedley et al., 2016), though Trevis and colleagues have provided the most explicit account of a theoretical model (Section 1.5.1.4). However, participants in this EFiT study described in Chapter 3 all had chronic tinnitus, while the data presented in this study demonstrated that those in the Non-Distressing Tinnitus group did not exhibit generalised failures of cognitive control as measured by the battery of core executive function tasks that were presented in Chapter 3. Therefore, a theoretical account of tinnitus that incorporates cognitive performance needs to highlight the central role of tinnitus distress, not the chronification of the percept (i.e., chronic tinnitus can be present without it being a measurable detriment to cognitive performance).

I previously suggested that the neurocognitive model has questionable validity (Section 5.2). In addition to experimental data from this thesis not supporting this model, there is also a theoretical aspect of this criticism that should be considered in future research: chronic tinnitus is not necessarily distressing tinnitus. The majority of people with chronic tinnitus do not report being distressed or having significant concentration difficulties or reporting detriments in cognitive performance. However, it should be noted that the theoretical position presented in the neurocognitive model may have been influenced by research that has suggested that 'non-help-seeking tinnitus sufferers' have detriments in cognitive performance (Jackson et al., 2014). Furthermore, the neurocognitive model also suggests that cognitive components act as a mechanism

for the chronification of tinnitus, meaning it cannot therefore be considered a cognitive model of tinnitus distress per se. This highlights the need to provide an adequate theoretical account for the role of cognition in tinnitus chronification in addition to tinnitus distress. Challenges associated with the incorporation of cognition (specifically attention) into contemporary theoretical accounts of tinnitus initiation, chronification, and distress are further discussed in Section 6.3.

6.2.1.2 General tinnitus models

The EDM may be considered complimentary to existing models of tinnitus by expanding on the nature of relationship between tinnitus and executive functions/attention. It aligns with established models through conceiving of the tinnitus generation and subsequent processing as a multi-stage process (McFerran et al., 2019), but departs in its mechanistic conceptualisation (Section 6.3). The following section provides a brief overview of how I perceive the results presented in this thesis to be compatible with previous research.

The 'neuropsychological model' (Jastreboff, 1990) is a prominent model of tinnitus. It suggests that initial generation of the tinnitus percept occurs in the auditory periphery, and posits that auditory and other cortical areas are involved in perception and conscious evaluation involving memory and attention. The work presented in this thesis appears consistent with separate stages for generation of the tinnitus related signal and proposed role for cortical areas (specifically prefrontal cortex through executive attention). Similarly, Rauschecker et al. (2010) describe a tinnitus signal originating in auditory pathways from lesion-induced plasticity that subsequently gains access to conscious perception through broken frontostriatal gating. Rauschecker et al. (2015) suggest this typically being 'conceptualised as an abnormal learning process' in terms of cognition. Various other models follow this signal, resultant negative thought dynamic (Cima et al., 2019; McKenna et al., 2014). Overall, the evidence presented in this thesis is consistent with elements of prominent contemporary models of tinnitus in providing support for models that differentiate a stimulus and its subsequent processing

that provide distress. However, the EDM posits a mechanism by which executive abilities interact with the tinnitus stimulus to facilitate tinnitus distress (Section 5.3). In this sense, the work presented in this thesis may be seen as a cognitive-behavioural link between the tinnitus percept and higher-level 'contents' of distressing thoughts concerning the percept.

The experimental data in this thesis also expands beyond executive attention. De Ridder et al. (2014a) suggested that the tinnitus percept may arise from parahippocampal auditory memory. Tinnitus may therefore be associated with long-term memory deficits, and the results from research synthesis and experimental work presented in this thesis may be viewed as behavioural evidence in support of this hypothesis. Clarke et al. (2020) demonstrated an association between having tinnitus and poorer cognitive performance on general learning and retrieval tasks. While, experimental data presented in Chapters 3 and 4 supported this finding by demonstrating poorer overall storage and retrieval in the retrieval practise paradigm. However, it is important to note that paradigm itself may cause an interaction through the observed poorer performance in executive abilities that may confound the conclusion of poorer long term memory. Nonetheless, this could be interpreted as behavioural evidence for tinnitus affecting long-term memory. It may be viewed theoretically as the tinnitus percept becoming a permanent 'entry' in a multiple-entry-modular memory system (Johnson, 1983).

6.2.1.3 General cognitive models

While general tinnitus models posit various roles for attention (Section 6.3), prominent general cognitive models may be useful in conceptualising the relationship between tinnitus, cognitive performance, and consequent distress. For example, Duncan (2010) employs a metaphor of executive abilities as mental programs, which are in many ways analogous to computer programs. An interesting element of this conception is the key role that is suggested for fluid ability. Whereas Diamond (2013) posits that fluid ability is essentially the emergence of cognitive flexibility from executive functioning in goal-

directed behaviour, Duncan proposes that fluid ability specifically represents the compiling element of a mental program; in an analogous way in which a computer must compile a program that it intends to execute. Fluid ability is therefore suggested to bring together and sequence cognitive processes to achieve a goal. The findings of Chapter 3 suggest that compilation of mental programs is therefore specifically poorer in those with distressing tinnitus, and while specific functions are independently associated with poorer performance in distressing tinnitus, the biggest effect is seen in fluid ability.

An interpretation that follows is that tinnitus represents a signal that interferes with the compilation of mental programs in those individuals who experience it as distressing, and not for those who do not. Furthermore, Duncan (2013) identifies executive abilities as arising from a “multiple-demand” system. The multiple-demand system shares an association between tinnitus and areas implicated in emotional integration of task information such as the orbital prefrontal cortex (Adjamian et al., 2014; Duncan, 2010), and therefore generalised impairment of the multiple-demand system could be proposed to account for the link to emotional and global psychological distressing components of tinnitus.

Load theory was proposed to unify early and late selection models of attention theory (Lavie, 2010; Murphy et al., 2016). Load theory posits the concepts of perceptual and cognitive load, with emergent attention being a function of both. Perceptual load works at the lower levels (i.e. bottom-up) and suggests that increased perceptual load decreases availability of cognitive load. To expound these concepts, Murphy et al. (2016) provide the example of a person reading while there is a fly in the room: if the text a person is reading has challenging typography (or other low-level features that made it difficult to parse), then the increased perceptual load would likely lead to the fly being ignored. However, if the perceptual qualities were less taxing, this would facilitate increased capacity for the fly to be attended to.

Following this conception, load theory suggests that perceptual aspects of the tinnitus percept (i.e. loudness, pitch, or timbre) would be a key determinant in how

much cognitive load they had to engage with the percept. This interpretation suggests that keeping busy with perceptually demanding tasks would result in not noticing the percept. The face validity of this proposal would appear to have an element of circular logic, because a primary complaint of people with distressing tinnitus is an inability to concentrate and therefore undertake perceptually challenging tasks. Furthermore, this interpretation of tinnitus and load theory also suggests counterintuitive predictions, such as people with a distressing tinnitus percept finding more perceptually challenging tasks easier.

However, experimental data from Chapter 4 of this thesis suggests that people with distressing tinnitus do no worse in a complicated shifting task, and this could be taken as evidence to support this view. Indeed, the results of analyses presented in Chapters 3 and 4 suggest possible differences in working memory (although small) between individuals with non-distressing and no tinnitus, which could be interpreted as increased load due to the tinnitus percept. However, as load theory was primarily developed within the visual domain, Murphy et al. (2017) note that sensory differences in vision and audition make it challenging to directly import load theory into the domain of auditory attention.

In summary, the results presented in this thesis could be viewed as supporting established general models of tinnitus and various prominent cognitive models. However, I suggest that many conceptions of the role of executive abilities/attention are not readily accounted for and required assessment for their role in tinnitus specifically through a model such as the EDM.

6.2.2 Clinical relevance

Appropriate counselling of tinnitus patients is becoming increasingly recognised as an important clinical task (Beukes et al., 2017; Liu et al., 2018; Taylor et al., 2020). The EDM provides a clear conceptual explanatory framework concerning the relationship between executive attention, cognitive performance, and tinnitus distress that can readily be adapted for clinical practise. Current clinical advice is informed by models that

propose tinnitus distress arises through persistent active monitoring and/or negative thoughts about the percept; to someone seeking advice or assistance, such descriptions could appear uncomfortably close to “ignore it and it will go away”. The clinical narrative offered by the EDM does not suggest that people with tinnitus are continually focussing/paying attention to the percept (Section 6.3); instead, it proposes that tinnitus distress can arise from disruption to executive functioning through external stressors, which increase the intrusion of the tinnitus percept. Therefore, it may be a more readily accepted conception for people with distressing tinnitus.

The importance of having a conceptual model relating to attention that makes sense to people with tinnitus has been highlighted by Colagrosso et al. (2018), who reported a qualitative study of factors modulating the tinnitus experience, where attention was suggested to be an important inter-individual factor in tinnitus distress. The attention factor became so important that it went on to become the central feature of their patient-driven model. Furthermore, the causal mapping of the EDM highlights clinically relevant target areas that clinicians can identify with people with distressing tinnitus, as a means of focussing on known treatments that may be useful in their specific instance. For example, reducing stress/improving sleep would result in less disruption to executive abilities and consequently less tinnitus distress (Section 6.4.3.3).

Work presented in this thesis suggests that a reliable measure of executive functioning performance may have prognostic clinical utility for people with recent onset tinnitus that becomes chronic, and may be predictive of their subsequent tinnitus distress. Clarke et al. (2020) noted that if existing screening tools are to be used to assess cognitive performance in people with tinnitus, the Montreal Cognitive Assessment might be favoured, as this taps executive abilities and is more sensitive to cognitive performance than presently available clinical alternatives such as the Mini Mental State Exam (Ciesielska et al., 2016; Devenney and Hodges, 2017; Nasreddine et al., 2005). However, subsequent findings in Chapters 3 and 4 suggest that the Matrix Reasoning task is a candidate for such a measure, while Chapter 5 suggests that a composite

measure, where shared executive functioning variance is extracted across tasks, may be better still.

6.2.3 Strengths and limitations

A lack of psychoacoustic measurements of the tinnitus percept might be considered a potential limitation of this work. These were omitted from the experimental design of Chapter 3 due to its focus on associations between executive functioning and tinnitus distress (best captured by self-report). Furthermore, the additional time required by participants to complete such measures in addition to a battery of cognitive tasks was a practical consideration at the time. Such considerations mean that psychoacoustic measures of tinnitus are often neglected in favour of patient-reported outcomes such as visual analogue scales, which is evidenced by the absence of psychoacoustic measures in the dataset used for cross-validation in Chapter 5.

While I now suggest that psychoacoustic loudness measures are necessary to advance future experimental work relating the causal relationship between tinnitus and cognitive performance, the structural causal modelling presented in Chapter 5 demonstrated that psychoacoustic measurements of the tinnitus percept were not necessary to estimate the effect of executive functioning on tinnitus distress and therefore do not diminish substantive conclusions of this thesis.

6.3 Executive functioning and the role of attention in tinnitus

In this section, I provide an overview of how executive functioning (giving rise to executive attention) has been conceptualised to date in the literature concerning tinnitus and cognition. I argue that a focus on the concept of attention (as a unitary construct) likely results in a conflation between the concepts of bottom-up and top-down attention, and that behavioural evidence to date concerns top-down attention and tinnitus distress. I also provide an overview of proposed mechanistic interactions between executive

attention and the tinnitus percept, as well as presenting challenges to the validity of these conceptions.

Attention is frequently described as being partitioned into bottom-up (stimulus-driven) and top-down (goal driven), and executive functions are the family of top-down mental processes that enable our ability to use top-down or executive attention (Diamond, 2013; Roberts et al., 2013). The behavioural evidence presented in this thesis suggests that executive attention is a crucial component in chronic distressing tinnitus. Despite previous research suggesting that both top-down and bottom-up attention may be important to tinnitus (Mohamad et al., 2016; Tegg-Quinn et al., 2016), Clarke et al. (2020) found that research to date concerning tinnitus and cognitive performance has essentially yielded behavioural evidence related to executive functioning (i.e. top-down attention).

Chun et al. (2011) have suggested that a colloquial understanding of attention has dominated scientific discourse. They argue that scientific study relating to attention has subsequently been impeded because 'attention has become a catch all term for how the brain controls its own information processing'. They also suggest the concept of attention (as a unitary construct or mechanism) should be abandoned, and instead considered as 'a characteristic and property of multiple perceptual and cognitive control mechanisms'.

Kandeeban et al. (2019) have suggested that tinnitus distress is a paradoxical attention to the sound, however, I propose that much discussion pertaining to tinnitus and cognitive performance to date has generally treated attention as a unitary construct, resulting in conceptions of a tinnitus percept/executive attention interaction whereby top-down mechanisms guide information processing that is relatively independent of conscious control (Kaya and Elhilali, 2017), such as auditory attention capture through deviant stimuli or perceived stimulus threat (Andersson et al., 2005; Cuny et al., 2004; Dornhoffer et al., 2006). In an effort to differentiate conceptions of top-down and bottom-up attention, and to understand the role of top-down/executive attention

specifically, I have identified various themes that are recurrent in attempting to explain the contribution of attention (as a unitary construct) to chronic and/or distressing tinnitus.

6.3.1 Attention amplifies tinnitus severity

Some mechanistic models of tinnitus suggest that attention may increase tinnitus severity (Noda et al., 2018; Sedley et al., 2016). However, the term 'tinnitus severity' is ambiguous as to whether the perceptual characteristics, perceived loudness, or related distress are severe. Therefore, a reasonable interpretation of this claim is that tinnitus percept severity itself can be modulated through top-down attention. Few mechanistic examples for how this might arise have been proposed, however, Sedley et al. (2016) postulate that 'focussed attention' increases the gain of reported prediction errors in pyramidal cells.

6.3.2 Attention originates and/or perpetuates the tinnitus percept

Some models suggest that attention plays a part in causing or sustaining tinnitus. For example, Sedley et al. (2016) suggest that attention has a role in sustaining the percept, and that top-down attention has a potential in the 'initial emergence' of tinnitus. They also note the hypothesised role of attention to 'play a major role in the reinforcement of tinnitus once initiated'. This conception of executive attentional roles in tinnitus is also expressed by Trevis et al. (2016), who posit that an individual's failure to switch attention away from the tinnitus percept causes its persistence and subsequent chronification.

6.3.3 Attention distorts the tinnitus percept

Some models suggest a "distorted" perception of tinnitus arises through an attentional mechanism. These conceptions propose that attentional modulations directly affect the perceived magnitude of the tinnitus percept. For example, McKenna et al. (2014) suggest that selective attention and monitoring of the tinnitus percept, combined with negative thoughts, leads to feelings of distress that may cause the percept to be

distorted. Searchfield et al. (2012) describe a model of tinnitus based on Adaptation Level Theory, where the perceived magnitude of tinnitus is influenced by 'attention, background noise, and residual factors'. It is interesting to note that the conception of tinnitus as a distorted perception appears to conflict with psychoacoustic measurement efforts of tinnitus that propose corrections to the percept loudness to make it appropriately "severe" (Henry and Meikle, 2000; Moore, 2012).

6.3.4 The spotlight of attention

Various metaphors have been used to describe attention, but perhaps the most persistent and enduring is the "spotlight" of attention (Bengson et al., 2012; Fernandez-Duque and Johnson, 1999). It is easy to understand the intuitive appeal of the spotlight analogy: an illuminating beam that can be wilfully directed, but with a restricted range and limited focus, where anything falling outside of the spotlight is not attended to.

A common theme amongst the listed conceptualisations of tinnitus and executive attention is that, even when inhibitory/executive functions are considered, they appear to invoke the spotlight of attention metaphor. For example, attention and inhibitory mechanisms are proposed as 'driving attention to the tinnitus perception' (Searchfield et al., 2012), or reducing the 'ability to switch attention away from the tinnitus sound' (Trevis et al., 2016). McKenna et al. (2014) suggest that negative thoughts draw more attention toward the tinnitus and increased attention fuels further negative thoughts.

The spotlight analogy resurfaces again and again; however, intuitive visual cognitive concepts may be problematic when conceiving of (phantom) phenomena in the auditory domain (Murphy et al., 2017). Nonetheless, the analogy remains common within the literature, with Zeng et al. (2020) postulating the existence of a 'tinnitus-enhanced "spotlight"' of attention to explain normal speech and auditory perception in a sample of participants with tinnitus.

6.3.5 Causal relationship between constructs

A further commonality in the outlined attentional themes is top-down attention *causing* something to happen to the tinnitus percept. Attention is proposed to cause:

- i. increased tinnitus percept severity
- ii. tinnitus percept chronification
- iii. distortion of the “true” tinnitus percept

However, tinnitus is simultaneously presumed to cause deficits in attention, which is a conception that I have suggested may arise from early peripheral models of tinnitus pathophysiology (Section 1.3.4). The conceptual trappings of this default language appears to be the rule rather than the exception within the literature, with tinnitus (presumably the percept) generally described as having an ‘effect’ on cognition (Cardon et al., 2019; Mohamad et al., 2016; Rossiter et al., 2006; Tegg-Quinn et al., 2016). The current conceptualisation of the relationship between attention and tinnitus may exhibit reifying logic, as simultaneously the tinnitus percept is tacitly assumed to be the cause of attentional difficulties and attentional focus is assumed to be the cause of a more severe and/or persistent tinnitus.

Tavanai and Mohammadkhani (2018) suggest a ‘birectional relationship attention and memory with tinnitus’ (Figure 6.1). Their conceptualisation hints at a confounding third variable (i.e. shared brain generators being the proximal cause of tinnitus and cognitive problems), but ultimately they also make familiar allusions of a tinnitus percept being the causal agent, and conclude that ‘many factors seem to play a role in evaluating the effects of tinnitus on cognition’.

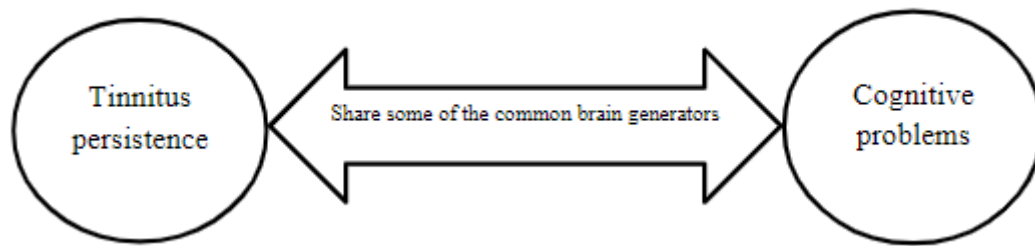


Figure 6—1. Illustration of reciprocal link between tinnitus and cognitive problems from Tavanai and Mohammadkhani (2018).

I suggest that these conceptions of the role of executive attention are mechanistically problematic because they fundamentally appear to propose that people with distressing tinnitus are wilfully and constantly directing their attention and focus to their tinnitus percept. However, experimental data from my thesis has shown that people with distressing tinnitus display poorer executive functioning that is necessary to facilitate an unwavering focus on the tinnitus percept. Furthermore, most people have non-distressing tinnitus and do not appear to focus on their tinnitus percept unless asked to, making the face validity of the claim that attention causes persistence of the percept problematic in the majority of cases.

Understanding the role of executive attention in tinnitus distress is important because non-cognitive researchers and clinicians (who are more likely to hold an unquestioned unitary view of attention) will inherit conceptual elements of the themes outlined above. Consequently, clinicians may advise patients that their tinnitus distress (severity?) would improve if they just stopped paying attention to it. For many people it will not. The conceptual view offered by the EDM provides an alternative clinical perspective to aid counselling by not suggesting that attention is being volitionally directed to the percept (Section 6.2.2).

McFerran et al. (2019) note that tinnitus has 'long been conceived as a two-stage process', with 'an initial ignition' in the auditory system, followed by a 'process of promotion'. Bottom-up attention may well have a role in generation and chronification of the tinnitus percept (Roberts et al., 2013). However, it is unclear to the extent that

behavioural evidence can be brought to bear on the role of bottom-up attention, as some authors have argued that even simple behavioural tasks necessarily involve executive processes (Cepeda et al., 2013). It is likely that the relationship between bottom-up attention and tinnitus will be more fruitfully investigated through imaging measures (e.g. electroencephalography or magnetoencephalography).

I propose that in addition to the two-stage process described by McFerran et al. (2019), subsequent distress should be considered the third. Executive attention, as I have demonstrated in this thesis, appears primarily linked to tinnitus distress. The EDM provides a novel conceptualisation in positing that the tinnitus percept is not the subject of focus for an attentional spotlight, but instead intrudes into conscious awareness because of inherent and/or disrupted executive abilities and causally contributes to distress.

An interesting possibility for further investigation of the role top-down attention within the auditory system may be through the objective measure of otoacoustic emissions. Differences have been reported in the magnitudes of otoacoustic emissions during tasks of selective auditory and visual attention (Walsh et al., 2015). Therefore, participants with tinnitus who are matched on hearing thresholds, but differ in their levels of distress, could undergo such a procedure to understand whether they differ in their ability to exert top-down attentional control within the auditory system.

6.4 Future work

6.4.1 EDM: implications and testable predictions

The conceptual view of the relationship between tinnitus distress and executive abilities offered by the EDM (Section 5.3) suggests various implications and predictions that can further scrutinised. For example, the model suggests that people with distressing tinnitus will experience more cognitive performance failures during their daily lives. The results of Chapters 3 and 4 support this with self-reported scores on the CFQ. A possible way to

further test this hypothesis would be to have participants track their executive performance failures and perceived tinnitus intrusiveness using methods such as ecological momentary assessment (Goldberg et al., 2017).

The EDM also makes predictions from observational data. For example, people with loudness tinnitus and better executive performance should not find their tinnitus distressing, while those with quieter tinnitus and poorer executive abilities should find their tinnitus equivalently or even more distressing. A further potential test of the EDM would be through a longitudinal study design, where a battery of tasks, measuring latent executive ability is given to people who have recently developed tinnitus and attend a tinnitus clinic. The EDM suggests that the latent measure of performance on such tasks will go on to significantly predict tinnitus distress should it persist.

The EDM also suggests that through stopping disruption to the executive abilities (e.g. through reducing global psychological distress or improving sleep) it should be possible to reduce tinnitus distress (see Section 6.4.3). The causal relationships considered alongside the EDM are important to understand, as they would inform the likelihood of success of various therapeutic intervention. For example, whether inherent poorer executive abilities are the cause may imply cognitive training is required, whereas disruption through stress would imply that removal of the disrupting stressor is required.

The SCMs presented in Chapter 5 (Section 5.5) demonstrated that qualities of the tinnitus percept were not needed to estimate the direct effect of executive abilities on tinnitus distress. However, future research could build on the work presented in this thesis by investigating qualities of the tinnitus percept such as loudness, pitch, and, timbre, and their relationship to executive functioning. This work could investigate direct and indirect effects of these variables on tinnitus distress by using the structural causal modelling tools outlined in this thesis, and experimental techniques such as simulated tinnitus (Hesser et al., 2013), and could to link to other interesting potential cognitive explanatory theories such as load theory (Lavie, 2010; Murphy et al., 2016).

The experimental data presented in this thesis suggest that subtle deficits in updating of working memory may also be independently associated with non-distressing tinnitus compared to those no tinnitus. This is an interesting area to explore further to see if non-distressing tinnitus actually impacts lower-level cognitive systems. Similarly, the involvement of long-term memory was hinted at through the RIF task. However, more extensive studies of working memory and long-term memory would be required to establish if these systems per se are altered in participants with tinnitus, or if the observed associations are present through executive abilities.

6.4.2 Wider applications of tinnitus and SCMs

Causal questions are vital for complex associative research, such as within the hearing sciences. For example, recent work has posited a causal link between hearing impairment and cognitive decline, and between hearing impairment and disrupted sleep (Clarke et al., 2019; Wayne and Johnsrude, 2015). While this thesis builds on a body of work that has suggested tinnitus causes cognitive deficits (Mohamad et al., 2016), understanding which variables affect others is of crucial importance for effective intervention.

This thesis presented and used a SCM framework that aims to make mathematically valid causal claims about observational data (Pearl, 2009; Pearl and Mackenzie, 2018). The previous chapter found that providing a conceptual model, and a rigorous SCM provided consistent casual estimates of executive abilities in distressing tinnitus across independent datasets.

Among the attractive features of mapping conceptual model to SCMs for research into tinnitus and cognition is the ability of local elements to be added to existing models (Section 5.5.2). Therefore, researchers working within the field of tinnitus and cognition can use the EDM and related SCM and augment this with their own hypothesised causal relations. After submitting these models to causal path analyses, they will be able to generate their own testable predictions of effects that are directly comparable to the work presented in this thesis. SCMs provide a principled, reproducible way to articulate

assumptions underpinning the relations of each of theoretically important and potentially confounding variables; this methodology can facilitate more effective communication among researchers on this complex topic.

Once theoretically underpinned causal models are established, further techniques for the analysis of SCMs - such as interventional and counterfactual effect estimates - can be leveraged, to investigate substantive theoretical claims that were previously only possible through expensive randomised control trials (Pearl et al., 2016). This is useful for making inferences regarding observational data because although tinnitus may be temporarily induced through various experimental methods, it is typically neither practical nor ethical to randomly induce in human participants, and will necessarily be confounded by variables that could affect cognitive performance such as hearing loss and age. A typical solution for experimental-inducement of tinnitus is to use animal models, however, there are challenges when extrapolating findings from animal models to humans (Eggermont, 2016). This is particularly true of models concerning distressing tinnitus because of the known involvement of higher-order processing and non-auditory pathways.

Furthermore, SCMs have wider applications to tinnitus research because even when appropriate experiments are theoretically possible, tinnitus research has noted funding limitations relative to other highly prevalent conditions (Cederroth et al., 2013); it is therefore imperative that researchers are able to make causal statements about data generated from observational studies if interventions are to be developed and tested. Application of methods from the SCM framework can offer alternatives to typical experimental methods and enable causal claims to be made from observational data. Furthermore, SCMs provide theoretically-informed alternatives to causally-blind machine learning approaches that are currently gaining interest within the field of tinnitus research (Simoes et al., 2019; van den Berge et al., 2017).

6.4.3 Executive function interventions

Diamond and Ling (2016) argue that the question of whether executive functioning can be improved is uninteresting, suggesting that it demonstrably can be 'at every age from infants through elders and via diverse approaches'. An underexplored question within the tinnitus literature is: what are the benefits of targeted executive function interventions for people with tinnitus?

The experimental and theoretical results of this thesis suggest that executive function interventions are unlikely to be effective for individuals who do not find their tinnitus distressing. However, for people with distressing tinnitus they may be beneficial. As a treatment strategy, benefit could arise through improving deficits caused by tinnitus distress. However, the EDM suggests that effective intervention on executive functions (or the causes that disrupt them) may reduce the intrusiveness of the tinnitus percept and therefore improve tinnitus distress. Various interventions aimed at achieving this effect could be explored.

6.4.3.1 *Pharmacological*

A pharmacological intervention arguably remain a desirable option because they aligns with observations that people suffering from tinnitus would be likely to take pills aimed at reducing 'loudness or annoyance' (McFerran et al., 2019). Pharmacological interventions for cognition related to tinnitus remains an unexplored domain. While many pharmacological options have been investigated in tinnitus research generally few have directly targeted cognitive performance (Baguley et al., 2012; Beebe Palumbo et al., 2015; Langguth et al., 2009; Tetteh et al., 2017).

A large research literature exists on the use of pharmacological compounds to improvement cognitive performance (Floresco and Jentsch, 2011; Greely et al., 2008; Mehlman, 2004; Suliman et al., 2016). However, to my awareness, only one published study and one unpublished clinical trial have investigated the overlap between tinnitus, cognitive performance, and pharmacological intervention.

Krings et al. (2015) reported no adverse effects when aiming to test the feasibility and preliminary effectiveness of using a putative neuroplasticity-enhancing drug, D-cycloserine; the authors hypothesised that the neuroplasticity-sensitizing properties of this compound 'would lead to greater improvements in the severity of tinnitus bother and cognitive functioning when given with the Brain Fitness Program'. They employed a double-blind randomised control trial where both groups undertook computer training, but one received D-cycloserine (n = 16) and the other placebo (n = 14). The authors reported that the group who received D-cycloserine demonstrated a greater overall improvement in tinnitus distress and cognitive deficits compared to baseline but the difference was not statistically significant. Despite the promising appearance of the reported results, the conclusions that can be drawn from this study are limited due to its small sample size, inclusion of two treatments, and absence of reported individual change data.

An unpublished clinical trial for participants with tinnitus was run in 2008 (<https://clinicaltrials.gov/ct2/show/NCT00591019>), which investigated the use of modafinil for 'attention problems associated with tinnitus'. Modafinil has been approved for the treatment of narcolepsy, and is thought to carry a low risk of harm. Its primary effects are enhancing concentration and alertness (Kredlow et al., 2019). The trial was terminated after enrolling only seven participants due to a lack of change in the primary and secondary outcome measures (<https://clinicaltrials.gov/ct2/show/results/NCT00591019>). However, the insights of this thesis suggest it would be worth revisiting a clinical trial into the efficacy of this intervention but in participants with distressing tinnitus only. With hindsight, there were various methodological issues with the study that meant it may not have been optimal for addressing an intervention concerning attention and tinnitus distress: For example, the primary (P50 amplitude) and secondary (simple reaction time) outcomes were not directly related to executive functioning or tinnitus distress, and no measure of tinnitus

distress (e.g. THI) was used to identify participants who may benefit from enhancement of executive functioning.

Given the broad scope of the literature concerning pharmacological cognitive enhancement, future research on this topic in relation to tinnitus should be driven by theory, with executive functions being highlighted in this thesis as a promising cognitive therapeutic domain. Pharmacology of executive functioning is an expansive and active area of research (Jack et al., 2018), and this it is an interesting prospect to seek pharmacological improvement of executive functions in people with distressing tinnitus; however, the appropriateness and feasibility of such interventions in a population with tinnitus remains to be explored.

6.4.3.2 Non-pharmacological

Cognitive and perceptual training methods are a potential, non-pharmacological intervention that have been subject to increasing scientific interest (Edwards et al., 2018). Such methods may have therapeutic potential in a population with tinnitus, and have already been suggested or applied in various diverse populations of patients with chronic conditions associated with impaired cognitive functioning including: eating disorders, attention-deficit/hyperactivity disorder, Parkinson's disease, multiple sclerosis, and the closely-related field of hearing impairment (Dardiotis et al., 2018; Díez-Cirarda et al., 2018; Goode et al., 2018; Henshaw and Ferguson, 2013; Juarascio et al., 2015). Perceptual training aims to improve bottom-up attention and has been of interest in tinnitus for a number of years (Hoare et al., 2014, 2010; Searchfield et al., 2007). However, the divide between cognitive and perceptual training is not always clear cut. For example, Wise et al. (2016) implemented training that despite being reported as perceptual training, likely tapped executive functions; the training game 'Terrain' was compared to popular computer game 'Tetris', and required participants to 'move through a virtual auditory landscape toward target sounds while ignoring distracter sounds, some of which were matched to the individual's tinnitus'. The authors reported improved scores on TFI and THI in both treatment and control groups. While these were greater in

the treatment group, the results appear to support the view that general executive function training through Tetris was also effective. Similarly, Kallogjeri et al. (2017) undertook an RCT of a proprietary cognitive training program, where the authors concluded that computer-based cognitive training is 'associated with self-reported changes in attention, memory, and perception of tinnitus'. They reported that although there was a post-intervention reduction in both groups, this was not significantly different between groups.

Future cognitive training studies may even be informed by the findings of this thesis and attempt to train specific executive functions, such as inhibitory control, or fluid ability (Webb et al., 2018). However, cognitive and executive function training is an active area of developing and contentious research claims concerning generalised transfer of training benefits (Chiu et al., 2017; Katz et al., 2018; Melby-Lervåg et al., 2016; Webb et al., 2018). For example, Diamond and Ling (2016) question the generalisability of such computer-based training methods. This is important because it remains an open question as to whether such training would generalise to improvements in executive ability for people with distressing tinnitus.

Domain-general improvements aside, the therapeutic benefit of training specific tasks that resemble appropriate skills (i.e. focusing on things other than a distracting sound) warrants further research due to the reported additional benefit of a training programme tailored to tinnitus (Wise et al., 2016). Furthermore, pharmacological and non-pharmacological approaches could be used together. For example, Krings et al. (2015) used a proprietary Brain Fitness program (Posit Science) in combination with a pharmacological intervention and suggested that this approach was well tolerated and caused greater improvements compared to cognitive training alone. However, the conclusions that can be made from currently available research are limited by the variable designs, methods, and group sample sizes used in studies to date.

6.4.3.3 *Indirect*

Indirect means of improving executive functions may also be fruitful. The EDM suggests that when executive functions are disrupted, then distress related to the tinnitus percept will increase because the tinnitus percept becomes more intrusive on conscious experience. Diamond and Ling (2016) suggest that executive functions 'are better when you are more relaxed and un-stressed, feel emotionally and socially nourished, and are in good health'. In light of this, they suggest that various interventions may indirectly improve executive functions if they challenge them and incorporate emotional, physical, and social health components.

The sample of participants in the EFiT study rated themselves as recently experiencing substantially more stress than both other experimental groups. Conversely, the group with non-distressing tinnitus were less stress than either group. This is interesting in the context of the EDM, as Diamond and Ling (2016) note that executive functions 'suffer first and most if you are stressed, sad, lonely, or not in good physical health'. Therefore, the specific executive functioning interventions described above could be combined with interventions that target global psychological distress, sleep, and general lifestyle factors.

6.5 Summary

This chapter presented a discussion of knowledge contribution concerning tinnitus and cognitive performance generated in this thesis. The primary strength of this thesis is that it used a variety of quantitative methods to reach convergent, theoretically-driven answer as to the role and nature of executive abilities in distressing tinnitus. Various quantitative methods were applied in combination with current cognitive theory to consistently link executive abilities to tinnitus distress. Key highlights from the thesis include: comprehensive synthesis of research literature to associate tinnitus with poorer executive performance, and identification of confounding factors within the literature;

presentation of novel experimental data that demonstrated poorer executive performance being associated with tinnitus distress, but also identification of methodological limitations that limit conclusions of an independent association; application of advanced methods to independently associate distress tinnitus with poorer executive performance, and non-distressing with improved performance; and the proposition and validation of a theoretically-driven model that links executive functioning to tinnitus distress in independent datasets. This chapter also discussed the proposed role of executive abilities/attention in tinnitus distress. Future work was suggested that included the testable implications of the proposed theoretical model, use of SCMs in future research, and possibilities for exploring interventions related to executive functioning for tinnitus distress.

Covid impact statement: Additional planned experimental work (that aimed to investigate the impact of an artificial tinnitus sound on the cognitive performance of individuals without tinnitus) was impacted by the Covid pandemic. University closures and testing restrictions led to considerable disruption of planned activities and the subsequent abandonment of this line of research. This negated a large amount of related work that I had already completed, including a considerable amount of study planning, material preparation, and successful applications for ethical approval.

References

- Abelson, R.P., 1985. A variance explanation paradox: When a little is a lot. *Psychol Bull* 97, 129–133. <https://doi.org/10.1037/0033-2909.97.1.129>
- Acrani, I.O., Pereira, L.D., 2010. Temporal resolution and selective attention of individuals with tinnitus. *Fono* 22, 233–238.
- Adjamian, P., Hall, D.A., Palmer, A.R., Allan, T.W., Langers, D.R.M., 2014. Neuroanatomical abnormalities in chronic tinnitus in the human brain. *Neurosci Biobehav Rev* 45, 119–133. <https://doi.org/10.1016/j.neubiorev.2014.05.013>
- Alam, N., Katarkar, A., Shah, P., Jalvi, R., Jain, A., Shah, M., 2012. Audiological, psychological and cognitive characteristics of tinnitus sufferers. *Indian J. Otol.* 18, 20. <https://doi.org/10.4103/0971-7749.98288>
- Althouse, A.D., 2016. Adjust for Multiple Comparisons? It's Not That Simple. *Ann Thorac Surg* 101, 1644–1645. <https://doi.org/10.1016/j.athoracsur.2015.11.024>
- Alvarez, J.A., Emory, E., 2006. Executive function and the frontal lobes: a meta-analytic review. *Neuropsychol Rev* 16, 17–42. <https://doi.org/10.1007/s11065-006-9002-x>
- Amaral, A.A., Langers, D.R.M., 2015. Tinnitus-related abnormalities in visual and salience networks during a one-back task with distractors. *Hear Res* 326, 15–29. <https://doi.org/10.1016/j.heares.2015.03.006>
- Anderson, M.C., 2003. Rethinking interference theory: Executive control and the mechanisms of forgetting. *J Mem Lang* 49, 415–445.
- Anderson, M.C., Bjork, R.A., Bjork, E.L., 1994. Remembering can cause forgetting: retrieval dynamics in long-term memory. *J Exp Psychol Learn Mem Cogn* 20, 1063–1087.
- Andersson, G., Bakhsh, R., Johansson, L., Kaldo, V., Carlbring, P., 2005. Stroop facilitation in tinnitus patients: an experiment conducted via the world wide web. *Cyberpsychol Behav* 8, 32–38. <https://doi.org/10.1089/cpb.2005.8.32>
- Andersson, G., Eriksson, J., Lundh, L.G., Lyttkens, L., 2000. Tinnitus and cognitive interference: a stroop paradigm study. *J Speech Lang Hear Res* 43, 1168–1173. <https://doi.org/10.1044/jslhr.4305.1168>
- Andersson, G., McKenna, L., 2006. The role of cognition in tinnitus. *Acta Otolaryngol Suppl* 39–43. <https://doi.org/10.1080/03655230600895226>
- Antony, M.M., Bieling, P.J., Cox, B.J., Enns, M.W., Swinson, R.P., 1998. Psychometric Properties of the 42-Item and 21-Item Versions of the Depression Anxiety Stress Scales in Clinical Groups and a Community Sample. *Psychol Assess* 10, 176–181. <https://doi.org/10.1037/1040-3590.10.2.176>
- Araneda, R., De Volder, A.G., Deggouj, N., Philippot, P., Heeren, A., Lacroix, E., Decat, M., Rombaux, P., Renier, L., 2015. Altered top-down cognitive control and auditory processing in tinnitus: evidences from auditory and visual spatial stroop. *Restor Neurol Neurosci* 33, 67–80. <https://doi.org/10.3233/RNN-140433>
- Attal, N., Masselin-Dubois, A., Martinez, V., Jayr, C., Albi, A., Fermanian, J., Bouhassira, D., Baudic, S., 2014. Does cognitive functioning predict chronic pain? Results from a prospective surgical cohort. *Brain* 137, 904–917. <https://doi.org/10.1093/brain/awt354>
- Baayen, R.H., Milin, P., 2010. Analyzing reaction times. *Int. J. Psychol. Res.* 3, 12–28.
- Baddeley, A., 2012. Working memory: theories, models, and controversies. *Annu Rev Psychol* 63, 1–29. <https://doi.org/10.1146/annurev-psych-120710-100422>
- Baguley, D., Andersson, G., McFerran, D., McKenna, L., 2012. Tinnitus: A Multidisciplinary Approach, 2 edition. ed. Wiley-Blackwell.
- Baguley, D., McFerran, D., Hall, D., 2013. Tinnitus. *Lancet* 382, 1600–1607. [https://doi.org/10.1016/S0140-6736\(13\)60142-7](https://doi.org/10.1016/S0140-6736(13)60142-7)
- Baguley, D.M., 2002. Mechanisms of tinnitus. *Br Med Bull* 63, 195–212. <https://doi.org/10.1093/bmb/63.1.195>

- Baldo, P., Doree, C., Molin, P., McFerran, D., Cecco, S., 2012. Antidepressants for patients with tinnitus. *Cochrane Database Syst Rev* CD003853. <https://doi.org/10.1002/14651858.CD003853.pub3>
- Banbury, S.P., Macken, W.J., Tremblay, S., Jones, D.M., 2001. Auditory distraction and short-term memory: phenomena and practical implications. *Hum Factors* 43, 12–29. <https://doi.org/10.1518/001872001775992462>
- Banich, M.T., Compton, R.J., 2018. *Cognitive Neuroscience*. Cambridge University Press.
- Barulli, D., Stern, Y., 2013. Efficiency, capacity, compensation, maintenance, plasticity: emerging concepts in cognitive reserve. *Trends Cogn Sci* 17, 502–509. <https://doi.org/10.1016/j.tics.2013.08.012>
- Baujat, B., Mahé, C., Pignon, J.-P., Hill, C., 2002. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. *Stat Med* 21, 2641–2652. <https://doi.org/10.1002/sim.1221>
- Beebe Palumbo, D., Joos, K., De Ridder, D., Vanneste, S., 2015. The Management and Outcomes of Pharmacological Treatments for Tinnitus. *Curr Neuropharmacol* 13, 692–700. <https://doi.org/10.2174/1570159x13666150415002743>
- Bengson, J.J., Lopez-Calderon, J., Mangun, G.R., 2012. The spotlight of attention illuminates failed feature-based expectancies. *Psychophysiology* 49, 1101–1108.
- Benjamin, D.J., Berger, J.O., Johannesson, M., Nosek, B.A., Wagenmakers, E.-J., Berk, R., Bollen, K.A., Brembs, B., Brown, L., Camerer, C., Cesarini, D., Chambers, C.D., Clyde, M., Cook, T.D., De Boeck, P., Dienes, Z., Dreber, A., Easwaran, K., Efferson, C., Fehr, E., Fidler, F., Field, A.P., Forster, M., George, E.I., Gonzalez, R., Goodman, S., Green, E., Green, D.P., Greenwald, A.G., Hadfield, J.D., Hedges, L.V., Held, L., Hua Ho, T., Hoijtink, H., Hruschka, D.J., Imai, K., Imbens, G., Ioannidis, J.P.A., Jeon, M., Jones, J.H., Kirchler, M., Laibson, D., List, J., Little, R., Lupia, A., Machery, E., Maxwell, S.E., McCarthy, M., Moore, D.A., Morgan, S.L., Munafó, M., Nakagawa, S., Nyhan, B., Parker, T.H., Pericchi, L., Perugini, M., Roulder, J., Rousseau, J., Savalei, V., Schönbrodt, F.D., Sellke, T., Sinclair, B., Tingley, D., Van Zandt, T., Vazire, S., Watts, D.J., Winship, C., Wolpert, R.L., Xie, Y., Young, C., Zinman, J., Johnson, V.E., 2018. Redefine statistical significance. *Nat Hum Behav* 2, 6–10. <https://doi.org/10.1038/s41562-017-0189-z>
- Ben-Shachar M, Makowski D, Lüdtke D, 2020. Compute and interpret indices of effect size.
- Best, J.R., Miller, P.H., 2010. A developmental perspective on executive function. *Child Dev* 81, 1641–1660. <https://doi.org/10.1111/j.1467-8624.2010.01499.x>
- Betensky, R.A., 2019. The p-Value Requires Context, Not a Threshold. *Am Stat* 73, 115–117. <https://doi.org/10.1080/00031305.2018.1529624>
- Beukes, E.W., Andersson, G., Allen, P.M., Manchaiah, V., Baguley, D.M., 2018. Effectiveness of Guided Internet-Based Cognitive Behavioral Therapy vs Face-to-Face Clinical Care for Treatment of Tinnitus: A Randomized Clinical Trial. *JAMA Otolaryngol Head Neck Surg* 144, 1126–1133. <https://doi.org/10.1001/jamaoto.2018.2238>
- Beukes, E.W., Baguley, D.M., Allen, P.M., Manchaiah, V., Andersson, G., 2017. Audiologist-Guided Internet-Based Cognitive Behavior Therapy for Adults With Tinnitus in the United Kingdom: A Randomized Controlled Trial. *Ear Hear*. <https://doi.org/10.1097/AUD.0000000000000505>
- Bhatt, J.M., Bhattacharyya, N., Lin, H.W., 2017. Relationships between tinnitus and the prevalence of anxiety and depression. *Laryngoscope* 127, 466–469. <https://doi.org/10.1002/lary.26107>
- Biswas, R., Hall, D.A., 2020. Epidemiology of Tinnitus. https://doi.org/10.1007/7854_2020_154
- Biswas, R., Lugo, A., Gallus, S., Akeroyd, M.A., Hall, D.A., 2019. Standardized questions in English for estimating tinnitus prevalence and severity, hearing difficulty and usage of healthcare resources, and their translation into 11 European languages. *Hear Res* 377, 330–338. <https://doi.org/10.1016/j.heares.2019.02.008>
- BNSSG Research and Evidence Team, W. of E.A.H.S.N. (WEAHSN), The National Institute for Health Research Collaboration for Leadership in Applied Health Research and

- Care West (NIHR CLAHRC West), 2018. NHS Hierarchy of Evidence [WWW Document]. URL <http://www.nhsevidencetoolkit.net/what-is-evidence/> (accessed 7.31.18).
- Brandimonte, M.A., Bruno, N., Collina, S., 2006. Cognition, in: P Pawlik And (Ed.), *Psychological Concepts: An International Historical Perspective*. Hove, UK: Psychology Press.
- British Society of Audiology, 2018. Pure tone air and bone conduction threshold audiometry with and without masking.
- Broadbent, D.E., Cooper, P.F., FitzGerald, P., Parkes, K.R., 1982. The Cognitive Failures Questionnaire (CFQ) and its correlates. *Br J Clin Psychol* 21 (Pt 1), 1–16. <https://doi.org/10.1080/00221300209602098>
- Brown, S.D., Heathcote, A., 2008. The simplest complete model of choice response time: linear ballistic accumulation. *Cogn Psychol* 57, 153–178.
- Bürkner, P.-C., 2017. brms: An R Package for Bayesian Multilevel Models Using Stan. *J. Stat. Softw. Artic.* 80, 1–28.
- Bushman, B.J., Wang, M.C., 2009. Vote-counting procedures in meta-analysis.
- Cabeza, R., Albert, M., Belleville, S., Craik, F.I.M., Duarte, A., Grady, C.L., Lindenberger, U., Nyberg, L., Park, D.C., Reuter-Lorenz, P.A., Rugg, M.D., Steffener, J., Rajah, M.N., 2018. Maintenance, reserve and compensation: the cognitive neuroscience of healthy ageing. *Nat Rev Neurosci.* <https://doi.org/10.1038/s41583-018-0068-2>
- Camerer, C.F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B.A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., Isaksson, S., Manfredi, D., Rose, J., Wagenmakers, E.-J., Wu, H., 2018. Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nat Hum Behav* 2, 637–644. <https://doi.org/10.1038/s41562-018-0399-z>
- Cardon, E., Jacquemin, L., Mertens, G., Van de Heyning, P., Vanderveken, O.M., Topsakal, V., De Hertogh, W., Michiels, S., Van Rompaey, V., Gilles, A., 2019. Cognitive Performance in Chronic Tinnitus Patients: A Cross-Sectional Study Using the RBANS-H. *Otol Neurotol* 40, e876–e882.
- Carroll, J., 2010. *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge University Press.
- Cederroth, C.R., Canlon, B., Langguth, B., 2013. Hearing loss and tinnitus--are funders and industry listening? *Nat Biotechnol* 31, 972–974. <https://doi.org/10.1038/nbt.2736>
- Cepeda, N.J., Blackwell, K.A., Munakata, Y., 2013. Speed isn't everything: complex processing speed measures mask individual differences and developmental changes in executive control. *Dev Sci* 16, 269–286. <https://doi.org/10.1111/desc.12024>
- Chalmers, I., Hedges, L.V., Cooper, H., 2002. A brief history of research synthesis. *Eval Health Prof* 25, 12–37. <https://doi.org/10.1177/0163278702025001003>
- Chambers, J.M., 2018. *Graphical Methods for Data Analysis*. CRC Press.
- Champely, S., 2020. Basic Functions for Power Analysis [R package pwr version 1.3-0]. Comprehensive R Archive Network (CRAN).
- Chiu, H.-L., Chu, H., Tsai, J.-C., Liu, D., Chen, Y.-R., Yang, H.-L., Chou, K.-R., 2017. The effect of cognitive-based training for the healthy older people: A meta-analysis of randomized controlled trials. *PLoS One* 12, e0176742. <https://doi.org/10.1371/journal.pone.0176742>
- Chun, M.M., Golomb, J.D., Turk-Browne, N.B., 2011. A taxonomy of external and internal attention. *Annu Rev Psychol* 62, 73–101. <https://doi.org/10.1146/annurev.psych.093008.100427>
- Chung, J.H., Lee, S.H., 2016. The pathophysiologic mechanism of tinnitus. *Hanyang Med. Rev.* 36, 81–85. <https://doi.org/10.7599/hmr.2013..81>
- Ciesielska, N., Sokołowski, R., Mazur, E., Podhorecka, M., Polak-Szabela, A., Kędziora-Kornatowska, K., 2016. Is the Montreal Cognitive Assessment (MoCA) test better suited than the Mini-Mental State Examination (MMSE) in mild cognitive

- impairment (MCI) detection among people aged over 60? Meta-analysis. *Psychiatr Pol* 50, 1039–1052. <https://doi.org/10.12740/PP/45368>
- Cima, R.F.F., 2018. Bothering tinnitus : Cognitive behavioral perspectives. *HNO* 66, 369–374. <https://doi.org/10.1007/s00106-018-0502-9>
- Cima, R.F.F., Mazurek, B., Haider, H., Kikidis, D., Lapira, A., Noreña, A., Hoare, D.J., 2019. A multidisciplinary European guideline for tinnitus: diagnostics, assessment, and treatment. *HNO*. <https://doi.org/10.1007/s00106-019-0633-7>
- Clarke, N.A., Akeroyd, M.A., Henshaw, H., Hoare, D.J., 2018. Association between subjective tinnitus and cognitive performance: protocol for systematic review and meta-analysis. *BMJ Open* 8, e023700. <https://doi.org/10.1136/bmjopen-2018-023700>
- Clarke, N.A., Henshaw, H., Akeroyd, M.A., Adams, B., Hoare, D.J., 2020. Associations Between Subjective Tinnitus and Cognitive Performance: Systematic Review and Meta-Analyses. *Trends Hear.* 24, 2331216520918416. <https://doi.org/10.1177/2331216520918416>
- Clarke, N.A., Hoare, D.J., Killan, E.C., 2019. Evidence for an Association Between Hearing Impairment and Disrupted Sleep: Scoping Review. *Am. J. Audiol.* 28, 1015–1024. https://doi.org/10.1044/2019_AJA-19-0026
- Cohen, J., 1988. *Statistical Power Analysis*, 2 edition. ed. Routledge.
- Cohen, J., Cohen, P., West, S.G., Aiken, L.S., 2003. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3rd ed. Routledge.
- Colagrosso, E.M.G., Fournier, P., Fitzpatrick, E.M., Hébert, S., 2018. A Qualitative Study on Factors Modulating Tinnitus Experience. *Ear Hear.* <https://doi.org/10.1097/AUD.0000000000000642>
- Correll, J., Mellinger, C., McClelland, G.H., Judd, C.M., 2020. Avoid Cohen’s “Small”, “Medium”, and “Large” for Power Analysis. *Trends Cogn Sci.* <https://doi.org/10.1016/j.tics.2019.12.009>
- Craik, F.I.M., 2007. The role of cognition in age-related hearing loss. *J Am Acad Audiol* 18, 539–547.
- Cumming, G., 2017. *Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. Routledge.
- Cuny, C., Noreña, A., El Massioui, F., Chéry-Croze, S., 2004. Reduced attention shift in response to auditory changes in subjects with tinnitus. *Audiol Neurotol* 9, 294–302. <https://doi.org/10.1159/000080267>
- Dardiotis, E., Nousia, A., Siokas, V., Tsouris, Z., Andravizou, A., Mentis, A.-F.A., Florou, D., Messinis, L., Nasios, G., 2018. Efficacy of computer-based cognitive training in neuropsychological performance of patients with multiple sclerosis: A systematic review and meta-analysis. *Mult Scler Relat Disord* 20, 58–66. <https://doi.org/10.1016/j.msard.2017.12.017>
- Das, S.K., Wineland, A., Kallogjeri, D., Piccirillo, J.F., 2012. Cognitive speed as an objective measure of tinnitus. *Laryngoscope* 122, 2533–2538. <https://doi.org/10.1002/lary.23555>
- Davis, A., El Refaie, A., 2000. The epidemiology of tinnitus, in: *The Handbook of Tinnitus*. Singular.
- Dawes, P., Emsley, R., Cruickshanks, K.J., Moore, D.R., Fortnum, H., Edmondson-Jones, M., McCormack, A., Munro, K.J., 2015. Hearing Loss and Cognition: The Role of Hearing Aids, Social Isolation and Depression. *PLoS One* 10, e0119616. <https://doi.org/10.1371/journal.pone.0119616>
- De Ridder, D., Vanneste, S., Freeman, W., 2014a. The Bayesian brain: phantom percepts resolve sensory uncertainty. *Neurosci Biobehav Rev* 44, 4–15. <https://doi.org/10.1016/j.neubiorev.2012.04.001>
- De Ridder, D., Vanneste, S., Langguth, B., Llinas, R., 2015. Thalamocortical Dysrhythmia: A Theoretical Update in Tinnitus. *Front Neurol* 6, 124. <https://doi.org/10.3389/fneur.2015.00124>
- De Ridder, D., Vanneste, S., Weisz, N., Londero, A., Schlee, W., Elgoyhen, A.B., Langguth, B., 2014b. An integrative model of auditory phantom perception:

- tinnitus as a unified percept of interacting separable subnetworks. *Neurosci Biobehav Rev* 44, 16–32. <https://doi.org/10.1016/j.neubiorev.2013.03.021>
- Degeest, S., Keppler, H., Corthals, P., 2017. The Effect of Tinnitus on Listening Effort in Normal-Hearing Young Adults: A Preliminary Study. *J Speech Lang Hear Res* 60, 1036–1045. https://doi.org/10.1044/2016_JSLHR-H-16-0090
- Del Bo, L., Forti, S., Ambrosetti, U., Costanzo, S., Mauro, D., Ugazio, G., Langguth, B., Mancuso, A., 2008. Tinnitus aurium in persons with normal hearing: 55 years later. *Otolaryngol Head Neck Surg* 139, 391–394. <https://doi.org/10.1016/j.otohns.2008.06.019>
- Delacre, M., Lakens, D., Leys, C., 2017. Why Psychologists Should by Default Use Welch's t-test Instead of Student's t-test. <https://doi.org/10.5334/irsp.82>
- Devenney, E., Hodges, J.R., 2017. The Mini-Mental State Examination: pitfalls and limitations. *Pr. Neurol* 17, 79–80. <https://doi.org/10.1136/practneurol-2016-001520>
- Devlieger, I., Rosseel, Y., 2019. Multilevel Factor Score Regression. *Multivar. Behav Res* 1–25.
- Diamond, A., 2013. Executive functions. *Annu Rev Psychol* 64, 135–168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- Diamond, A., Ling, D.S., 2016. Conclusions about interventions, programs, and approaches for improving executive functions that appear justified and those that, despite much hype, do not. *Dev Cogn Neurosci* 18, 34–48. <https://doi.org/10.1016/j.dcn.2015.11.005>
- Dienes, Z., 2011. Bayesian Versus Orthodox Statistics: Which Side Are You On? *Perspect Psychol Sci* 6, 274–290. <https://doi.org/10.1177/1745691611406920>
- Dienes, Z., Mclatchie, N., 2018. Four reasons to prefer Bayesian analyses over significance testing. *Psychon Bull Rev* 25, 207–218. <https://doi.org/10.3758/s13423-017-1266-z>
- Dienes, Zoltan, Dr, 2008. *Understanding Psychology as a Science: An Introduction to Scientific and Statistical Inference*, 2008 edition. ed. Palgrave Macmillan.
- Díez-Cirarda, M., Ibarretxe-Bilbao, N., Peña, J., Ojeda, N., 2018. Neurorehabilitation in Parkinson's Disease: A Critical Review of Cognitive Rehabilitation Effects on Cognition and Brain. *Neural Plast* 2018, 2651918. <https://doi.org/10.1155/2018/2651918>
- Distefano, C., Zhu, M., Mindrila, D., 2008. Understanding and Using Factor Scores: Considerations for the Applied Researcher 14.
- Dornhoffer, J., Danner, C., Mennemeier, M., Blake, D., Garcia-Rill, E., 2006. Arousal and attention deficits in patients with tinnitus. *Int Tinnitus J* 12, 9–16.
- Duncan, J., 2013. The structure of cognition: attentional episodes in mind and brain. *Neuron* 80, 35–50. <https://doi.org/10.1016/j.neuron.2013.09.015>
- Duncan, J., 2010. *How Intelligence Happens*. Yale University Press.
- Duncan, J., Emslie, P., Williams, P., Johnson, R., Freer, C., 1996. Intelligence and the Frontal Lobe: The Organization of Goal-Directed Behavior. *Cognit. Psychol.* 30, 257–303.
- Duncan, J., Phillips, L., McLeod, P. (Eds.), 2005. *Measuring the Mind: Speed, control, and age*, First Edition. ed. OUP Oxford.
- Ecker, U.K.H., Lewandowsky, S., Oberauer, K., Chee, A.E.H., 2010. The components of working memory updating: an experimental decomposition and individual differences. *J Exp Psychol Learn Mem Cogn* 36, 170–189. <https://doi.org/10.1037/a0017891>
- Ecker, U.K.H., Oberauer, K., Lewandowsky, S., 2014. Working memory updating involves item-specific removal. *J Mem Lang* 74, 1–15. <https://doi.org/10.1016/j.jml.2014.03.006>
- Edwards, J.D., Fausto, B.A., Tetlow, A.M., Corona, R.T., Valdés, E.G., 2018. Systematic review and meta-analyses of useful field of view cognitive training. *Neurosci Biobehav Rev* 84, 72–91. <https://doi.org/10.1016/j.neubiorev.2017.11.004>

- Eggermont, J.J., 2016. Can Animal Models Contribute to Understanding Tinnitus Heterogeneity in Humans? *Front Aging Neurosci* 8, 265. <https://doi.org/10.3389/fnagi.2016.00265>
- Eggermont, J.J., Roberts, L.E., 2015. Tinnitus: animal models and findings in humans. *Cell Tissue Res* 361, 311–336. <https://doi.org/10.1007/s00441-014-1992-8>
- Ellermeier, W., Zimmer, K., 2014. The psychoacoustics of the irrelevant sound effect. *Acoust Sci Technol* 35, 10–16. <https://doi.org/10.1250/ast.35.10>
- Ellis, P.D., 2010. *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*, 1 edition. ed. Cambridge University Press.
- Etz, A., Gronau, Q.F., Dablander, F., Edelsbrunner, P.A., Baribault, B., 2018. How to become a Bayesian in eight easy steps: An annotated reading list. *Psychon Bull Rev* 25, 219–234. <https://doi.org/10.3758/s13423-017-1317-5>
- Eysenck, M.W., Derakshan, N., 2011. New perspectives in attentional control theory. *Individ Dif* 50, 955–960. <https://doi.org/10.1016/j.paid.2010.08.019>
- Fauci, A.S., Lane, H.C., Redfield, R.R., 2020. Covid-19 — Navigating the Uncharted. *N Engl J Med* 382, 1268–1269. <https://doi.org/10.1056/NEJMe2002387>
- Fernandez-Duque, D., Johnson, M.L., 1999. Attention Metaphors: How Metaphors Guide the Cognitive Psychology of Attention. *Cogn Sci* 23, 83–116.
- Floresco, S.B., Jentsch, J.D., 2011. Pharmacological enhancement of memory and executive functioning in laboratory animals. *Neuropsychopharmacology* 36, 227–250. <https://doi.org/10.1038/npp.2010.158>
- Franke, G.R., 2001. Applications of Meta-Analysis for Marketing and Public Policy: A Review. *J. Public Policy Mark.* 20, 186–200. <https://doi.org/10.1509/jppm.20.2.186>
- Friedman, L., 2001. Why vote-count reviews don't count. *Biol Psychiatry* 49, 161–162. [https://doi.org/10.1016/S0006-3223\(00\)01075-1](https://doi.org/10.1016/S0006-3223(00)01075-1)
- Friedman, N.P., Miyake, A., 2017. Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex* 86, 186–204. <https://doi.org/10.1016/j.cortex.2016.04.023>
- Friedman, N.P., Miyake, A., Corley, R.P., Young, S.E., Defries, J.C., Hewitt, J.K., 2006. Not all executive functions are related to intelligence. *Psychol Sci* 17, 172–179. <https://doi.org/10.1111/j.1467-9280.2006.01681.x>
- Friston, K., 2010. The free-energy principle: a unified brain theory? *Nat Rev Neurosci* 11, 127–138. <https://doi.org/10.1038/nrn2787>
- Fuller, T., Cima, R., Langguth, B., Mazurek, B., Vlaeyen, J.W., Hoare, D.J., 2020. Cognitive behavioural therapy for tinnitus. *Cochrane Database Syst Rev* 1, CD012614. <https://doi.org/10.1002/14651858.CD012614.pub2>
- Funder, D.C., Ozer, D.J., 2019. Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Adv. Methods Pract. Psychol. Sci.* 2, 156–168. <https://doi.org/10.1177/2515245919847202>
- Gaeta, L., Brydges, C.R., 2020. An Examination of Effect Sizes and Statistical Power in Speech, Language, and Hearing Research. *J Speech Lang Hear Res* 1–9. https://doi.org/10.1044/2020_JSLHR-19-00299
- Gates, G.A., Cooper, J.C., 1991. Incidence of hearing decline in the elderly. *Acta Otolaryngol* 111, 240–248.
- Gigerenzer, G., Krauss, S., Vitouch, O., 2004. The Null Ritual: What You Always Wanted to Know About Significance Testing but Were Afraid to Ask, in: Kaplan, D. (Ed.), *The SAGE Handbook of Quantitative Methodology for the Social Sciences*. SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks California 91320 United States of America, pp. 392–409. <https://doi.org/10.4135/9781412986311.n21>
- Gignac, G.E., Szodorai, E.T., 2016. Effect size guidelines for individual differences researchers. *Individ Dif* 102, 74–78. <https://doi.org/10.1016/j.paid.2016.06.069>
- Gilmore, C., Cragg, L., 2018. Chapter 14 - The Role of Executive Function Skills in the Development of Children's Mathematical Competencies, in: Henik, A., Fias, W. (Eds.), *Heterogeneity of Function in Numerical Cognition*. Academic Press, pp. 263–286.

- Goldberg, R.L., Piccirillo, M.L., Nicklaus, J., Skillington, A., Lenze, E., Rodebaugh, T.L., Kallogjeri, D., Piccirillo, J.F., 2017. Evaluation of Ecological Momentary Assessment for Tinnitus Severity. *JAMA Otolaryngol Head Neck Surg* 143, 700–706. <https://doi.org/10.1001/jamaoto.2017.0020>
- Goldstein, S., Naglieri, J.A. (Eds.), 2014. *Handbook of Executive Functioning*, 2014 ed. edition. ed. Springer.
- Goode, A.P., Coeytaux, R.R., Maslow, G.R., Davis, N., Hill, S., Namdari, B., LaPointe, N.M.A., Befus, D., Lallinger, K.R., Bowen, S.E., Kosinski, A., McBroom, A.J., Sanders, G.D., Kemper, A.R., 2018. Nonpharmacologic Treatments for Attention-Deficit/Hyperactivity Disorder: A Systematic Review. *Pediatrics* 141. <https://doi.org/10.1542/peds.2018-0094>
- Goodrich B, Gabry J, Ali I, Brilleman S, 2020. *rstanarm: Bayesian applied regression modeling via Stan*.
- Goodwin, P.E., Johnson, R.M., 1980. The loudness of tinnitus. *Acta Otolaryngol* 90, 353–359.
- Gough, D., 2007. Weight of Evidence: a framework for the appraisal of the quality and relevance of evidence. *Res. Pap. Educ.* 22, 213–228. <https://doi.org/10.1080/02671520701296189>
- Grange, J., 2015. *trimr*. Github.
- Grange, J.A., Becker, R.B., 2019. The Effect of Aging on Response Congruency in Task Switching: A Meta-Analysis. *J Gerontol B Psychol Sci Soc Sci* 74, 389–396. <https://doi.org/10.1093/geronb/gbx122>
- Grant, M.J., Booth, A., 2009. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Info Libr J* 26, 91–108. <https://doi.org/10.1111/j.1471-1842.2009.00848.x>
- Greely, H., Sahakian, B., Harris, J., Kessler, R.C., Gazzaniga, M., Campbell, P., Farah, M.J., 2008. Towards responsible use of cognitive-enhancing drugs by the healthy. *Nature* 456, 702–705. <https://doi.org/10.1038/456702a>
- Groome, D., Sterkaj, F., 2010. Retrieval-induced forgetting and clinical depression. *Cogn. Emot.* 24, 63–70. <https://doi.org/10.1080/02699930802536219>
- Grosz, M.P., Rohrer, J.M., Thoemmes, F., 2020. The Taboo Against Explicit Causal Inference in Nonexperimental Psychology. <https://doi.org/10.31234/osf.io/8hr7n>
- Haig, B.D., 2017. Tests of Statistical Significance Made Sound. *Educ Psychol Meas* 77, 489–506. <https://doi.org/10.1177/0013164416667981>
- Hall, D.A., Fackrell, K., Li, A.B., Thavayogan, R., Smith, S., Kennedy, V., Tinoco, C., Rodrigues, E.D., Campelo, P., Martins, T.D., Lourenço, V.M., Ribeiro, D., Haider, H.F., 2018. A narrative synthesis of research evidence for tinnitus-related complaints as reported by patients and their significant others. *Health Qual Life Outcomes* 16, 61. <https://doi.org/10.1186/s12955-018-0888-9>
- Hallam, R., Rachman, S., Hinchcliffe, R., 1984. Psychological aspects of tinnitus 3, 31–53.
- Hallam, R.S., McKenna, L., Shurlock, L., 2004. Tinnitus impairs cognitive efficiency. *Int J Audiol* 43, 218–226. <https://doi.org/10.1080/14992020400050030>
- Handscorn, L., Shorter, G.W., Hoare, D.J., Hall, D.A., 2019. Evaluation of a Cognitive Behavioral Model of Tinnitus Distress: A Cross-Sectional Study Using Structural Equation Modeling. *Ear Hear*.
- Hartshorne, J.K., Germine, L.T., 2015. When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychol Sci* 26, 433–443. <https://doi.org/10.1177/0956797614567339>
- Hedge, C., Powell, G., Sumner, P., 2018. The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behav Res Methods* 50, 1166–1186. <https://doi.org/10.3758/s13428-017-0935-1>
- Heeren, A., Maurage, P., Perrot, H., De Volder, A., Renier, L., Araneda, R., Lacroix, E., Decat, M., Deggouj, N., Philippot, P., 2014. Tinnitus specifically alters the top-down executive control sub-component of attention: evidence from the Attention Network Task. *Behav Brain Res* 269, 147–154. <https://doi.org/10.1016/j.bbr.2014.04.043>

- Heitz, R.P., 2014. The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Front Neurosci* 8, 150. <https://doi.org/10.3389/fnins.2014.00150>
- Heller, A.J., 2003. Classification and epidemiology of tinnitus. *Otolaryngol Clin North Am* 36, 239–248.
- Heller, M.F., Bergman, M., 1953. Tinnitus aurium in normally hearing persons. *Ann Otol Rhinol Laryngol* 62, 73–83. <https://doi.org/10.1177/000348945306200107>
- Henry, J.A., 2016. "Measurement" of Tinnitus. *Otol Neurotol* 37, e276–85. <https://doi.org/10.1097/MAO.0000000000001070>
- Henry, J.A., Dennis, K.C., Schechter, M.A., 2005. General review of tinnitus: prevalence, mechanisms, effects, and management. *J Speech Lang Hear Res* 48, 1204–1235. [https://doi.org/10.1044/1092-4388\(2005/084\)](https://doi.org/10.1044/1092-4388(2005/084))
- Henry, J.A., Meikle, M.B., 2000. Psychoacoustic measures of tinnitus. *J Am Acad Audiol* 11, 138–155.
- Henshaw, H., Ferguson, M.A., 2013. Efficacy of individual computer-based auditory training for people with hearing loss: a systematic review of the evidence. *PLoS One* 8, e62836. <https://doi.org/10.1371/journal.pone.0062836>
- Hernán, M.A., 2018. The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data. *Am J Public Health* 108, 616–619. <https://doi.org/10.2105/AJPH.2018.304337>
- Hershberger, S.L., Marcoulides, G.A., 2013. The problem of equivalent structural models. *Struct. Equ. Model. Second Course 2nd Ed* 2, 3–39.
- Hesser, H., Molander, P., Jungermann, M., Andersson, G., 2013. Costs of suppressing emotional sound and countereffects of a mindfulness induction: an experimental analog of tinnitus impact. *PLoS One* 8, e64540. <https://doi.org/10.1371/journal.pone.0064540>
- Higgins, J.P.T., Thompson, S.G., Deeks, J.J., Altman, D.G., 2003. Measuring inconsistency in meta-analyses. *BMJ* 327, 557–560. <https://doi.org/10.1136/bmj.327.7414.557>
- Hill, N.L., Mogle, J., Wion, R., Munoz, E., DePasquale, N., Yevchak, A.M., Parisi, J.M., 2016. Subjective Cognitive Impairment and Affective Symptoms: A Systematic Review. *Gerontologist* 56, e109–e127. <https://doi.org/10.1093/geront/gnw091>
- Hoare, D.J., Stacey, P.C., Hall, D.A., 2010. The efficacy of auditory perceptual training for tinnitus: a systematic review. *Ann Behav Med* 40, 313–324. <https://doi.org/10.1007/s12160-010-9213-5>
- Hoare, D.J., Van Labeke, N., McCormack, A., Sereda, M., Smith, S., Al Taher, H., Kowalkowski, V.L., Sharples, M., Hall, D.A., 2014. Gameplay as a source of intrinsic motivation in a randomized controlled trial of auditory training for tinnitus. *PLoS One* 9, e107430. <https://doi.org/10.1371/journal.pone.0107430>
- Hoekstra, C.E., Rynja, S.P., van Zanten, G.A., Rovers, M.M., 2011. Anticonvulsants for tinnitus. *Cochrane Database Syst Rev* CD007960. <https://doi.org/10.1002/14651858.CD007960.pub2>
- Hunter, A., Edmondson-Jones, M., Somerset, S., Hall, D.A., 2016. A systematic review of the reporting of tinnitus prevalence and severity. *Hear Res* 337, 219. <https://doi.org/10.1016/j.heares.2016.05.009>
- Hunter, J.E., Schmidt, F.L., 2004. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. SAGE.
- Husain, F.T., Akrofi, K., Carpenter-Thompson, J.R., Schmidt, S.A., 2015. Alterations to the attention system in adults with tinnitus are modality specific. *Brain Res* 1620, 81–97. <https://doi.org/10.1016/j.brainres.2015.05.010>
- Imbens, G.W., Rubin, D.B., 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.
- Imel, Z.E., Sheng, E., Baldwin, S.A., Atkins, D.C., 2015. Removing very low-performing therapists: A simulation of performance-based retention in psychotherapy. *Psychotherapy* 52, 329–336. <https://doi.org/10.1037/pst0000023>
- Ivansic, D., Guntinas-Lichius, O., Müller, B., Volk, G.F., Schneider, G., Dobel, C., 2017. Impairments of Speech Comprehension in Patients with Tinnitus-A Review. *Front Aging Neurosci* 9, 224.

- Jack, B., Louk, V., Bart, E., Paul, W., 2018. The pharmacology of executive functioning. *Behav Pharmacol* 29, 557–558. <https://doi.org/10.1097/FBP.0000000000000444>
- Jackson, J.G., Coyne, I.J., Clough, P.J., 2014. A preliminary investigation of potential cognitive performance decrements in non-help-seeking tinnitus sufferers. *Int J Audiol* 53, 88–93. <https://doi.org/10.3109/14992027.2013.846481>
- Jain, A., Bansal, R., Kumar, A., Singh, K.D., 2015. A comparative study of visual and auditory reaction times on the basis of gender and physical activity levels of medical first year students. *Int J Appl Basic Med Res* 5, 124–127.
- Jastreboff, P.J., 1990. Phantom auditory perception (tinnitus): mechanisms of generation and perception. *Neurosci Res* 8, 221–254.
- Jewsbury, P.A., Bowden, S.C., Duff, K., 2016. The Cattell–Horn–Carroll Model of Cognition for Clinical Assessment. *J Psychoeduc Assess* 35, 547–567. <https://doi.org/10.1177/0734282916651360>
- Johnson, D.P., Whisman, M.A., 2013. Gender differences in rumination: A meta-analysis. *Individ Dif* 55, 367–374.
- Johnson, M.K., 1983. A Multiple-Entry, Modular Memory System, in: Bower, G.H. (Ed.), *Psychology of Learning and Motivation*. Academic Press, pp. 81–123. [https://doi.org/10.1016/S0079-7421\(08\)60097-3](https://doi.org/10.1016/S0079-7421(08)60097-3)
- Johnson, V.E., 2013. Revised standards for statistical evidence. *Proc Natl Acad Sci U A* 110, 19313–19317. <https://doi.org/10.1073/pnas.1313476110>
- Juarascio, A.S., Manasse, S.M., Espel, H.M., Kerrigan, S.G., Forman, E.M., 2015. Could training executive function improve treatment outcomes for eating disorders? *Appetite* 90, 187–193. <https://doi.org/10.1016/j.appet.2015.03.013>
- Kaiser, H.F., Rice, J., 1974. Little Jiffy, Mark Iv. *Educ Psychol Meas* 34, 111–117.
- Kallogjeri, D., Piccirillo, J.F., Spitznagel, E., Jr, Hale, S., Nicklaus, J.E., Hardin, F.M., Shimony, J.S., Coalson, R.S., Schlaggar, B.L., 2017. Cognitive Training for Adults With Bothersome Tinnitus: A Randomized Clinical Trial. *JAMA Otolaryngol Head Neck Surg* 143, 443–451. <https://doi.org/10.1001/jamaoto.2016.3779>
- Kaltenbach, J.A., Godfrey, D.A., Neumann, J.B., McCaslin, D.L., Afman, C.E., Zhang, J., 1998. Changes in spontaneous neural activity in the dorsal cochlear nucleus following exposure to intense sound: relation to threshold shift. *Hear Res* 124, 78–84. [https://doi.org/10.1016/S0378-5955\(98\)00119-1](https://doi.org/10.1016/S0378-5955(98)00119-1)
- Kandeean, S., Maudoux, A., Ribeiro de Paula, D., Zheng, J.Y., Cabay, J.E., Gómez, F., Chronik, B.A., Ridder, D., Vanneste, S., Soddu, A., 2019. Tinnitus distress: a paradoxical attention to the sound? *J Neurol*. <https://doi.org/10.1007/s00415-019-09390-1>
- Katz, B., Shah, P., Meyer, D.E., 2018. How to play 20 questions with nature and lose: Reflections on 100 years of brain-training research. *Proc Natl Acad Sci U A* 115, 9897–9904. <https://doi.org/10.1073/pnas.1617102114>
- Kaya, E.M., Elhilali, M., 2017. Modelling auditory attention. *Philos Trans R Soc Lond B Biol Sci* 372. <https://doi.org/10.1098/rstb.2016.0101>
- Khan, R.A., Husain, F.T., 2020. Tinnitus and cognition: Can load theory help us refine our understanding? *Laryngoscope Investig. Otolaryngol.* 19, 1045. <https://doi.org/10.1002/lio2.501>
- Kline, R.B., 2016. *Principles and Practice of Structural Equation Modeling*, 4th ed. Guilford Publications.
- Kock, A., Gemünden, H.G., 2009. A guideline to meta-analysis. TIM working paper series. Technische Universität Berlin.
- Kover, S.T., Atwood, A.K., 2013. Establishing equivalence: methodological progress in group-matching design and analysis. *Am J Intellect Dev Disabil* 118, 3–15.
- Kredlow, M.A., Keshishian, A., Oppenheimer, S., Otto, M.W., 2019. The Efficacy of Modafinil as a Cognitive Enhancer: A Systematic Review and Meta-Analysis. *J Clin Psychopharmacol* 39, 455–461.
- Krings, J.G., Wineland, A., Kallogjeri, D., Rodebaugh, T.L., Nicklaus, J., Lenze, E.J., Piccirillo, J.F., 2015. A novel treatment for tinnitus and tinnitus-related cognitive difficulties using computer-based cognitive training and D-cycloserine. *JAMA*

- Otolaryngol Head Neck Surg 141, 18–26.
<https://doi.org/10.1001/jamaoto.2014.2669>
- Kruschke, J., 2014. Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. Academic Press.
- Kruschke, J.K., 2010. What to believe: Bayesian methods for data analysis. Trends Cogn Sci 14, 293–300. <https://doi.org/10.1016/j.tics.2010.05.001>
- Kruschke, J.K., Liddell, T.M., 2018. Bayesian data analysis for newcomers. Psychon Bull Rev 25, 155–177. <https://doi.org/10.3758/s13423-017-1272-1>
- Langguth, B., Goodey, R., Azevedo, A., Bjorne, A., Cacace, A., Crocetti, A., Del Bo, L., De Ridder, D., Diges, I., Elbert, T., Flor, H., Herraiz, C., Ganz Sanchez, T., Eichhammer, P., Figueiredo, R., Hajak, G., Kleinjung, T., Landgrebe, M., Londero, A., Lainez, M.J.A., Mazzoli, M., Meikle, M.B., Melcher, J., Rauschecker, J.P., Sand, P.G., Struve, M., Van de Heyning, P., Van Dijk, P., Vergara, R., 2007. Consensus for tinnitus patient assessment and treatment outcome measurement: Tinnitus Research Initiative meeting, Regensburg, July 2006. Prog Brain Res 166, 525–536. [https://doi.org/10.1016/S0079-6123\(07\)66050-6](https://doi.org/10.1016/S0079-6123(07)66050-6)
- Langguth, B., Salvi, R., Elgoyhen, A.B., 2009. Emerging pharmacotherapy of tinnitus. Expert Opin Emerg Drugs 14, 687–702.
<https://doi.org/10.1517/14728210903206975>
- Lavie, N., 2010. Attention, Distraction, and Cognitive Control Under Load 19, 143–148.
<https://doi.org/10.1177/0963721410370295>
- Lederer, D.J., Bell, S.C., Branson, R.D., Chalmers, J.D., Marshall, R., Maslove, D.M., Ost, D.E., Punjabi, N.M., Schatz, M., Smyth, A.R., Stewart, P.W., Suissa, S., Adjei, A.A., Akdis, C.A., Azoulay, É., Bakker, J., Ballas, Z.K., Bardin, P.G., Barreiro, E., Bellomo, R., Bernstein, J.A., Brusasco, V., Buchman, T.G., Chokroverty, S., Collop, N.A., Crapo, J.D., Fitzgerald, D.A., Hale, L., Hart, N., Herth, F.J., Iwashyna, T.J., Jenkins, G., Kolb, M., Marks, G.B., Mazzone, P., Moorman, J.R., Murphy, T.M., Noah, T.L., Reynolds, P., Riemann, D., Russell, R.E., Sheikh, A., Sotgiu, G., Swenson, E.R., Szczesniak, R., Szymusiak, R., Teboul, J.-L., Vincent, J.-L., 2019. Control of Confounding and Reporting of Results in Causal Inference Studies. Guidance for Authors from Editors of Respiratory, Sleep, and Critical Care Journals. Ann Am Thorac Soc 16, 22–28.
<https://doi.org/10.1513/AnnalsATS.201808-564PS>
- Lee, S.J., Kim, H., Kim, L.-S., Kim, J.-H., Park, K.W., 2019. Effects of frontal-executive dysfunction on self-perceived hearing handicap in the elderly with mild cognitive impairment. PLoS One 14, e0210014.
- Lee, S.-Y., Lee, J.Y., Han, S.-Y., Seo, Y., Shim, Y.J., Kim, Y.H., 2019. Neurocognition of Aged Patients With Chronic Tinnitus: Focus on Mild Cognitive Impairment. Clin Exp Otorhinolaryngol. <https://doi.org/10.21053/ceo.2018.01914>
- Leong, S.L., Tchen, S., Robertson, I.H., Alsalman, O., To, W.T., Vanneste, S., 2020. The potential interruptive effect of tinnitus-related distress on attention. Sci Rep 10, 11911.
- Levine, R.A., 1999. Somatic (craniocervical) tinnitus and the dorsal cochlear nucleus hypothesis. Am J Otolaryngol 20, 351–362.
- Lewis, M., Kuerbis, A., 2016. An Overview of Causal Directed Acyclic Graphs for Substance Abuse Researchers. J Drug Alcohol Res 5, 1–8.
<https://doi.org/10.4303/jdar/235992>
- Li, Z., Gu, R., Zeng, X., Qi, M., Cen, J., Zhang, S., Gu, J., Chen, Q., 2018. Eyes and Ears: Cross-Modal Interference of Tinnitus on Visual Processing. Front Psychol 9, 1779. <https://doi.org/10.3389/fpsyg.2018.01779>
- Lima, D.O., Araújo, A.M.G.D. de, Branco-Barreiro, F.C.A., Carneiro, C. da S., Almeida, L.N.A., Rosa, M.R.D. da, 2019. Auditory attention in individuals with tinnitus. Braz J Otorhinolaryngol. <https://doi.org/10.1016/j.bjorl.2019.01.011>
- Lindeløv, J.K., 2019. Reaction time distributions: an interactive overview [WWW Document]. Simple List My Tutor. Hosted GitHub. URL
<https://lindeloev.github.io/shiny-rt/>
- Lipsey, M.W., Wilson, D.B., 2000. Practical Meta-Analysis. SAGE Publications.

- Liu, Y.-Q., Chen, Z.-J., Li, G., Lai, D., Liu, P., Zheng, Y., 2018. Effects of Educational Counseling as Solitary Therapy for Chronic Primary Tinnitus and Related Problems. *Biomed Res Int* 2018, 6032525. <https://doi.org/10.1155/2018/6032525>
- MacLeod, C.M., 2014. The Stroop Effect, in: Luo, R. (Ed.), *Encyclopedia of Color Science and Technology*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–6. https://doi.org/10.1007/978-3-642-27851-8_67-1
- Macmillan, M., 2002. *An Odd Kind of Fame: Stories of Phineas Gage (A Bradford Book)*, New Ed edition. ed. MIT Press.
- Makowski, D., Ben-Shachar, M., Lüdtke, D., 2019. bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. *JOSS* 4, 1541. <https://doi.org/10.21105/joss.01541>
- Manning, C., Grush, L., Thielman, E., Roberts, L., Henry, J.A., 2019. Comparison of Tinnitus Loudness Measures: Matching, Rating, and Scaling. *Am J Audiol* 28, 137–143. https://doi.org/10.1044/2018_AJA-17-0115
- Mather, N., Jaffe, L.E., 2016. *Woodcock-Johnson IV: Reports, Recommendations, and Strategies*. John Wiley & Sons.
- Matsunaga, M., 2010. How to Factor-Analyze Your Data Right: Do's, Don'ts, and How-To's. *Int. J. Psychol. Res.* 3, 97–110.
- Mayo, D.G., 2018. *Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars*. Cambridge University Press. <https://doi.org/10.1017/9781107286184>
- Mazurek, B., Boecking, B., Brüeggemann, P., 2019. Association Between Stress and Tinnitus-New Aspects. *Otol Neurotol* 40, e467–e473. <https://doi.org/10.1097/MAO.0000000000002180>
- McCombe, A., Baguley, D., Coles, R., McKenna, L., McKinney, C., Windle-Taylor, P., British Association of Otolaryngologists, H. and N.S., 2001. Guidelines for the grading of tinnitus severity: the results of a working group commissioned by the British Association of Otolaryngologists, Head and Neck Surgeons, 1999. *Clin Otolaryngol Allied Sci* 26, 388–393. <https://doi.org/10.1046/j.1365-2273.2001.00490.x>
- McElreath, R., 2016. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press.
- McFerran, D., Hoare, D.J., Carr, S., Ray, J., Stockdale, D., 2018. Tinnitus services in the United Kingdom: a survey of patient experiences. *BMC Health Serv Res* 18, 110.
- McFerran, D.J., Stockdale, D., Holme, R., Large, C.H., Baguley, D.M., 2019. Why Is There No Cure for Tinnitus? *Front Neurosci* 13, 802. <https://doi.org/10.3389/fnins.2019.00802>
- McGrath, R.E., Meyer, G.J., 2006. When effect sizes disagree: the case of r and d. *Psychol Methods* 11, 386–401. <https://doi.org/10.1037/1082-989X.11.4.386>
- McGrew, K.S., 2009. CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence* 37, 1–10. <https://doi.org/10.1016/j.intell.2008.08.004>
- McKENNA, L., 1997. *PSYCHOLOGICAL ASPECTS OF AUDITORY DISORDERS: COGNITIVE FUNCTIONING AND PSYCHOLOGICAL STATE*. The City University, London.
- McKenna, L., Handscomb, L., Hoare, D.J., Hall, D.A., 2014. A scientific cognitive-behavioral model of tinnitus: novel conceptualizations of tinnitus distress. *Front Neurol* 5, 196. <https://doi.org/10.3389/fneur.2014.00196>
- McMillan, G.P., Cannon, J.B., 2019. Bayesian Applications in Auditory Research. *J Speech Lang Hear Res* 62, 577–586. https://doi.org/10.1044/2018_JSLHR-H-ASTM-18-0228
- McMillan, G.P., Thielman, E.J., Wypych, K., Henry, J.A., 2014. A Bayesian perspective on tinnitus pitch matching. *Ear Hear* 35, 687–694. <https://doi.org/10.1097/AUD.000000000000081>
- Mehlman, M.J., 2004. Cognition-enhancing drugs. *Milbank Q* 82, 483–506, table of contents. <https://doi.org/10.1111/j.0887-378X.2004.00319.x>

- Meikle, M.B., Vernon, J., Johnson, R.M., 1984. The perceived severity of tinnitus. Some observations concerning a large population of tinnitus clinic patients. *Otolaryngol Head Neck Surg* 92, 689–696. <https://doi.org/10.1177/019459988409200617>
- Melby-Lervåg, M., Redick, T.S., Hulme, C., 2016. Working Memory Training Does Not Improve Performance on Measures of Intelligence or Other Measures of “Far Transfer”: Evidence From a Meta-Analytic Review. *Perspect Psychol Sci* 11, 512–534. <https://doi.org/10.1177/1745691616635612>
- Mervis, C.B., Klein-Tasman, B.P., 2004. Methodological issues in group-matching designs: alpha levels for control variable comparisons and measurement characteristics of control and target variables. *J Autism Dev Disord* 34, 7–17.
- Mervis, C.B., Robinson, B.F., 1999. Methodological issues in cross-syndrome comparisons: matching procedures, sensitivity (Se), and specificity (Sp). *Monogr Soc Res Child Dev* 64, 115–30; discussion 131–9.
- Michl, L.C., McLaughlin, K.A., Shepherd, K., Nolen-Hoeksema, S., 2013. Rumination as a mechanism linking stressful life events to symptoms of depression and anxiety: longitudinal evidence in early adolescents and adults. *J Abnorm Psychol* 122, 339–352.
- Milner, B., 1963. Effects of Different Brain Lesions on Card Sorting: The Role of the Frontal Lobes. *Arch Neurol* 9, 90–100.
- Miyake, A., Friedman, N.P., 2012. The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Curr Dir Psychol Sci* 21, 8–14. <https://doi.org/10.1177/0963721411429458>
- Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., Wager, T.D., 2000. The Unity and Diversity of Executive Functions and Their Contributions to Complex “Frontal Lobe” Tasks: A Latent Variable Analysis. *Cogn Psychol* 41, 49–100. <https://doi.org/10.1006/cogp.1999.0734>
- Moeyaert, M., Ugille, M., Natasha Beretvas, S., Ferron, J., Bunuan, R., Van den Noortgate, W., 2017. Methods for dealing with multiple outcomes in meta-analysis: a comparison between averaging effect sizes, robust variance estimation and multilevel meta-analysis. *Int J Soc Res Methodol* 20, 559–572. <https://doi.org/10.1080/13645579.2016.1252189>
- Mohamad, N., 2015. Tinnitus: priority research questions and the impact of the condition for working memory and attention. University of Nottingham.
- Mohamad, N., Hoare, D.J., Hall, D.A., 2016. The consequences of tinnitus and tinnitus severity on cognition: A review of the behavioural evidence. *Hear Res* 332, 199–209. <https://doi.org/10.1016/j.heares.2015.10.001>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., PRISMA Group, 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 6, e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- Møller, A.R., Langguth, B., De Ridder, D., Kleinjung, T., 2010. Textbook of Tinnitus. Springer Science & Business Media.
- Moore, B.C.J., 2012. The Psychophysics of Tinnitus, in: Eggermont, J.J., Zeng, F.-G., Popper, A.N., Fay, R.R. (Eds.), Tinnitus. Springer New York, New York, NY, pp. 187–216. https://doi.org/10.1007/978-1-4614-3728-4_9
- Munn, Z., Stern, C., Aromataris, E., Lockwood, C., Jordan, Z., 2018. What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. *BMC Med Res Methodol* 18, 5. <https://doi.org/10.1186/s12874-017-0468-4>
- Murayama, K., Miyatsu, T., Buchli, D., Storm, B.C., 2014. Forgetting as a consequence of retrieval: a meta-analytic review of retrieval-induced forgetting. *Psychol Bull* 140, 1383–1409. <https://doi.org/10.1037/a0037505>
- Murphy, G., Groeger, J.A., Greene, C.M., 2016. Twenty years of load theory—Where are we now, and where should we go next? *Psychon Bull Rev* 23, 1316–1340. <https://doi.org/10.3758/s13423-015-0982-5>
- Murphy, S., Spence, C., Dalton, P., 2017. Auditory perceptual load: A review. *Hear Res* 352, 40–48. <https://doi.org/10.1016/j.heares.2017.02.005>

- Muth, C., Oravecz, Z., Gabry, J., 2018. User-friendly Bayesian regression modeling: A tutorial with rstanarm and shinystan. *TQMP* 14, 99–119. <https://doi.org/10.20982/tqmp.14.2.p099>
- Nagaraj, M.K., Bhaskar, A., Prabhu, P., 2020. Assessment of auditory working memory in normal hearing adults with tinnitus. *Eur Arch Otorhinolaryngol* 277, 47–54.
- Nasreddine, Z.S., Phillips, N.A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J.L., Chertkow, H., 2005. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 53, 695–699. <https://doi.org/10.1111/j.1532-5415.2005.53221.x>
- Newman, C.W., Sandridge, S.A., Jacobson, G.P., 1998. Psychometric adequacy of the Tinnitus Handicap Inventory (THI) for evaluating treatment outcome. *J Am Acad Audiol* 9, 153–160.
- NICE, 2020. Tinnitus: assessment and management.
- Nigg, J.T., 2017. Annual Research Review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *J Child Psychol Psychiatry* 58, 361–383. <https://doi.org/10.1111/jcpp.12675>
- Noda, K., Kitahara, T., Doi, K., 2018. Sound Change Integration Error: An Explanatory Model of Tinnitus. *Front Neurosci* 12, 831. <https://doi.org/10.3389/fnins.2018.00831>
- Nolen-Hoeksema, S., 1987. Sex differences in unipolar depression: evidence and theory. *Psychol Bull* 101, 259–282. <https://doi.org/10.1037/0033-2909.101.2.259>
- Noreña, A.J., 2011. An integrative model of tinnitus based on a central gain controlling neural sensitivity. *Neurosci Biobehav Rev* 35, 1089–1109. <https://doi.org/10.1016/j.neubiorev.2010.11.003>
- Noreña, A.J., Eggermont, J.J., 2003. Changes in spontaneous neural activity immediately after an acoustic trauma: implications for neural correlates of tinnitus. *Hear Res* 183, 137–153.
- Nucci, M., Mapelli, D., Mondini, S., 2012. Cognitive Reserve Index questionnaire (CRIq): a new instrument for measuring cognitive reserve. *Aging Clin Exp Res* 24, 218–226. <https://doi.org/10.3275/7800>
- Nunnari, D., De Cola, M.C., Costa, A., Rifici, C., Bramanti, P., Marino, S., 2016. Exploring cognitive reserve in multiple sclerosis: New findings from a cross-sectional study. *J Clin Exp Neuropsychol* 38, 1158–1167. <https://doi.org/10.1080/13803395.2016.1200538>
- Nyberg, L., Lövdén, M., Riklund, K., Lindenberger, U., Bäckman, L., 2012. Memory aging and brain maintenance. *Trends Cogn Sci* 16, 292–305. <https://doi.org/10.1016/j.tics.2012.04.005>
- Oberauer, K., 2019. Working Memory Capacity Limits Memory for Bindings. *J Cogn* 2, 40. <https://doi.org/10.5334/joc.86>
- Odum, M., 2011. Factor Scores, Structure and Commuality Coefficients: A Primer. Online Submiss.
- Opdebeeck, C., Martyr, A., Clare, L., 2016. Cognitive reserve and cognitive function in healthy older people: a meta-analysis. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn* 23, 40–60. <https://doi.org/10.1080/13825585.2015.1041450>
- Open Science Collaboration, 2015. PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science* 349, aac4716. <https://doi.org/10.1126/science.aac4716>
- Pearce, N., Lawlor, D.A., 2016. Causal inference-so much more than statistics. *Int J Epidemiol* 45, 1895–1903. <https://doi.org/10.1093/ije/dyw328>
- Pearl, J., 2009. Causality: Models, Reasoning and Interference, 2Rev e. edition. ed. Cambridge University Press.
- Pearl, J., Glymour, M., Jewell, N.P., 2016. Causal Inference in Statistics - A Primer, Pap/Psc edition. ed. Wiley.
- Pearl, J., Mackenzie, D., 2018. The Book of Why: The New Science of Cause and Effect. Allen Lane.

- Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J.K., 2019. PsychoPy2: Experiments in behavior made easy. *Behav Res Methods* 51, 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- Peker, S., Sirin, A., 2016. Parallels between phantom pain and tinnitus. *Med Hypotheses* 91, 95–97. <https://doi.org/10.1016/j.mehy.2016.04.023>
- Perfect, T.J., Maylor, E.A., 2000. *Models of cognitive aging*, *Debates in psychology*. Oxford University Press, Oxford.
- Perneger, T.V., 1998. What's wrong with Bonferroni adjustments. *BMJ* 316, 1236–1238. <https://doi.org/10.1136/bmj.316.7139.1236>
- Petersen, S.E., Posner, M.I., 2012. The attention system of the human brain: 20 years after. *Annu Rev Neurosci* 35, 73–89. <https://doi.org/10.1146/annurev-neuro-062111-150525>
- Phillips, J.S., McFerran, D.J., Hall, D.A., Hoare, D.J., 2018. The natural history of subjective tinnitus in adults: A systematic review and meta-analysis of no-intervention periods in controlled trials. *Laryngoscope* 128, 217–227. <https://doi.org/10.1002/lary.26607>
- Pierce, K.J., Kallogjeri, D., Piccirillo, J.F., Garcia, K.S., Nicklaus, J.E., Burton, H., 2012. Effects of severe bothersome tinnitus on cognitive function measured with standardized tests. *J Clin Exp Neuropsychol* 34, 126–134. <https://doi.org/10.1080/13803395.2011.623120>
- Polanin, J.R., Snijlsteit, B., 2016. Converting between effect sizes. <https://doi.org/10.4073/cmpn.2016.3>
- Quintana, D.S., 2017. Statistical considerations for reporting and planning heart rate variability case-control studies. *Psychophysiology* 54, 344–349. <https://doi.org/10.1111/psyp.12798>
- Quintana, D.S., 2015. From pre-registration to publication: a non-technical primer for conducting a meta-analysis to synthesize correlational data. *Front Psychol* 6, 1549. <https://doi.org/10.3389/fpsyg.2015.01549>
- Quintana, D.S., Williams, D.R., 2018. Bayesian alternatives for common null-hypothesis significance tests in psychiatry: a non-technical guide using JASP. *BMC Psychiatry* 18, 178. <https://doi.org/10.1186/s12888-018-1761-4>
- R Core Team, 2018. *R: A language and environment for statistical computing*.
- Rabbitt, P., 2004. *Methodology Of Frontal And Executive Function*. Psychology Press.
- Ratcliff, R., 1993. Methods for dealing with reaction time outliers. *Psychol Bull* 114, 510–532.
- Ratcliff, R., Smith, P.L., Brown, S.D., McKoon, G., 2016. Diffusion Decision Model: Current Issues and History. *Trends Cogn Sci* 20, 260–281.
- Rauschecker, J.P., Leaver, A.M., Mühlau, M., 2010. Tuning out the noise: limbic-auditory interactions in tinnitus. *Neuron* 66, 819–826. <https://doi.org/10.1016/j.neuron.2010.04.032>
- Rauschecker, J.P., May, E.S., Maudoux, A., Ploner, M., 2015. Frontostriatal Gating of Tinnitus and Chronic Pain. *Trends Cogn Sci* 19, 567–578. <https://doi.org/10.1016/j.tics.2015.08.002>
- Raykov, T., Penev, S., 1999. On Structural Equation Model Equivalence. *Multivar. Behav Res* 34, 199–244. <https://doi.org/10.1207/S15327906Mb340204>
- Reiser, B., Faraggi, D., 1999. Confidence Intervals for the Overlapping Coefficient: The Normal Equal Variance Case. *J. R. Stat. Soc. Ser. Stat.* 48, 413–418.
- Revelle, W., 2013. *An introduction to psychometric theory with applications in R*.
- Revelle, W.R., 2017. *psych: Procedures for Personality and Psychological Research*.
- Roberts, L.E., Husain, F.T., Eggermont, J.J., 2013. Role of attention in the generation and modulation of tinnitus. *Neurosci Biobehav Rev* 37, 1754–1773. <https://doi.org/10.1016/j.neubiorev.2013.07.007>
- Rock, P.L., Roiser, J.P., Riedel, W.J., Blackwell, A.D., 2014. Cognitive impairment in depression: a systematic review and meta-analysis. *Psychol Med* 44, 2029–2040. <https://doi.org/10.1017/S0033291713002535>

- Rohrer, J.M., 2018. Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. *Adv. Methods Pract. Psychol. Sci.* 1, 27–42. <https://doi.org/10.1177/2515245917745629>
- Roldán-Tapia, L., García, J., Cánovas, R., León, I., 2012. Cognitive reserve, age, and their relation to attentional and executive functions. *Appl Neuropsychol Adult* 19, 2–8. <https://doi.org/10.1080/09084282.2011.595458>
- Román, P., Soriano, M.F., Gómez-Ariza, C.J., Bajo, M.T., 2009. Retrieval-induced forgetting and executive control. *Psychol Sci* 20, 1053–1058. <https://doi.org/10.1111/j.1467-9280.2009.02415.x>
- Rose, S., Laan, M.J. van der, 2009. Why match? Investigating matched case-control study designs with causal effect estimation. *Int J Biostat* 5, Article 1.
- Rosenhall, U., Möller, C., Hederstierna, C., 2013. Hearing of 75-year old persons over three decades: has hearing changed? *Int J Audiol* 52, 731–739. <https://doi.org/10.3109/14992027.2013.808381>
- Rosenthal, R., DiMatteo, M.R., 2001. Meta-analysis: recent developments in quantitative methods for literature reviews. *Annu Rev Psychol* 52, 59–82. <https://doi.org/10.1146/annurev.psych.52.1.59>
- Rossiter, S., Stevens, C., Walker, G., 2006. Tinnitus and its effect on working memory and attention. *J Speech Lang Hear Res* 49, 150–160. [https://doi.org/10.1044/1092-4388\(2006/012\)](https://doi.org/10.1044/1092-4388(2006/012))
- Ruscio, J., Mullen, T., 2012. Confidence Intervals for the Probability of Superiority Effect Size Measure and the Area Under a Receiver Operating Characteristic Curve. *Multivar. Behav Res* 47, 201–223. <https://doi.org/10.1080/00273171.2012.658329>
- Salthouse, T.A., 1996. The processing-speed theory of adult age differences in cognition. *Psychol Rev* 103, 403–428. <https://doi.org/10.1037/0033-295x.103.3.403>
- Sandi, C., 2013. Stress and cognition. *Wiley Interdiscip Rev Cogn Sci* 4, 245–261. <https://doi.org/10.1002/wcs.1222>
- Savastano, M., 2008. Tinnitus with or without hearing loss: are its characteristics different? *Eur Arch Otorhinolaryngol* 265, 1295–1300. <https://doi.org/10.1007/s00405-008-0630-z>
- Schaette, R., McAlpine, D., 2011. Tinnitus with a normal audiogram: physiological evidence for hidden hearing loss and computational model. *J Neurosci* 31, 13452–13457. <https://doi.org/10.1523/JNEUROSCI.2156-11.2011>
- Schaette, R., Turtle, C., Munro, K.J., 2012. Reversible induction of phantom auditory sensations through simulated unilateral hearing loss. *PLoS One* 7, e35238. <https://doi.org/10.1371/journal.pone.0035238>
- Schäfer, T., Schwarz, M.A., 2019. The Meaningfulness of Effect Sizes in Psychological Research: Differences Between Sub-Disciplines and the Impact of Potential Biases. *Front Psychol* 10, 813. <https://doi.org/10.3389/fpsyg.2019.00813>
- Schick-Makaroff, K., MacDonald, M., Plummer, M., Burgess, J., Neander, W., 2016. What Synthesis Methodology Should I Use? A Review and Analysis of Approaches to Research Synthesis. *AIMS Public Health* 3, 172–215. <https://doi.org/10.3934/publichealth.2016.1.172>
- Schmidt, C.J., Kerns, R.D., Griest, S., Theodoroff, S.M., Pietrzak, R.H., Henry, J.A., 2014. Toward development of a tinnitus magnitude index. *Ear Hear* 35, 476–484.
- Schmiedek, F., Hildebrandt, A., Lövdén, M., Wilhelm, O., Lindenberger, U., 2009. Complex span versus updating tasks of working memory: the gap is not that deep. *J Exp Psychol Learn Mem Cogn* 35, 1089–1096. <https://doi.org/10.1037/a0015730>
- Schultz, I.Z., Sepehry, A.A., Greer, S.C., 2018. Impact of Tinnitus on Cognitive Function in Forensic Neuropsychology Context. *Psychol Inj Law*. <https://doi.org/10.1007/s12207-018-9321-2>
- Searchfield, G.D., Kobayashi, K., Sanders, M., 2012. An adaptation level theory of tinnitus audibility. *Front Syst Neurosci* 6, 46.

- Searchfield, G.D., Morrison-Low, J., Wise, K., 2007. Object identification and attention training for treating tinnitus. *Prog Brain Res* 166, 441–460.
[https://doi.org/10.1016/S0079-6123\(07\)66043-9](https://doi.org/10.1016/S0079-6123(07)66043-9)
- Sedley, W., 2019. Tinnitus: Does Gain Explain? *Neuroscience* 407, 213–228.
<https://doi.org/10.1016/j.neuroscience.2019.01.027>
- Sedley, W., Friston, K.J., Gander, P.E., Kumar, S., Griffiths, T.D., 2016. An Integrative Tinnitus Model Based on Sensory Precision. *Trends Neurosci* 39, 799–812.
<https://doi.org/10.1016/j.tins.2016.10.004>
- Sereda, M., Xia, J., El Refaie, A., Hall, D.A., Hoare, D.J., 2018. Sound therapy (using amplification devices and/or sound generators) for tinnitus. *Cochrane Database Syst Rev* 12, CD013094. <https://doi.org/10.1002/14651858.CD013094.pub2>
- Shallice, T., 1982. Specific impairments of planning. *Philos Trans R Soc Lond B Biol Sci* 298, 199–209. <https://doi.org/10.1098/rstb.1982.0082>
- Shallice, T., Burgess, P., 1996. The domain of supervisory processes and temporal organization of behaviour. *Philos Trans R Soc Lond B Biol Sci* 351, 1405–11; discussion 1411–2. <https://doi.org/10.1098/rstb.1996.0124>
- Shamseer, L., Moher, D., Clarke, M., Gherzi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L.A., 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 349, g7647. <https://doi.org/10.1136/bmj.g7647>
- Shore, S., Zhou, J., Koehler, S., 2007. Neural mechanisms underlying somatic tinnitus. *Prog Brain Res* 166, 107–123. [https://doi.org/10.1016/S0079-6123\(07\)66010-5](https://doi.org/10.1016/S0079-6123(07)66010-5)
- Shore, S.E., Roberts, L.E., Langguth, B., 2016. Maladaptive plasticity in tinnitus--triggers, mechanisms and treatment. *Nat Rev Neurol* 12, 150–160.
<https://doi.org/10.1038/nrneurol.2016.12>
- Signorelli, F., Turjman, F. (Eds.), 2014. Tinnitus: Causes, Treatment and Short & Long-Term Health Effects (Acoustics Research and Technology), UK ed. edition. ed. Nova Science Pub Inc.
- Simoes, J., Neff, P., Schoisswohl, S., Bulla, J., Schecklmann, M., Harrison, S., Vesala, M., Langguth, B., Schlee, W., 2019. Toward Personalized Tinnitus Treatment: An Exploratory Study Based on Internet Crowdsensing. *Front Public Health* 7, 157.
<https://doi.org/10.3389/fpubh.2019.00157>
- Simonsohn, U., Nelson, L.D., Simmons, J.P., 2014. P-curve: a key to the file-drawer. *J Exp Psychol Gen* 143, 534–547. <https://doi.org/10.1037/a0033242>
- Simonsohn, U., Simmons, J.P., Nelson, L.D., 2015. Better P-curves: Making P-curve analysis more robust to errors, fraud, and ambitious P-hacking, a Reply to Ulrich and Miller (2015). *J Exp Psychol Gen* 144, 1146–1152.
<https://doi.org/10.1037/xge0000104>
- Skrondal, A., Laake, P., 2001. Regression among factor scores. *Psychometrika* 66, 563–575.
- Spiegel, D.P., Linford, T., Thompson, B., Petoe, M.A., Kobayashi, K., Stinear, C.M., Searchfield, G.D., 2015. Multisensory attention training for treatment of tinnitus. *Sci Rep* 5, 10802. <https://doi.org/10.1038/srep10802>
- Staplin, N., Herrington, W.G., Judge, P.K., Reith, C.A., Haynes, R., Landray, M.J., Baigent, C., Emberson, J., 2017. Use of Causal Diagrams to Inform the Design and Interpretation of Observational Studies: An Example from the Study of Heart and Renal Protection (SHARP). *Clin J Am Soc Nephrol* 12, 546–552.
<https://doi.org/10.2215/CJN.02430316>
- Stegenga, J., 2011. Is meta-analysis the platinum standard of evidence? *Stud Hist Philos Biol Biomed Sci* 42, 497–507. <https://doi.org/10.1016/j.shpsc.2011.07.003>
- Stern, Y., 2013. Cognitive reserve: implications for assessment and intervention. *Folia Phoniatr Logop* 65, 49–54. <https://doi.org/10.1159/000353443>
- Stern, Y., 2009. Cognitive reserve. *Neuropsychologia* 47, 2015–2028.
<https://doi.org/10.1016/j.neuropsychologia.2009.03.004>
- Sterne, J.A.C., Sutton, A.J., Ioannidis, J.P.A., Terrin, N., Jones, D.R., Lau, J., Carpenter, J., Rücker, G., Harbord, R.M., Schmid, C.H., Tetzlaff, J., Deeks, J.J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D.G., Moher, D., Higgins, J.P.T.,

2011. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 343, d4002. <https://doi.org/10.1136/bmj.d4002>
- Stevens, C., Walker, G., Boyer, M., Gallagher, M., 2007. Severe tinnitus and its effect on selective and divided attention. *Int J Audiol* 46, 208–216. <https://doi.org/10.1080/14992020601102329>
- Storm, B.C., Levy, B.J., 2012. A progress report on the inhibitory account of retrieval-induced forgetting. *Mem Cogn.* 40, 827–843. <https://doi.org/10.3758/s13421-012-0211-7>
- Stuss, D.T., 2011. Functions of the frontal lobes: relation to executive functions. *J Int Neuropsychol Soc* 17, 759–765. <https://doi.org/10.1017/S1355617711000695>
- Suliman, N.A., Mat Taib, C.N., Mohd Moklas, M.A., Adenan, M.I., Hidayat Baharudin, M.T., Basir, R., 2016. Establishing Natural Nootropics: Recent Molecular Enhancement Influenced by Natural Nootropic. *Evid Based Complement Altern. Med* 2016, 4391375. <https://doi.org/10.1155/2016/4391375>
- Sumowski, J.F., Leavitt, V.M., 2013. Cognitive reserve in multiple sclerosis. *Mult Scler* 19, 1122–1127. <https://doi.org/10.1177/1352458513498834>
- Tavanai, E., Mohammadkhani, G., 2018. A different view on the link between tinnitus and cognition; is there a reciprocal link? *Int J Neurosci* 1–17. <https://doi.org/10.1080/00207454.2018.1503185>
- Taylor, J.A., Thompson, D.M., Hall, D.A., Walker, D.-M., McMurran, M., Casey, A., Featherstone, D., MacDonald, C., Stockdale, D., Hoare, D.J., 2020. The TinMan study: feasibility trial of a psychologically informed, audiologist-delivered, manualised intervention for tinnitus. *Int J Audiol* 1–10.
- Tegg-Quinn, S., Bennett, R.J., Eikelboom, R.H., Baguley, D.M., 2016. The impact of tinnitus upon cognition in adults: A systematic review. *Int J Audiol* 55, 533–540. <https://doi.org/10.1080/14992027.2016.1185168>
- Teng, E.L., Hasegawa, K., Homma, A., Imai, Y., Larson, E., Graves, A., Sugimoto, K., Yamaguchi, T., Sasaki, H., Chiu, D., 1994. The Cognitive Abilities Screening Instrument (CASI): a practical test for cross-cultural epidemiological studies of dementia. *Int Psychogeriatr* 6, 45–58; discussion 62. <https://doi.org/10.1017/s1041610294001602>
- Tennant, P.W.G., Harrison, W.J., Murray, E.J., Arnold, K.F., Berrie, L., Fox, M.P., Gadd, S.C., Keeble, C., Ranker, L.R., Textor, J., Tomova, G.D., Gilthorpe, M.S., Ellison, G.T.H., 2019. Use of directed acyclic graphs (DAGs) in applied health research: review and recommendations. <https://doi.org/10.1101/2019.12.20.19015511>
- Tetteh, H., Lee, M., Geoffrey Lau, C., Yang, Sunggu, Yang, Sungchil, 2017. Tinnitus: Prospects for Pharmacological Interventions With a Seesaw Model. *Neuroscientist* 1073858417733415. <https://doi.org/10.1177/1073858417733415>
- Textor, J., van der Zander, B., Gilthorpe, M.S., Liskiewicz, M., Ellison, G.T., 2016. Robust causal inference using directed acyclic graphs: the R package “dagitty.” *Int J Epidemiol* 45, 1887–1894.
- Tonndorf, J., 1987. The analogy between tinnitus and pain: a suggestion for a physiological basis of chronic tinnitus. *Hear Res* 28, 271–275. [https://doi.org/10.1016/0378-5955\(87\)90054-2](https://doi.org/10.1016/0378-5955(87)90054-2)
- Treitz, F.H., Heyder, K., Daum, I., 2007. Differential course of executive control changes during normal aging. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn* 14, 370–393. <https://doi.org/10.1080/13825580600678442>
- Trevis, K.J., McLachlan, N.M., Wilson, S.J., 2017a. A systematic review and meta-analysis of psychological functioning in chronic tinnitus. *Clin Psychol Rev.* <https://doi.org/10.1016/j.cpr.2017.12.006>
- Trevis, K.J., McLachlan, N.M., Wilson, S.J., 2016. Cognitive Mechanisms in Chronic Tinnitus: Psychological Markers of a Failure to Switch Attention. *Front Psychol* 7, 1262. <https://doi.org/10.3389/fpsyg.2016.01262>
- Trevis, K.J., Tailby, C., Grayden, D.B., McLachlan, N.M., Jackson, G.D., Wilson, S.J., 2017b. Identification of a Neurocognitive Mechanism Underpinning Awareness of Chronic Tinnitus. *Sci Rep* 7, 15220. <https://doi.org/10.1038/s41598-017-15574-4>

- van den Berge, M.J.C., Free, R.H., Arnold, R., de Kleine, E., Hofman, R., van Dijk, J.M.C., van Dijk, P., 2017. Cluster Analysis to Identify Possible Subgroups in Tinnitus Patients. *Front Neurol* 8, 115. <https://doi.org/10.3389/fneur.2017.00115>
- Vanneste, S., Faber, M., Langguth, B., De Ridder, D., 2016. The neural correlates of cognitive dysfunction in phantom sounds. *Brain Res* 1642, 170–179. <https://doi.org/10.1016/j.brainres.2016.03.016>
- Verhaeghen, P., 2013. *The Elements of Cognitive Aging: Meta-Analyses of Age-Related Differences in Processing Speed and Their Consequences*. Oxford University Press.
- Viechtbauer, W., 2010. Conducting Meta-Analyses in R with the metafor Package. *J. Stat. Softw. Artic.* 36, 1–48. <https://doi.org/10.18637/jss.v036.i03>
- Viechtbauer, W., Cheung, M.W.-L., 2010. Outlier and influence diagnostics for meta-analysis. *Res Synth Methods* 1, 112–125. <https://doi.org/10.1002/jrsm.11>
- Waechter, S., Brännström, K.J., 2015. The impact of tinnitus on cognitive performance in normal-hearing individuals. *Int J Audiol* 54, 845–851. <https://doi.org/10.3109/14992027.2015.1055836>
- Wagenmakers, E.-J., van der Maas, H.L.J., Grasman, R.P.P.P., 2007. An EZ-diffusion model for response time and accuracy. *Psychon Bull Rev* 14, 3–22.
- Walsh, K.P., Pasanen, E.G., McFadden, D., 2015. Changes in otoacoustic emissions during selective auditory and visual attention. *J Acoust Soc Am* 137, 2737–2757. <https://doi.org/10.1121/1.4919350>
- Wattamwar, K., Qian, Z.J., Otter, J., Leskowitz, M.J., Caruana, F.F., Siedlecki, B., Spitzer, J.B., Lalwani, A.K., 2017. Increases in the Rate of Age-Related Hearing Loss in the Older Old. *JAMA Otolaryngol Head Neck Surg* 143, 41–45. <https://doi.org/10.1001/jamaoto.2016.2661>
- Wayne, R.V., Johnsrude, I.S., 2015. A review of causal mechanisms underlying the link between age-related hearing loss and cognitive decline. *Ageing Res Rev* 23, 154–166. <https://doi.org/10.1016/j.arr.2015.06.002>
- Webb, S.L., Loh, V., Lampit, A., Bateman, J.E., Birney, D.P., 2018. Meta-Analysis of the Effects of Computerized Cognitive Training on Executive Functions: a Cross-Disciplinary Taxonomy for Classifying Outcome Cognitive Factors. *Neuropsychol Rev* 28, 232–250. <https://doi.org/10.1007/s11065-018-9374-8>
- Weed, D.L., 2005. Weight of evidence: a review of concept and methods. *Risk Anal* 25, 1545–1557. <https://doi.org/10.1111/j.1539-6924.2005.00699.x>
- Wegner, I., Hall, D.A., Smit, A.L., McFerran, D., Stegeman, I., 2018. Betahistine for tinnitus. *Cochrane Database Syst Rev* 12, CD013093. <https://doi.org/10.1002/14651858.CD013093.pub2>
- Weschler, D., 2011. *Wechsler Abbreviated Scaled of Intelligence-Second Edition*. Pearson, San Antonio.
- Weschler, D., 2010. *Wechsler Adult Intelligence Scale -Fourth Edition*. Pearson, San Antonio.
- Wetzels, R., Matzke, D., Lee, M.D., Rouder, J.N., Iverson, G.J., Wagenmakers, E.-J., 2011. Statistical Evidence in Experimental Psychology: An Empirical Comparison Using 855 t Tests. *Perspect Psychol Sci* 6, 291–298. <https://doi.org/10.1177/1745691611406923>
- Whelan, R., 2008. Effective Analysis of Reaction Time Data. *Psychol Rec* 58, 475–482. <https://doi.org/10.1007/BF03395630>
- Whitmer, A.J., Banich, M.T., 2010. Trait rumination and inhibitory deficits in long-term memory. *Cogn. Emot.* 24, 168–179. <https://doi.org/10.1080/02699930802645762>
- Whitmer, A.J., Gotlib, I.H., 2013. An attentional scope model of rumination. *Psychol Bull* 139, 1036–1061. <https://doi.org/10.1037/a0030923>
- Wicherts, J.M., Veldkamp, C.L.S., Augusteijn, H.E.M., Bakker, M., van Aert, R.C.M., van Assen, M.A.L.M., 2016. Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking. *Front Psychol* 7, 1832. <https://doi.org/10.3389/fpsyg.2016.01832>

- Wickham, H., Averick, M., Bryan, J., Chang, W., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the Tidyverse. *J. Open Source Softw.* 4, 1686. <https://doi.org/10.21105/joss.01686>
- Wilhelm, O., Hildebrandt, A., Oberauer, K., 2013. What is working memory capacity, and how can we measure it? *Front Psychol* 4, 433. <https://doi.org/10.3389/fpsyg.2013.00433>
- Wilkinson, P.O., Croudace, T.J., Goodyer, I.M., 2013. Rumination, anxiety, depressive symptoms and subsequent depression in adolescents at risk for psychopathology: a longitudinal cohort study. *BMC Psychiatry* 13, 250.
- Wise, K., Kobayashi, K., Magnusson, J., Welch, D., Searchfield, G.D., 2016. Randomized Controlled Trial of a Perceptual Training Game for Tinnitus Therapy. *Games Health J* 5, 141–149. <https://doi.org/10.1089/g4h.2015.0068>
- Wright, B.C., 2017. What Stroop tasks can tell us about selective attention from childhood to adulthood. *Br J Psychol* 108, 583–607. <https://doi.org/10.1111/bjop.12230>
- Yang, Y., Cao, S., Shields, G.S., Teng, Z., Liu, Y., 2017. The relationships between rumination and core executive functions: A meta-analysis. *Depress Anxiety* 34, 37–50.
- Yntema, D.B., 1963. Keeping track of several things at once. *Hum Factors* 5, 7–17. <https://doi.org/10.1177/001872086300500102>
- Zeng, F.-G., Richardson, M., Turner, K., 2020. Tinnitus Does Not Interfere with Auditory and Speech Perception. *J Neurosci* 40, 6007–6017.

Appendices

Appendix A – WoE tool

Evidence was graded according to broad domains of Methodological Quality (WoE A), Methodological Relevance (WoE B), and Topic Relevance (WoE C). Considered domains are given low, medium, or high ratings to provide a final total score of study bias and quality (WoE D).

	High	Medium	Low
WoE A: Methodological Quality • Not review specific • Generic evaluation of study quality • Judgement of coherence and integrity of evidence within study	Explicit and detailed Methods and Results section for data collection and analysis. Well-designed and described. Clear interpretation of findings. Adequate statistical power, estimated with a priori power calculation.	Satisfactory Methods and Results sections. Enough details to determine what was done. Some design and/or potential interpretation issues.	Poorly designed or described. Poorly controlled or inadequate statistical power
WoE B: Methodological Relevance • Review specific • Judgement about the appropriateness of study design/method to answer the review question	No risk of sample selection bias as relevant confounds have been factored into participant selection or within data analysis (i.e. hearing impairment, depression, anxiety, medication usage, previous head injury, or visual acuity of participants if using visual cognitive measures).	Slight risk of sample selection bias but most confounds (> 50%) have been factored into participant selection or within data analysis.	High risk of sample selection bias as most confounds (< 50%) have not been factored into participant selection or within data analysis.
WoE C: Topic Relevance • Review specific • Judgement about the relevance of study focus to answer the review question	Little risk of effect detection bias as cognitive measures used in the study are established paradigms, standardised, or commercially available material. An established multi-item questionnaire has been used when measuring tinnitus severity.	Moderate risk of effect detection bias as established cognitive measures have been modified for the study. An established multi-item questionnaire has been used when measuring tinnitus severity.	High risk of effect detection bias as completely bespoke cognitive measures have been used. An established multi-item questionnaire has not been used when measuring tinnitus severity.

Appendix B – Retrieval-Practise Paradigm Category-Exemplars

Category	Category-Exemplar Pair	Final Test Prompt	Practise Category
drinks	drinks-vodka	drinks-v	Rp-
drinks	drinks-bourbon	drinks-b	Rp-
drinks	drinks-rum	drinks-r	Rp-
drinks	drinks-ale	drinks-a	Rp+
drinks	drinks-gin	drinks-g	Rp+
drinks	drinks-whiskey	drinks-w	Rp+
weapons	weapons-sword	weapons-s	Rp-
weapons	weapons-bomb	weapons-b	Rp-
weapons	weapons-rifle	weapons-r	Rp-
weapons	weapons-pistol	weapons-p	Rp+
weapons	weapons-tank	weapons-t	Rp+
weapons	weapons-club	weapons-c	Rp+
fish	fish-catfish	fish-c	Rp-
fish	fish-bluegill	fish-b	Rp-
fish	fish-trout	fish-t	Rp-
fish	fish-flounder	fish-f	Rp+
fish	fish-herring	fish-h	Rp+
fish	fish-guppy	fish-g	Rp+
fruit	fruit-tomato	fruit-t	Rp-
fruit	fruit-orange	fruit-o	Rp-
fruit	fruit-strawberry	fruit-s	Rp-
fruit	fruit-lemon	fruit-l	Rp+
fruit	fruit-banana	fruit-b	Rp+
fruit	fruit-pineapple	fruit-p	Rp+
professions	professions-engineer	professions-e	Np
professions	professions-nurse	professions-n	Np
professions	professions-accountant	professions-a	Np
professions	professions-plumber	professions-p	Np
professions	professions-farmer	professions-f	Np
professions	professions-dentist	professions-d	Np
metals	metals-iron	metals-i	Np
metals	metals-silver	metals-s	Np
metals	metals-aluminium	metals-a	Np
metals	metals-brass	metals-b	Np
metals	metals-nickel	metals-n	Np
metals	metals-gold	metals-g	Np
trees	trees-birch	trees-b	Np
trees	trees-hickory	trees-h	Np
trees	trees-elm	trees-e	Np
trees	trees-spruce	trees-s	Np
trees	trees-dogwood	trees-d	Np
trees	trees-redwood	trees-r	Np
insects	insects-beetle	insects-b	Np

insects	insects-fly	insects-f	Np
insects	insects-roach	insects-r	Np
insects	insects-mosquito	insects-m	Np
insects	insects-hornet	insects-h	Np
insects	insects-grasshopper	insects-g	Np