Quantitative appraisal of trial methods in reviews: an investigation of antipsychotic drug effects in dementia

Hulshof, Tessa

DOI:
10.33612/diss.193114503

Quantitative appraisal of trial methods in reviews: an investigation of antipsychotic drug effects in dementia

Some studies described in this thesis were financially supported by the College ter Beoordeling van Geneesmiddelen – Medicines Evaluation Board (CBG-MEB).

CBG MEB | COLLEGE TER BEOORDELING VAN GENEESMIDDELEN

Publication of this thesis was financially supported by:

rijksuniversiteit groningen

Research Institute SHARE

Cover photo	Henk van den Tillaart
Cover lay-out	Loes Kema
Printed by	proefschriften.nl

rijksuniversiteit
groningen

Quantitative appraisal of trial methods in reviews: an investigation of antipsychotic drug effects in dementia

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. C. Wijmenga
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

woensdag 15 december 2021 om 16.15 uur

door

Teresa Anneleen Hulshof

geboren op 19 december 1988
te Zevenaar

Promotor
Prof. dr. S.U. Zuidema

Copromotor
Dr. H.J. Luijendijk

Beoordelingscommissie
Prof. dr. W.P. Achterberg
Prof. dr. P. Denig
Prof. dr. F. Verhey

Contents

Chapter 1.

General introduction and outline of the thesis

Introduction

Around 50 million people worldwide have dementia, and it is estimated that in 2050 this number has tripled.[1,2] Dementia is a syndrome of a combination of symptoms associated with cognitive decline. Decline in memory, thinking, problem-solving, and language make daily activities more difficult for the patients.

There are many types of dementia, with Alzheimer dementia as the most common type. It contributes to two-thirds of dementia cases.[3,4] Alzheimer dementia impacts the part of the brain that is associated with learning. Therefore, patients start having trouble remembering new information.[5] Another well-known type of dementia is vascular dementia. This type causes a decline in skills and memory due to reduced blood flow to the brain, which leads to deprivation of oxygen and nutrients in the brain.[6] Other types of dementia include Lewy body dementia with abnormal deposits of protein in the brain cells, and frontotemporal dementia characterized by nerve cell loss in the frontal and temporal lobes of the brain.[6]

Neuropsychiatric symptoms and antipsychotics

When dementia progresses, 60% to 98% of patients develop neuropsychiatric symptoms (NPS) such as agitation or psychosis, especially in later stages of the disease.[7–9] Other common NPS are depression, apathy, aggression, delusions, hallucinations, and sleep disturbances.[7–9] These symptoms are not only distressing for patients but also for their families and caregivers.[7,8] The prevalence of NPS depends on the severity and the type of dementia.[10]

Some patients may have had another psychiatric disorder before being diagnosed with dementia, and some psychiatric disorders increase the risk of developing dementia later in life.[11,12] The symptoms of these psychiatric disorders should not be confused with NPS caused by dementia. Also, NPS such as agitation and psychosis can occur as a result of underlying causes such as pain or delirium.[13]

To treat agitation and psychosis associated with dementia itself, antipsychotics are widely prescribed.[14] Even though guidelines recommend nonpharmacological interventions as the first treatment option for NPS.[8,15] Before considering the use of antipsychotics, a thorough problem analysis should be undertaken and other treatments should have had an insufficient effect.[15]

First-, and second-generation antipsychotics

There are two types of antipsychotics to choose from: first-generation or conventional antipsychotics, and second-generation or atypical antipsychotics.[16] First-generation

antipsychotics were developed in the 1950s and are primarily used to treat psychosis (in particular in schizophrenia).[17] They have also been effective in the treatment of other related psychotic disorders.[17,18] Chlorpromazine was the first developed antipsychotic. Other frequently prescribed antipsychotics are perphenazine and haloperidol.[18] These first-generation antipsychotics are associated with a high risk of side effects, such as extrapyramidal symptoms.[17]

Second-generation antipsychotics were marketed in the 1980s and gained popularity due to an apparent lower risk of neurological side effects.[17] However, later research showed that second-generation antipsychotics are associated with an increased risk of developing metabolic side effects and mortality.[17,19] In general, aripiprazole, olanzapine, risperidone, and quetiapine are the most frequently prescribed atypical antipsychotics.[18]

Randomized controlled trials

The efficacy of antipsychotics for dementia has been investigated in randomized controlled trials (RCTs). In RCTs, a new antipsychotic drug is usually compared to a placebo to see if it is safe and effective. In some cases, researchers perform a head-to-head trial when they want to compare a new antipsychotic drug to an existing antipsychotic and investigate if the new antipsychotic performs better. Placebo-controlled trials can be conducted with fewer patients than trials that compare antipsychotic drugs to each other.

A few steps should be performed with extra attention, starting with the determination of a research question. This may seem like a small step, but it is important that the research question is clear and focused.[20,21] For antipsychotic trials in patients with dementia this means that the target symptom(s) and the interventions should be defined in detail, and outcomes should be clinically relevant.[20]

Besides defining the exact target population, it is also necessary to include a sufficient number of participants, i.e. a sample size that is large enough, to identify a clinically relevant treatment effect. Larger sample sizes provide a more precisely estimated effect, i.e. smaller confidence intervals,[22] but are also more time consuming and cost more money.

Trial methods

To yield valid results, RCTs should be conducted properly. An important step is the randomization procedure. In RCTs, treatment groups need to be comparable in order to yield an unbiased estimation of the treatment effect. To ensure that chance instead of patient characteristics determine treatment assignment, randomization

is performed.[23] The randomization procedure consists of two aspects, the random sequence generation and the concealment of allocation, that make the treatment a participants gets unpredictable and uncontrolable.[23] Otherwise, if the persons enrolling participants know or are able to find out what the next allocation is, they might decide not to enroll a participant, postpone the enrollment, or change the allocation to the preferred treatment.

Currently, the preferred method to generate the random allocation sequence is a computerized random number generator. Restrictions on the random sequence generation that decrease unpredictability, such as use of blocked randomization with a small set block size, should be clearly stated by the authors of the trial.[23] Concealment of allocation can be done with opaque envelopes, and sequences only known by pharmacists or statisticians not involved in randomization for example. If both randomized sequence generation and allocation concealment are executed successfully, baseline differences between the control group and the intervention group are minimized in particular when the sample size is large.

Blinding is a procedure that prevents people involved in the trial, such as study participants, caregivers, or outcome assessors, from knowing which intervention the participant received.[23,24] Participants of a trial may report more favorable results on a subjective outcome, e.g. a symptom questionnaire, if they are aware they are receiving the active or new treatment. The same applies to unblinded caregivers, health care workers and research staff assessing outcomes.[23] The investigators should therefore explicitly report the blinding status of all persons involved in outcome assessment, especially if these outcomes are more or less susceptible to measurement error.[23]

When designing a trial, recruiting a representative sample is also essential.[25] RCTs may yield accurate estimates for the participants in the trial, but if they do not represent the defined study population well the trial might not yield valid information for that target population.[25] For instance, if patients who belong to the target population but have an increased risk of side effects are not enrolled in the trial, the trial will not yield a risk of side effects that is valid for the entire target population. This is commonly called lack of generalizability.

Reviews of trials

Multiple individual RCTs can be included in a systematic review and their results pooled in a meta-analysis for the highest level of evidence.[26,27] Combining several trials increases the chance of finding an effect (more statistical power), and the estimated intervention effect will be more precise.[28] In addition, it might also settle controversies from conflicting results of studies.[28]

Performing a systematic review or meta-analysis starts with a thorough literature search. The results of this search are used to select the appropriate studies. These studies are then assessed with a risk of bias tool. Researchers use these tools to assess the design and performance of individual elements of an RCT. It is recommended that two reviewers assess risk of bias independently, discuss discrepancies and reach a consensus about the final assessment.[29] Completed assessments are used to determine the overall risk of bias of the individual study, considering all relevant sources of bias.[29]

However, a number of problems in these standardized methods for reviews need to be acknowledged. These errors can be divided into two main problems. The first problem is the clinical relevance of the enrolled population, the measured outcomes, and the reported effect size. Up to now, this has not received the attention it deserves in reviews of antipsychotics in dementia.

The second problem is the subjective assessment of risk of bias despite the use of risk of bias tools and two independent assessors. The adequate application of commonly used risk of bias tools depends on the level of methodological knowledge of the reviewers. Some are known for high interrater variability, such as the Cochrane Collaboration risk of bias tool,[30] and others are known for the limited number of items, such as the Jadad scale.[31] The use of multiple terms for the same type of bias makes it difficult to rate the risk of bias properly too.[32]

Research aims

In this thesis, I critically appraised the methods and results of antipsychotic drug trials for patients with dementia. First, the clinical relevance of treatment effects of antipsychotics in dementia was assessed. Next, the influence of bias on the estimated treatment effects was investigated.[33] The aims of the different studies included in this thesis are explained below.

Part 1: Assessing clinical relevance of effects

The work for this thesis started with a standard systematic review. At the time, a review of RCTs about the risk of mortality of conventional antipsychotics in dementia was

not yet available. Only meta-analyses of cohort studies had been published. These cohort studies could be biased because none of the studies took severity of illness into account,[34,35] even though patients with terminal illness often receive haloperidol to treat symptoms of delirium.[36,37] Therefore, my colleagues and I, performed a meta-analysis of trials to investigate whether conventional antipsychotics have an effect on the risk of mortality in elderly patients with dementia.

Next, we took a closer look at the patient population in terms of target symptom, intervention of interest, comparison intervention, and outcomes (PICO). In trials on the efficacy of antipsychotics for dementia, these have not always been defined clearly or applied consistently. For instance, some trials investigating the effect on agitation included only agitated patients, but others included patients with neuropsychiatric symptoms in general. If the effects from trials with different patient populations are pooled, the results may not be meaningful. In addition, even if all enrolled participants had agitation, some trials used agitation specific outcome scales, but others generic outcomes scales. The outcome scale did not always to fit the target symptom either. The same issues arose in trials that investigated the effect of antipsychotics on psychosis in dementia. Therefore, we investigated whether the definition of the study populations and the choice of outcome scale affected the pooled efficacy of antipsychotic trials in dementia.

Older antipsychotic studies are often small. Studies based on small sample sizes may yield non-statistically significant results, but these results can still be clinically relevant. On the other hand, newer studies based on large sample sizes may yield statistically significant but clinically irrelevant treatment effects.[38–40] Large sample size fallacy occurs when such statistically significant results from large trials are interpreted as relevant for medical practice.[41] Sample sizes of antipsychotic trials in dementia have increased over the years.[19,42] We assessed the variation of the sample sizes, the size of the reported treatment effects and the association between study characteristics and sample size in these trials.

Part 2: Quantifying bias in trials

Systematic errors in research methodology may result in biased treatment effect, i.e. the true effect of the treatment on the outcome of interest is under- or overestimated.[33] In part two of this thesis bias in antipsychotic trials in dementia was studied. There are three types of bias commonly acknowledged in observational studies: selection bias, confounding and information bias. This categorization can be applied to trials as well.[32]

In a trial, patient withdrawal i.e. deselection from a study can lead to incomplete outcome data. This is especially problematic when the withdrawal differs between treatment groups and these participants are omitted from reports or analyses.[43] To avoid this commonly acknowledged source of selection bias, the investigators should include information on all participants who underwent randomization in the groups to which they were originally allocated (intention-to-treat analysis).[23]

Drop-out of eligible patient related to the investigated treatment can also occur before randomization during a so-called run-in period. To decrease drop-out and placebo response, and hence increase effect size and power, some RCTs include a run-in period.[44,45] A run-in phase is performed between screening for eligibility and randomization, it takes a few days to months, and patients can receive a wash-out of already used drugs, the investigated active drug, or placebo. However, efficiently executed run-in periods may lead to exclusion (deselection) of certain patients from randomization and thus limit the validity of trial results for the defined study population.[45,46] This is a sources of selection bias that has not received much attention in the medical literature so far. We studied the effects of antipsychotics in trials with and without run-in.

Confounding occurs when a factor that affects treatment and outcome is not adjusted for. This form of bias can be avoided with randomization as described above. Flawed randomization procedures can give rise to baseline imbalances. When my colleagues and I conducted the meta-analysis about mortality risk of conventional antipsychotics, it seemed that baseline imbalances were present in the individual trials. Therefore, we assessed the presence of baseline imbalances and investigated if these imbalances affected study results in antipsychotic trials for dementia.

Finally, information bias can occur when knowledge of treatment status affects the measurement of the outcome. For example, when an active drug has specific side effects, treatment status can sometimes be guessed. Trials are susceptible to this type of bias when subjective rating scales are used. Objective outcomes, such as the use of rescue medication, may provide more valid results. We assessed the effectiveness and side effects of antipsychotics in neuropsychiatric patients with dementia, using subjective and objective measures of these outcomes.

Thesis outline

This thesis presents six studies. First, part 1 presents three studies about the clinical relevance of the reported effects. **Chapter 2** provides an assessment of the mortality risk of conventional antipsychotics in elderly patients with dementia in randomized controlled trials. In **chapter 3** the pooled efficacy of antipsychotics

measured in patients with dementia and agitation or psychosis was assessed. We investigated whether a broad definition of population and outcome yielded different effects compared to a target-specific definition. **Chapter 4** focuses on large sample size fallacy. The variation in sample size, size of treatment effect and general study characteristics related to sample size were investigated.

Part 2 then reports three studies related to quantification of risk of bias. **Chapter 5** presents a study about the association of run-in periods with reported treatment effects in trials of antipsychotics in dementia. In **chapter 6** the presence of baseline imbalances was assessed as well as the association of baseline imbalances with reduction of neuropsychiatric symptoms, the risk of extrapyramidal symptoms and risk of mortality. **Chapter 7** describes the effectiveness and side effects of antipsychotics in terms of subjective and objective outcome measures.

Finally, **chapter 8** provides a general discussion, and a scientific summary can be found in the appendices.

References

1. World Health Organization (WHO). Dementia: Key facts. https://www.who.int/news-room/fact-sheets/detail/dementia. Published 2019. Accessed May 25, 2020.
2. Alzheimer's Research UK. Global prevalence. https://www.dementiastatistics.org/statistics/global-prevalence/. Published 2018. Accessed May 25, 2020.
3. Alzheimer's Research UK. Different types of dementia. https://www.dementiastatistics.org/statistics/different-types-of-dementia/. Published 2018. Accessed May 25, 2020.
4. World Health Organization (WHO). Dementia: Key facts.
5. Alzheimer's association. Types of dementia. https://www.alz.org/alzheimers-dementia/what-is-dementia/types-of-dementia. Published 2020. Accessed June 22, 2020.
6. Alzheimer's association. Types of dementia.
7. Sink K, Holden K, Yaffe K. Pharmacological treatment of neuropsychiatric symptoms of dementia: A review of the evidence. J Am Med Dir Assoc. 2006;7(3):201-202. doi:10.1016/j.jamda.2005.12.024
8. Phan S V., Osae S, Morgan JC, Inyang M, Fagan SC. Neuropsychiatric Symptoms in Dementia: Considerations for Pharmacotherapy in the USA. Drugs R D. 2019;19(2):93-115. doi:10.1007/s40268-019-0272-1
9. Lyketsos C, Lopez O, Jones B, Fitzpatrick A, Breitner J, Dekosky S. Prevalence of Neuropsychiatric Symptoms in Dementia and Mild Cognitive Impairment: Results From the Cardiovascular Health Study. JAMA. 2002;288(12):1475-1483.
10. Bergh S, Selbæk G. The prevalence and the course of neuropsychiatric symptoms in patients with dementia. Nor Epidemiol. 2012;22(2):225-232. doi:10.5324/nje.v22i2.1570
11. Li G, Wang L, Shofer J, et al. Temporal Relationship Between Depression and Dementia. Arch Gen Psychiatry. 2011;68(9):970. doi:10.1001/archgenpsychiatry.2011.86
12. Zilkens R, Bruce D, Duke J, Spilsbury K, Semmens J. Severe Psychiatric Disorders in Mid-Life and Risk of Dementia in Late- Life (Age 65-84 Years): A Population Based Case-Control Study. Curr Alzheimer Res. 2014;11(7):681-693. doi:10.2174/1567205011666140812115004
13. Mühlbauer V, Luijendijk H, Dichter M, Möhler R, Zuidema S, Köpke S. Antipsychotics for Agitation and Psychosis in People with Alzheimer's Disease and Vascular Dementia (Protocol).; 2019. doi:10.1002/14651858.CD013304. www.cochranelibrary.com
14. Zuidema SU, Johansson A, Selbaek G, et al. A consensus guideline for antipsychotic drug use for dementia in care homes. Bridging the gap between scientific evidence and clinical practice. Int Psychogeriatrics. 2015;27(11):1849-1859. doi:10.1017/S1041610215000745

15. Zuidema S, Smalbrugge M, Bil W, et al. Multidisciplinary Guideline problem behaviour in dementia [dutch]. Verenso, NIP. https://www.verenso.nl/richtlijnen-en-praktijkvoering/richtlijnendatabase/probleemgedrag-bij-mensen-met-dementie. Published 2018.
16. Eisenberg J. First-Generation Versus Second-Generation Antipsychotics in Adults: Comparative Effectiveness. In: Comparative Effectiveness Review Summary Guides for Clinicians [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013. https://www.ncbi.nlm.nih.gov/books/NBK142851/.
17. Guzman F. First vs Second-Generation Antipsychotics. Psychopharmacology Institute. https://psychopharmacologyinstitute.com/publication/first-vs-second-generation-antipsychotics-2082. Published 2019. Accessed July 27, 2020.
18. Abou-Setta A, Mousavi S, Spooner C, et al. First-Generation Versus Second-Generation Antipsychotics in Adults: Comparative Effectiveness. Edmonton, Alberta Canada; 2007. http://www.ncbi.nlm.nih.gov/pubmed/23741760.
19. Schneider L, Dagerman K, Insel P. Risk of Death With Atypical Antipsychotic Drug Treatment for Dementia. J Am Med Assoc. 2005;294(15):1934-1943.
20. McKenzie J, Brennan S, Ryan R, Thomson H, Johnston R, Thomas J. Chapter 3: Defining the criteria for including studies and how they will be grouped for the synthesis. In: Higgins J, Thomas J, Chandler J, et al., eds. Cochrane Handbook for Systematic Reviews of Interventions. Version 6. Cochrane; 2019. https://training.cochrane.org/handbook/current/chapter-03.
21. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol. 2011;64(4):395-400. doi:10.1016/j.jclinepi.2010.09.012
22. Schünemann H, Vist G, Higgins J, et al. Chapter 15: Interpreting results and drawing conclusions. In: Higgins J, Thomas J, Chandler J, et al., eds. Cochrane Handbook for Systematic Reviews of Interventions. Version 6. Cochrane; 2019. https://training.cochrane.org/handbook/current/chapter-15.
23. Schulz K, Altman D, Moher D. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol. 2010;63(8):1-6. doi:10.1016/j.ijsu.2010.09.006
24. Wood L, Egger M, Gluud L, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ. 2008;336(7644):601-605. doi:10.1136/bmj.39465.451748.AD
25. Stuart E, Bradshaw C, Leaf P. Assessing the Generalizability of Randomized Trial Results to Target Populations. Prev Sci. 2015;16(3):475-485.
26. Guyatt G, Sackett D, Sinclair J, et al. Users' Guides to the Medical

Literature IX. A Method for Grading Health Care Recommendations. JAMA. 1995;274(22):1800-1804. doi:10.1001/jama.1995.03530220066035

27. Greenhalgh T. How to Read a Paper: Getting your bearings (deciding what the paper is about). BMJ. 1997;315(7102):243-246. doi:10.1136/bmj.315.7102.243
28. Deeks J, Higgins J, Altman D. Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins J, Thomas J, Chandler J, et al., eds. Chochrane Handbook for Systematic Reviews of Interventions. Version 6. Cochrane; 2019. https://training.cochrane.org/handbook/current/chapter-10.
29. NHMRC. Guidelines for Guidelines: Assessing risk of bias. https://www.nhmrc.gov.au/guidelinesforguidelines/develop/assessing-risk-bias. Published 2019. Accessed July 30, 2020.
30. The Cochrane Collaboration. The Cochrane Collaboration's tool for assessing risk of bias. 2005:1-4.
31. Jadad A, Moore A, Carroll D, et al. Assessing the quaility of reports of randomized clinical trials: Is blinding necessary? Control Clin Trials. 1996;17:1-12.
32. Luijendijk HJ, Page MJ, Burger H, Koolman X. Assessing risk of bias: a proposal for a unified framework for observational studies and randomized trials. BMC Med Res Methodol. 2020;20(1):237. doi:10.1186/s12874-020-01115-7
33. Barratt H, Kirwan M, Shantikumar S. Biases and Confounding. Public Health Action Support Team CIC. https://www.healthknowledge.org.uk/public-health-textbook/research-methods/1a-epidemiology/biases. Published 2018. Accessed July 30, 2020.
34. Luijendijk H, De Bruin N, Koolman X. Haloperidol in Elderly Users: Not A Cause of Death, but Caused by Impeding Death? Am J Epidemiol. 2013;177(11 Suppl):S1-S181.
35. Luijendijk H, De Bruin N, Hulshof T, Koolman A. Terminal illness and the increased mortality risk of conventional antipsychotics in observational studies: a systematic review. Pharmacoepidemiol Drug Saf. 2016;25(2):113-122.
36. Maltoni M, Scarpi E, Rosati M, et al. Palliative sedation in end-of-life care and survival: a systematic review. J Clin Oncol. 2012;30(12):1378-1383. doi:10.1200/JCO.2011.37.3795
37. Mercadante S, Fulfaro F, Casuccio A. Pattern of drug use by advanced cancer patients followed at home. J Palliat Care. 2001;17(1):37-40.
38. Chan A, Hrobjartsson A, Jorgensen K, Gotzsche P, Altman D. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. BMJ. 2008;337:1-8. doi:10.1136/bmj.a2299
39. Bhardwaj SS, Camacho F, Derrow A, Fleischer AB, Feldman SR. Statistical Significance and Clinical Relevance. 2014;140.
40. Page P. Beyond statistical significance: clinical interpretation of rehabilitation research literature. Int J Sports Phys Ther. 2014;9(5):726-736.

41. Lantz B. The large sample size fallacy. Scandanavian J Caring Sci. 2013;27(1):487-492.
42. Cox Gad S. Trial design. In: Clinical Trials Handbook. Hoboken, New Jersey: John Wiley & Sons; 2009:284.
43. Higgins J, Savovic J, Page M, Elbers R, Sterne J. Chapter 8: Assessing risk of bias in a randomized trial. In: Higgins J, Thomas J, Chandler J, et al., eds. Cochrane Handbook for Systematic Reviews of Interventions. Version 6. Cochrane; 2019. https://training.cochrane.org/handbook/current/chapter-08.
44. Ulmer M, Robinaugh D, Friedberg J, Lipsitz S, Natarajan S. Usefulness of a run-in period to reduce drop-outs in a randomized controlled trial of a behavioral intervention. Contemp Clin Trials. 2008;29(5):705-710.
45. Laursen DRT, Paludan-Müller AS, Hróbjartsson A. Randomized clinical trials with run-in periodsfrequency, characteristics and reporting. Clin Epidemiol. 2019;11:169-184. doi:10.2147/CLEP.S188752
46. Agid O, Siu C, Potkin S, Kapur S, Watsky E, Vanderburg D. Meta-regression analysis of placebo response in antipsychotic trials 1970-2010. Am J Psychiatry. 2013;170(11):1335-1344.

Part I.

Assessment of clinically relevant effects

Chapter 2.

The mortality risk of conventional antipsychotics in elderly patients: a systematic review and meta-analysis of randomized placebo-controlled trials

J Am Med Dir Assoc. 2015 Oct 1;16(10):817-24

Hulshof, T.A.
Zuidema, S.U.
Ostelo, R.W.J.G.
Luijendijk, H.J.

Abstract

Background

Numerous observational studies have reported an increased risk of mortality for conventional antipsychotics in elderly patients, and for haloperidol in particular. Subsequently, health authorities have warned against use of conventional antipsychotics in dementia. Experimental evidence is lacking.

Objective

To assess the mortality risk of conventional antipsychotics in elderly patients with a meta-analysis of trials.

Methods

Original studies were identified in electronic databases, online trial registers, and hand searched references of published reviews. Two investigators found 28 potentially eligible studies, and they selected 17 randomized placebo-controlled trials in elderly patients with dementia, delirium, or a high risk of delirium. Two investigators independently abstracted trial characteristics and deaths, and 3 investigators assessed the risk of bias. Deaths were pooled with RevMan to obtain risk differences and risk ratios.

Results

Data of 17 trials with a total of 2387 participants were available. Thirty-two deaths occurred. The pooled risk difference of 0.1% was not statistically significant (95% confidence interval (CI) -1.0%-1.2%). The risk ratio was 1.07 (95% CI 0.54-2.13). Eleven of 17 trials tested haloperidol (n= 1799). The risk difference was 0.4% (95% CI 0.9%-1.6%), the risk ratio was 1.25 (95% CI 0.59-2.65).

Conclusions

This meta-analysis of placebo-controlled randomized trials does not show that conventional antipsychotics in general or haloperidol in particular increase the risk of mortality in elderly patients. It questions the observational findings and the warning based on these findings.

Introduction

Haloperidol and other conventional antipsychotics are commonly used to reduce hallucinations, delusions, and aggression in elderly patients with dementia or delirium despite their well-known extrapyramidal and cardiac side effects.[1] However, in 2005, a meta-analysis of randomized trials suggested that use of haloperidol in patients with dementia increased the risk of mortality compared with placebo (odds ratio 1.68; 95% confidence interval [CI] 0.72-3.92).[2]

Multiple large cohort studies have since confirmed that conventional antipsychotics are associated with a higher risk of mortality than atypical antipsychotics and no use.[3] The association was present in general elderly populations, residents of nursing homes, and in patients with and patients without dementia. In several studies, haloperidol in particular increased the risk of mortality.[4,5] In 2008, the US Food and Drug Administration and the UK Commission for Drug Safety warned against use of conventional antipsychotics in elderly patients with dementia.[6,7] Health care professionals were advised to consider other, nonpharmacological, management options. The cohort studies that reported the mortality risk of conventional antipsychotics used extensive administrative databases incorporating sociodemographic data, medical diagnoses, and filed prescriptions.

Some studies applied advanced statistical techniques to adjust for confounders. Nevertheless, results of observational studies may be biased, even if the studies are of high quality. One source of bias might be that none of the studies took severity of illness into account.[8] This is a potentially strong confounder because haloperidol and chlorpromazine are often used to treat the symptoms of delirium in terminally ill patients.[9,10] These 2 drugs accounted for more than half of the conventional antipsychotics used.[8]

Evidence from experimental data is scarce. Two meta-analyses of randomized controlled trials (RCTs) have reported the risk of various adverse effects of conventional antipsychotics in patients with dementia, but not the risk of mortality.[1,11] Moreover, the risk of mortality presented in the 2005 meta-analysis was based on an unplanned subanalysis of 2 trials.[2] More trials that tested haloperidol and other conventional antipsychotics in patients with dementia have been published.[1,11] Information from trials in delirium may be valuable as well. Delirium, like dementia, is characterized by cognitive impairment and is indicative of frailty in an elderly patient. Many patients with delirium have a history of premorbid cognitive disorders or dementia, and patients with behavioral or psychological symptoms in dementia may have delirium.[12,13] Also, the use of haloperidol to prevent delirium in frail elderly patients has been advocated in recent years, and tested in trials.[14]

In general, the study period of trials is too short and the number of participants too small to detect infrequent adverse events such as deaths. However, the observational studies have suggested that deaths due to conventional antipsychotic use are rather common during the first 180 days of use (4.2%-7.3% of users),[3] and the relative risk of dying is highest in the first month when compared with atypical antipsychotics.[8] Trials to test antipsychotics for neuropsychiatric symptoms of dementia usually last 3 months or longer. The aim of this study was to perform a systematic review and meta-analysis of randomized trials to establish the mortality risk of conventional antipsychotics compared with placebo in elderly patients with dementia or delirium. We investigated (1) the conventional antipsychotics that were available in the study periods of the cohort studies (1994-2010), and (2) haloperidol, because this drug is the most popular conventional antipsychotic for psychosis and aggression related to dementia and delirium.

Methods

We set out to perform a systematic review and meta-analysis of RCTs using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) method to rate the quality of the evidence.[15]

Setting and Participants

We included RCTs that tested the efficacy of a conventional antipsychotic compared with placebo in participants aged 65 years or older who had diagnosed dementia, or delirium, or were frail and at risk of delirium. We excluded RCTs among patients with schizophrenia, advanced cancer, or terminal illness, and studies with multiple drugs in an intervention arm.

Intervention

The following drugs were considered to be conventional antipsychotics[16]: chlorpromazine, chlorprothixene, droperidol, flupentixol, fluphenazine, haloperidol, levomepromazine, loxapine, mesoridazine, molindone, pericyazine, perphenazine, pimozide, prochlorperazine, thioridazine, thiothixene, trifluoperazine, and zuclopenthixol.

Outcome Measure

Primary outcome measure was the number of participants who died between the start and the end of the study. Deaths of participants after the end of the study were excluded from the analyses.

Search Strategy

Two investigators performed the literature search and selected the studies (TAH, HJL). Three sources were used to identify studies: (1) electronic databases, (2) references of published systematic reviews and meta-analyses, and (3) trial registration Web sites. The electronic databases covered PubMed, CINAHL, and Embase. To search RCTs, the search strings ['generic name conventional antipsychotic' AND trial] and [dementia OR delirium] were used to find studies. Second, published systematic reviews and meta-analyses also were identified with PubMed, CINAHL, and Embase databases. The references in these systematic reviews were hand-searched. Title and abstract of potentially eligible studies were retrieved in PubMed. Third, RCTs were searched in the trial registries clinicaltrials.gov and controlled-trials.com for all the conventional antipsychotics mentioned previously. There were no restrictions with respect to publication date, language, or duration of the study. If studies seemed potentially eligible given title and abstract, full articles of published studies and online protocols of unpublished studies were retrieved. These articles and protocols were reviewed for definitive eligibility.

Data Extraction

Two reviewers (TAH, HJL) independently extracted the following data from the included studies: setting, type of patients, treatment groups, number of randomized patients in each treatment group, mean dose and range of administered haloperidol, study period, dropouts per group, and number of deaths per group. When mortality rates or other data were not reported, the corresponding author was contacted by e-mail and asked to provide the missing information. Only data from the first part of crossover studies (before actual crossover) was included.

GRADE and Risk of Bias

We followed the GRADE recommendation to rate the quality of overall evidence according to 5 items: risk of bias, inconsistency, indirectness, imprecision, and other considerations.[15] The items are to be graded as either "serious risk" or "no serious risk."

Three reviewers (TAH, SUZ, HJL) independently assessed the risk of bias in the RCTs with the Cochrane Collaboration risk of bias assessment tool.[17] This tool covers 6 items: (1) random sequence generation; (2) allocation concealment; (3) blinding of participants, health care personnel, and outcome assessors; (4) incomplete outcome data addressed adequately; (5) absence of selective reporting; and (6) absence of other potential sources of bias such as commercial funding. The reviewers also assessed (7) absence of baseline differences between treatment groups, because lack of baseline differences is the goal of randomization (items 1 and 2). Characteristics

that predict risk of dying, such as age, sex, race, and history of (cerebro)vascular disease, were of particular interest to the aim of our study. The last item was (8) low overall dropout (<20%)[18] and comparable dropout across treatment arms (<5% difference). Each item was scored as low, high, or unclear risk of bias. Disagreements were resolved by consensus.

Statistical Analyses

Results of the included RCTs were pooled to obtain an estimate of the risk of mortality compared with placebo of (1) all conventional antipsychotics tested in the included trials, and (2) haloperidol. We considered patients with dementia, delirium, or at risk of delirium homogeneous in terms of frailty.

Deaths in treatment groups testing different dosages of a drug within one trial were combined. A pooled risk difference (RD) was calculated with Review Manager (RevMan, Cochrane Collaboration, Cophenhagen, Denmark) version 5.3 software by one author, and reproduced by another in Stata (Stata Corp, College Station, TX). RevMan software also calculated heterogeneity, presented as I^2. A fixed effects model was used because heterogeneity was found to be below 30%.[19] *P* values and 95% CIs were calculated around the risk estimates.

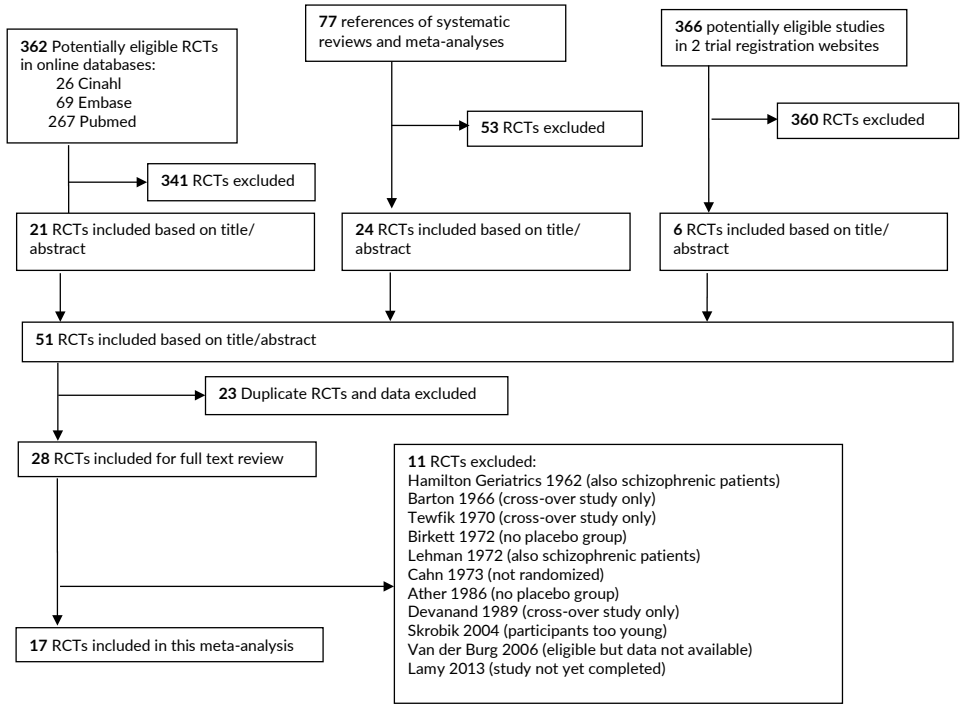


Figure 1. Flow diagram of the literature search.

Using the data of all trials and haloperidol trials, 5 subanalyses were performed in which pooled RDs were calculated for (1) RCTs in participants with dementia, (2) RCTs in participants at risk of delirium, (3) RCTs with no baseline differences, (4) RCTs with adequate blinding, and (5) RCTs with low overall dropout and comparable dropout across treatment groups. We did not study the effect of items 4 to 6 of the Cochrane assessment tool because these items, although indicative for the quality of a trial, relate to efficacy outcomes instead of mortality risk. In addition, risk ratios (RRs) were calculated in RevMan to enable comparison with the results of previous studies. Because of zero events in most intervention groups, we used the overall unweighted number of events and participants randomized to a conventional antipsychotic or placebo. Also, a hazard ratio was calculated in Stata to take into account that trial duration varied between dementia and delirium-prevention trials, and exposure duration was shorter than trial duration in delirium-prevention trials. Person-days were calculated by multiplying the number of patients with the exposure duration in days. Patients who died or dropped out during the study were considered to have contributed person-days during half of the exposure period. Finally, we used RevMan software to create a funnel plot, and GRADEpro software version 3.6 (McMaster University, Hamilton, Canada) to make an evidence profile that presents the quality of the evidence.

Results

We identified 362 potentially relevant RCTs in the electronic databases, and checked 77 references of systematic reviews and meta-analyses (see Figure 1). After screening titles and abstracts, full text articles or online protocols of 28 RCTs were retrieved. Finally, 17 RCTs met inclusion criteria.[20-36] Data of one additional eligible, unpublished RCT found on a trial registration website was not provided by the corresponding authors (NCT00250237). We reran the search in June 2014, and found no additional studies. Conventional antipsychotics were compared with placebo in 17 RCTs that included 2387 participants.[20-36] Table 1 presents the study characteristics of these trials. They were published between 1962 and 2012. Fourteen studies were performed in patients with Alzheimer or vascular dementia. The follow-up periods varied between 2.5 and 16.0 weeks. The other 3 studies were delirium-prevention trials in elderly patients who had undergone noncardiac surgery. These trials had 5 to 14 days of follow-up. We did not find a placebo-controlled trial that tested the efficacy of conventional antipsychotics to reduce psychotic symptoms of current delirium.

We assessed the risk of bias in the 17 RCTs (Table 2). In 14 studies, the random sequence generation or allocation concealment was not (adequately) described.

Twelve RCTs did not provide baseline characteristics per treatment group, or showed possible baseline differences in the risk of dying. Blinding procedures were unclear in 9 RCTs. Eleven trials had a high overall dropout rate (>20%) or incomparable dropout rates between treatment arms (>5%).

Five articles specifically mentioned the mortality rates of the treatment groups.[20,26,32,35,36] For one RCT, the rates were retrieved from a previously published systematic review.[2,29] Eleven RCTs did not mention the number of deaths, but 3 authors provided the information on request.[27,31,33] In 3 other studies, the number of deaths in the groups was likely to be zero because no dropouts were reported.[21,22,25] For the 5 remaining trials, the number of deaths was assumed to be zero because deaths were not reported despite elaborate information on other adverse effects.[23,27,31,33,34]

A total of 32 deaths occurred in the 17 RCTs: 17 in the active treatment groups, and 15 in the placebo groups. The (rounded) mortality rate was 1.4% for the active treatment groups and 1.3% for the placebo groups. The RD of 0.1% was not statistically significant (95% CI -1.0%-1.2%; *P*= .84). Figure 2 presents the forest plot for the pooled analysis of all trials. Heterogeneity of the included trials in these analyses was low (I^2= 0%). Apart from one of the smaller trials, which had a relatively high rate of deaths in the placebo group,[26] the mortality risks were not substantially different between studies. The RD was 0.3% (95% CI -0.01-0.01; *P*= .61) when we excluded this trial from the analysis.

We repeated the meta-analysis for the 11 trials with one or more haloperidol groups (n= 1799). The RD was 0.4% and again not statistically significant (95% CI -0.9%-1.6%; *P*= .57) (see Figure 1 in the appendix). Heterogeneity was low for this subset of trials (I^2= 0%).

Table 3 shows the results of the sensitivity analyses. The point estimate was somewhat higher in the dementia trials (0.5%), but lower in delirium-prevention trials (-0.4%). The risks were also lower in trials with low risk because of baseline differences (0.0%), inadequate blinding (0.0%), or dropout (-0.6%). A similar pattern occurred for the haloperidol trials (see Table 1 in the appendix).

The RR for all conventional antipsychotics was 1.07 (95% CI 0.54-2.13), and the hazard ratio was 1.06 (95% CI 0.63-1.78). The RR for haloperidol was 1.25 (95% CI 0.59-2.65), and the hazard ratio 1.21 (95% CI 0.66-2.22).

Table 1. Characteristics of randomized placebo-controlled trials that tested conventional antipsychotics in elderly patients

Author, Year	n Randomized to Conventional Antipsychotic or Placebo	Drugs Tested	Setting, Country	Daily dosage, mg, Mean [range]	Trial Duration, wk	Age, Mean	Women, %	Total Drop-out, %
Dementia trials								
Hamilton, 1962	27	Trifluoperazine	1 hospital, USA	[4.0 – 8.0]	8	71.0	NA	4
Sugerman, 1964	18	Haloperidol	1 hospital, USA	[0.5 – 3.0]	6	71.5*	94.4	0
Rada, 1976	42	Thiothixene	1 hospital, USA	[6.0 – 15]	4	75.5	50.0	0
Barnes, 1982	53°	Thioridazine Loxapine	Nursing homes, USA	62.5 10.5	8	83.0	NA	9
Petrie, 1982	61±	Haloperidol Loxapine	1 hospital, USA	4.6 [2.0 – 10] 21.9 [10 – 50]	8	72.7	50.8	39
Stotsky, 1984	358	Thioridazine	5 hospitals/nursing homes, USA	[10 – 200]	4	76.2	54.8	0
Finkel, 1995	35	Thiothixene	1 nursing home, USA	4.6 [0.25 – 18]	11	85.0	86.0	11
Auchus, 1997	12	Haloperidol	1 out-patient center, USA	3.0	6	75.6	66.7	25
Devanand, 1998	66	Haloperidol	1 memory clinic, USA	[0.5 – 3.0]	6	72.1	64.8	9
De Deyn, 1999	229	Haloperidol	51 nursing homes, Europe/Canada	1.2 [0.5 – 4.0]	12	81.0*	56.0	53
Allain, 2000	204	Haloperidol	116 hospitals/nursing homes, Europe	3.5 [2.0 – 6.0]	3	79.6	64.0	23
Teri, 2000	70	Haloperidol	21 community centers, USA	1.8 [0.0 – 3.0]	16	75.6	54.7	36
Pollock, 2002	54	Perphenazine	1 nursing home, USA	6.5	2.5	79.7	66.6	56
Tariot, 2006	193	Haloperidol	47 nursing homes, USA	1.9 [0.5 – 12]	10	83.2	73.0	54
Delirium prevention trials								
Kaneko 1999	78	Haloperidol	1 hospital, Japan	5.0	0.7†	72.8	35.9	3
Kalisvaart, 2005	430	Haloperidol	1 hospital, Netherlands	1.5‡	2	79.2	80.0	8
Wang, 2012	457	Haloperidol	2 hospitals, China	1.7§	1	74.2	36.9	0

NA: data not available; *Median; †5 days; ‡Medication was started at admission and continued until 3 days after surgery. A maximum delay of surgery of 72 hours was permitted; §Mean total dose administered in first twelve hours after entry into trial.
°Thiordazine: n=17, and loxapine: n=19;
±Haloperidol: n=20, and loxapine: n=19;

The quality of the overall evidence is summarized in a GRADE evidence profile (Table 4). We graded our confidence in the evidence provided by the review and meta-analysis of the included trials as high. Some risk of bias seemed present in the individual trials, but the sensitivity analyses among studies with low risk of bias (sensitivity analyses 3 to 5) showed that, if anything, the pooled RDs of 0.1% for all trials and 0.4% for haloperidol are overestimations of the true mortality risks. Also no inconsistency was observed. Precision was considered sufficient (see Discussion). One completed but unpublished study was identified, but the funnel plot did not indicate publication bias for all conventional antipsychotics (Figure 2 in the appendix) or for haloperidol (see Figure 3 in appendix).

Table 2. Risk of Bias

	Random sequence generation	Allocation concealment	Blinding of participants, personnel and outcome	Incomplete outcome data addressed	Absence of selective reporting	Absence of commercial funding	No baseline differencesbetween treatment groups	Drop-out rates < 20% and the same in both arms
Dementia								
Hamilton 1962	U	U	L	H	H	U	U	L
Sugerman 1964	U	U	U	L	H	H	U	L
Rada 1976	U	U	U	H	U	U	U	U
Barnes 1982	U	U	L	H	U	H	U	H
Petrie 1982	H	U	U	H	H	H	U	H
Stotsky 1984	U	U	U	H	U	U	U	L
Finkel 1995	U	U	L	H	H	H	U	L
Auchus 1997	H	U	U	L	L	L	L	U
Devanand 1998	H	U	U	H	L	L	L	H
De Deyn 1999	L	L	L	H	L	H	H	H
Allain 2000	H	U	U	L	U	H	U	H
Teri 2000	U	L	L	L	U	H	L	H
Pollock 2002	U	U	U	H	H	L	L	H
Tariot 2006	H	U	L	L	U	H	H	H
Delirium								
Kaneko 1999	H	L	U	H	H	U	U	H
Kalisvaart 2005	L	L	L	U	U	L	L	L
Wang 2012	L	L	L	L	L	L	H	L

Post hoc Analyses

Of one of the included delirium-prevention trials, we included deaths occurring in the study period of 7 days in our meta-analysis.[36] The publication also provided 28-day mortality rates: 2/229 for haloperidol versus 6/228 for placebo.[36] When we used these rates in the meta-analysis of all trials, the pooled RD changed slightly from 0.1% (95% CI -1.0%-1.2%) to 0.0% (95% CI -1.0%-1.6%).

In addition, 2 trials tested thioridazine, which was discontinued worldwide because of concerns of cardiotoxicity and retinopathy 3 years before the warning against conventional antipsychotics. Excluding the thioridazine groups from the analysis did not change the pooled RD either (RD 0.1%; 95% CI -1.1%-1.4%; *P*= .84).

Table 3. Results of sensitivity analyses

Selected trials	*RCTs included*	*RD, %*	*95% CI, %*	*P-value*
1.1 Dementia trials	20-30,32,33,35	0.5	-1.3-2.2	0.59
1.2 Delirium prevention trials	31,34,36	-0.4	-1.3-0.5	0.37
1.3 No baseline differences	27,28,30,33,34	0.0	-1.4-1.4	1.00
1.4 Adequate blinding	20,23,26,29,30,34-36	0.0	-1.5-1.5	0.96
1.5 Drop-out low and similar across groups	20,21,25,26,34,36	-0.6	-1.6-0.4	0.23

RD: risk difference

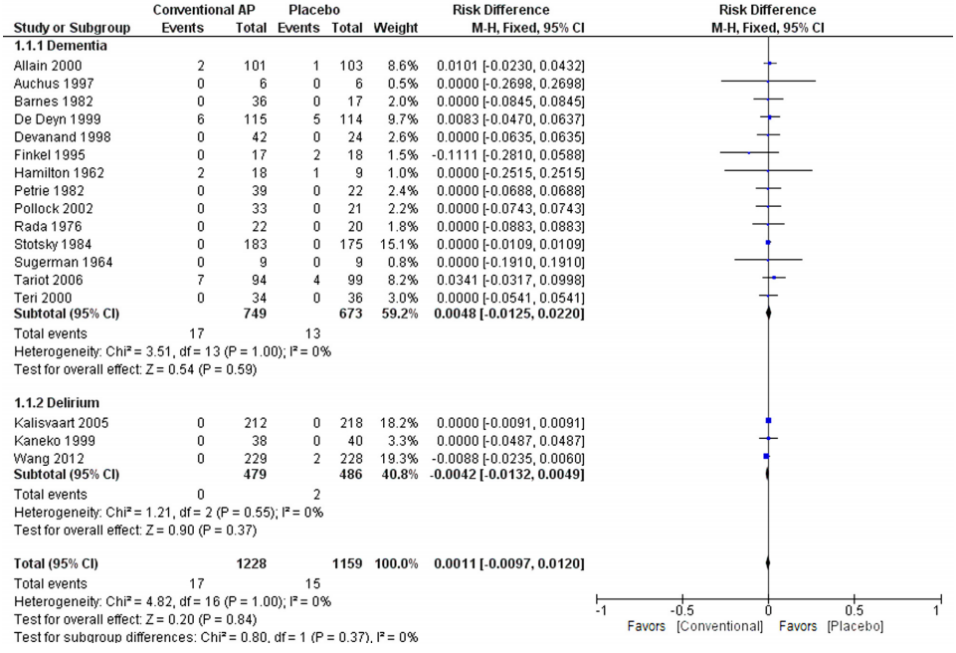


Figure 2. Forest plot of all trials.

Discussion

Our meta-analysis of 17 randomized trials in 2387 patients with dementia or at risk of delirium did not confirm that conventional antipsychotics or haloperidol in particular have a higher mortality rate than placebo. The RD was 0.1% for all conventional antipsychotics (95% CI -1.0%-1.2%), and 0.4% for haloperidol (95% CI -0.9%-1.6%) compared with placebo.

Many cohort studies have reported an association between conventional antipsychotics and an increased risk of mortality when compared with atypical antipsychotics or no use in elderly patients.[3] The RD varied between 4.2% and 7.3% for the first 6 months of use, and the hazard ratio between 1.26 and 1.47. However, none of the studies adjusted for severity of illness.[8] Therefore, terminal illness and comorbid delirium may have consistently confounded the observational findings. A strength of our study is that we provide a systematic review of randomized trials. The goal of randomization is to distribute baseline characteristics evenly among treatment groups. We found an RD of 0.1% and hazard ratio of 1.06 for all conventional antipsychotics. It is possible that study populations of the cohort studies and trials differed. The cohort studies involved any elderly patient who had received a conventional antipsychotic in hospital, a nursing home, or sometimes at home. The prevalence of cerebrovascular disease, heart failure, dementia, and depression was generally high. Many cohort studies excluded patients with schizophrenia. Trials on the other hand excluded patients who are terminally ill. They included patients treated for symptoms of dementia or for prevention of delirium after noncardiac surgery. These patient groups can be considered frail because of their age, cognitive impairment, somatic morbidity, and functional disability.

Table 4. GRADE evidence profile

Quality assessment					
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**
Conventional antipsychotics versus placebo					
17	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision
Haloperidol versus placebo					
11	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision

[1]If anything, the risk estimate is an overestimation of the true risk.

One previous meta-analysis of trials reported a mortality risk of haloperidol in elderly patients.[2] The main focus of this meta-analysis was to compare mortality rates of atypical antipsychotics with placebo in patients with dementia. An ad hoc analysis of 2 trials that had a third haloperidol group showed that the risk of mortality was higher for haloperidol than placebo (RR 2.07; 95% CI 0.78-5.51).[2] We found an RR that was substantially lower (RR 1.25; 95% CI 0.59-2.65) based on 11 trials. The previous meta-analysis reported a mortality rate of 6.2% for haloperidol, and 3.8% for placebo, whereas in our meta-analysis these rates were 1.7% and 1.3%, respectively. Not just the higher number of trials that we included may explain the disparate findings. The previous meta-analysis included data from a subgroup of trial participants with schizophrenia and other psychotic disorders not related to dementia. In this trial, the mortality rate of haloperidol versus placebo was higher among all participants (RR 2.3) than in the patients with Alzheimer dementia (RR 1.8).[35] Moreover, the previous meta-analysis was based on 2 RCTs with a high risk of bias owing to baseline differences among treatment groups and high dropout. For example, the haloperidol groups had higher proportions of men or persons with vascular dementia[29,35] and these factors are predictive of death in patients with dementia.[37] Our sensitivity analyses yielded lower risk estimates for trials with low risk of bias due to either baseline differences or dropout. That these trials were adequately blinded seems less important for an objective outcome, such as death.[38]

In general, the quality of the included trials seemed poor. Fifteen of 17 individual trials scored unclear or high risk of bias on 4 or more of the 8 quality items that we scored. Only 2, more recently published, trials scored low risk of bias on most items.[34,36] Nine trials scored unclear risk because of inadequate blinding, probably because a relatively new requirement is that not only the patient and outcome

	No of patients		Effect		Quality	Im-portance
Other considerations	Conventional Antipsychotics	Control	Relative (95% CI)	Absolute		
none[1]	17/1228 (1.4%)	15/1159 (1.3%)	RR 1.07 (0.54 to 2.13)	1 fewer per 1000 (from 0.01 fewer to 0.01 more)	++++ HIGH	
		0%		-		
none[1]	15/900 (1.7%)	12/899 (1.3%)	RR 1.25 (0.59 to 2.65)	4 fewer per 1000 (from 0.01 fewer to 0.02 more)	++++ HIGH	
		0%		-		

assessor are blinded (double-blind), but also the health care professionals attending to the patients. Way of blinding, for example through identical tablets of haloperidol and placebo, was often not specified in the older publications. However, publication date may not fully explain the overall poor study quality. Twelve of 17 trials scored an unclear or high risk of bias owing to baseline differences despite randomization, because for instance tables with baseline characteristics per groups were missing. Dropout rates were also high independent of publication date. Noteworthy is that investigators of one trial performed discontinuation visits to reduce bias due to dropout.[30] Although the individual trials may not have been free of bias, as part of our GRADE rating, we rated the risk of bias in the overall evidence as "not serious." This is related to the direction of the presumed bias.[18] Our sensitivity analyses among studies with low risk of bias showed that, if anything, the pooled RD of 0.1% for conventional antipsychotics and RD of 0.4% for haloperidol are overestimations of the true mortality risks.

We rated the power of the meta-analyses and thereby precision of the results as sufficient. As mentioned previously, cohort studies have found an RD of at least 4.2% for conventional versus atypical antipsychotics in elderly populations.[3] None of the cohort studies reported risk differences for haloperidol separately, but in general the RRs of haloperidol were higher for haloperidol than all conventional antipsychotics combined. A post hoc calculation showed that the power of our meta-analysis of all trials was 100% to detect an RD of 4.2% or higher, assuming a placebo rate of 1.3% (as we found in our meta-analysis). Put differently, the total number of patients included in our meta-analysis (n= 2387 for all conventional antipsychotics, and 1799 for haloperidol) exceeded the number needed for a single adequately powered trial given an alpha of 0.05, and beta of 0.2 (n= 1762).[39,40] For the subgroup of patients with dementia, to which the warning against conventional antipsychotics refers, one cohort study reported a statistically significant mortality risk of 2.6%.[3] The power of our analysis of dementia trials was 80.0% to detect an RD of at least 2.6%, assuming a placebo rate of 1.9%.

The power of our analyses was not sufficient to detect an RD of 1%. This is the mortality risk that Schneider et al.[2] found for atypical antipsychotics versus placebo, and it led to a warning from health authorities against use of atypical antipsychotics in patient with dementia.[41] It is, however, questionable whether an RD of 1% is considered clinically relevant in daily practice. In the years following the warning, use of atypical antipsychotics decreased only slightly from 18% to 17% in Canadian nursing home patients with dementia, and 14% to 11% in American outpatients with dementia, and increased again after the warning against conventional antipsychotics was issued.[42,43] Perhaps clinicians, patients, and family do not consider a 1% increase

in risk of mortality clinically relevant. Patients with dementia already have a poor prognosis, especially when psychosis or aggression co-occurs. The opportunity of diminishing these symptoms with an antipsychotic and thereby increasing quality of life may outweigh the small increase in risk of dying.

The results of this meta-analysis refute the observational association between conventional antipsychotics in general, or haloperidol in particular. They do not confirm an increased mortality risk for conventional antipsychotics in elderly patients with dementia either. Except for one delirium-prevention trial, the trials that we used were available in 2008. Hence, our findings question the scientific support for the warning against conventional antipsychotics in patients with dementia that was based on the observational findings.

References

1. Lanctot K, Best T, Mittmann N, et al. Efficacy and safety of neuroleptics in behavioral disorders associated with dementia. Clin Psychiatry 1998;59: 549e563.
2. Schneider L, Dagerman K, Insel P. Risk of death with atypical antipsychotic drug treatment for dementia. J Am Med Assoc 2005;294:1934e1943.
3. Pratt N, Roughead E, Salter A, Ryan P. Choice of observational study design impacts on measurement of antipsychotic risks in the elderly: A systematic review. BMC Med Res Methodol 2012;12:72.
4. Hollis J, Forrester L, Brodaty H, et al. Risk of death associated with antipsychotic drug dispensing in residential aged care facilities. Aust N Z J Psychiatry 2007; 41:751e758.
5. Rossom R, Rector T, Lederle F, Dysken M. Are all commonly prescribed antipsychotics associated with greater mortality in elderly male veterans with dementia? J Am Geriatr Soc 2010;58:1027e1034.
6. US Food and Drug Administration. Information for healthcare professionals: Conventional antipsychotics. Available at: http://www.fda.gov/drugs/drug safety/postmarketdrugsafetyinformationforpatientsandproviders/ucm124830. htm; 2008. Accessed June 12, 2013.
7. Drug Safety Update. Conventional (typical) antipsychotics: Increased mortalityin dementia. MHRA. Available at: http://webarchive.nationalarchives.gov.uk/ 20141205150130/http://www.mhra.gov.uk/Safetyinformation/DrugSafetyUpdate/ CON087985; 2008. Accessed June 12, 2013.
8. Luijendijk H, De Bruin N, Koolman X. Haloperidol in elderly users: Not a cause of death, but caused by impeding death? Am J Epidemiol 2013;177:S1eS181.
9. Maltoni M, Scarpi E, Rosati M, et al. Palliative sedation in end-of-life care and survival: A systematic review. J Clin Oncol 2012;30:1378e1383.
10. Mercadante S, Fulfaro F, Casuccio A. Pattern of drug use by advanced cancer patients followed at home. J Palliat Care 2001;17:37e40.
11. Lonergan E, Luxenberg J, Colford J, Birks J. Haloperidol for agitation in dementia (Review). Cochrane Libr 2002;12:1e42.
12. Hasegawa N, Hashimoto M, Yuuki S, et al. Prevalence of delirium among outpatients with dementia. Int Psychogeriatr 2013;25:1877e1883.
13. Sandberg O, Gustafson Y, Brannstrom B, Bucht G. Prevalence of dementia, delirium and psychiatric symptoms in various care settings for the elderly. Scand J Soc Med 1998;26:56e62.
14. Friedman J, Soleimani L, McGonigle D, et al. Pharmacological treatments of non-substance-withdrawal delirium: A systematic review of prospective trials. Am J Psychiatry 2014;171:151e159.
15. The Grading of Recommendations Assessment Development and Evaluation.GRADE working group. GRADE.

Available at: http://www.gradeworkinggroup. org/publications/JCE_series.htm. Accessed April 15, 2013.
16. Wikipedia. Typical antipsychotic. Wikipedia, Free Encycl. 2014. Available at: http://en.wikipedia.org/wiki/Typical_antipsychotic. Accessed July 8, 2014.
17. Higgins J, Altman D. Assessing risk of bias in included studies. In: Higgins JP,Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.0.1 [updated September 2008]. The Cochrane Collaboration; 2008. Available from www.cochrane-handbook.org. Accessed April 21, 2015.
18. Guyatt G, Oxman A, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidencedstudy limitations (risk of bias). J Clin Epidemiol 2011;64:407e415.
19. Deeks J, Higgins J, Altman D. Analysing data and undertaking meta-analyses. In: Higgins JP, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration; 2011. Available from www.cochrane-handbook.org. Accessed April 21, 2015.
20. Hamilton L, Bennett J. The use of trifluoperazine in geriatric patients with chronic brain syndrome. J Am Geriatr Soc 1962;10:140e147.
21. Sugerman A, Williams B, Alderstein A. Haloperidol in the psychiatric disorders of old age. Am J Psychiatry 1964;120:1190e1192.
22. Rada R, Kellner R. Thiothixene in the treatment of geriatric patients with chronic organic brain syndrome. J Am Geriatr Soc 1976;24:105e107.
23. Barnes R, Veith R, Okimoto J, et al. Efficacy of antipsychotic medications in behaviorally disturbed dementia patients. Am J Psychiatry 1982;139: 1170e1174.
24. Petrie W, Ban T, Berney S, et al. Loxapine in psychogeriatrics: A placebo- and standard-controlled clinical investigation. J Clin Psychopharmacol 1982;2: 122e126.
25. Stotsky B. Multicenter study comparing thioridazine with diazepam and placebo in elderly, nonpsychotic patients with emotional and behavioral disorders. Clin Ther 1984;6:546e559.
26. Finkel S, Lyons J, Anderson R, Sherrell K. A randomized, placebo-controlled trial of thiothixene in agitated, demented nursing home patients. Int J Geriatr Psychiatry 1995;10:129e136.
27. Auchus A, Bissey-Black C. Pilot study of haloperidol, and placebo agitation in Alzheimer's disease. Neuropsychiatry Clin Neurosci 1997;9:591e593.
28. Devanand D, Sackeim HA, Marder K, et al. A randomized, placebo-controlled dose-comparison trial of haloperidol for psychosis and disruptive behaviors in Alzheimer's Disease. Am J Psychiatry 1998;155:1512e1520.
29. De Deyn P, Rabheru K, Rasmussen A. A randomized trial of risperidone, placebo, and haloperidol for behavioral symptoms of dementia. Neurology 1999; 53:945e956.
30. Teri L, Logsdon RG, Peskind E. Treatment of agitation in AD: A randomized, placebo-controlled clinical trial. Neu-

rology 2000;55:1271e1277.
31. Kaneko T, Cai J, Ishikura T, Kobayashi M. Prophylactic consecutive administration of haloperidol can reduce the occurrence of postoperative delirium in gastrointestinal surgery. Yonago Acta Med 1999;42:179e184.
32. Allain H, Dautzenberg P, Maurer K, et al. Double blind study of tiapride versus haloperidol and placebo in agitation and aggressiveness in elderly patients with cognitive impairment. Psychopharmacology (Berl) 2000;148: 361e366.
33. Pollock B, Mulsant B, Rosen J, et al. Comparison of citalopram, perphenazine, and placebo for the acute treatment of psychosis and behavioral disturbances in hospitalized, demented patients. Am J Psychiatry 2002;159: 460e465.
34. Kalisvaart K, de Jonghe J, Bogaards M, et al. Haloperidol prophylaxis for elderly hip-surgery patients at risk for delirium: A randomized placebo-controlled study. J Am Geriatr Soc 2005;53:1658e1666.
35. Tariot P, Schneider L, Katz I, et al. Quetiapine treatment of psychosis associated with dementia: A double-blind, randomized, placebo-controlled clinical trial. Am J Geriatr Psychiatry 2006;14:767e776.
36. Wang W, Li H, Wang D, et al. Haloperidol prophylaxis decreases delirium incidence in elderly patients after noncardiac surgery: A randomized controlled trial. Crit Care Med 2012;40:731e739.
37. Molsa P, Marttila R, Rinne U. Long-term survival and predictors of mortality in Alzheimer's disease and multi-infarct dementia. Acta Neurol Scand 1995;91: 159e164.
38. Wood L, Egger M, Gluud L, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: Metaepidemiological study. BMJ 2008;336:601e605.
39. Brant R. Inference for proportions: Comparing two independent samples. Available at: http://www.stat.ubc.ca/wrollin/stats/ssize/b2.html; 2014. Accessed July 8, 2014.
40. Guyatt G, Oxman A, Kunz R, et al. GRADE guidelines: 6. Rating the quality of evidence imprecision. J Clin Epidemiol 2011;64:1283e1293.
41. US Food and Drug Administration. Public Health Advisory: Deaths withantipsychotics in elderly patients with behavioral disturbances. Available at: http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationfor PatientsandProviders/DrugSafetyInformationforHeathcareProfessionals/Public HealthAdvisories/ucm053171.htm; 2013. Accessed June 12, 2013.
42. Kales H, Zivin K, Kim H, et al. Trends in antipsychotic use in dementia 1999e2007. Arch Gen Psychiatry 2011;68:190e197.
43. Alessi-Severini S, Dahl M, Schultz J, et al. Prescribing of psychotropic medications to the elderly population of a Canadian province: A retrospective study using administrative databases. PeerJ 2013;1:e168.

Appendix

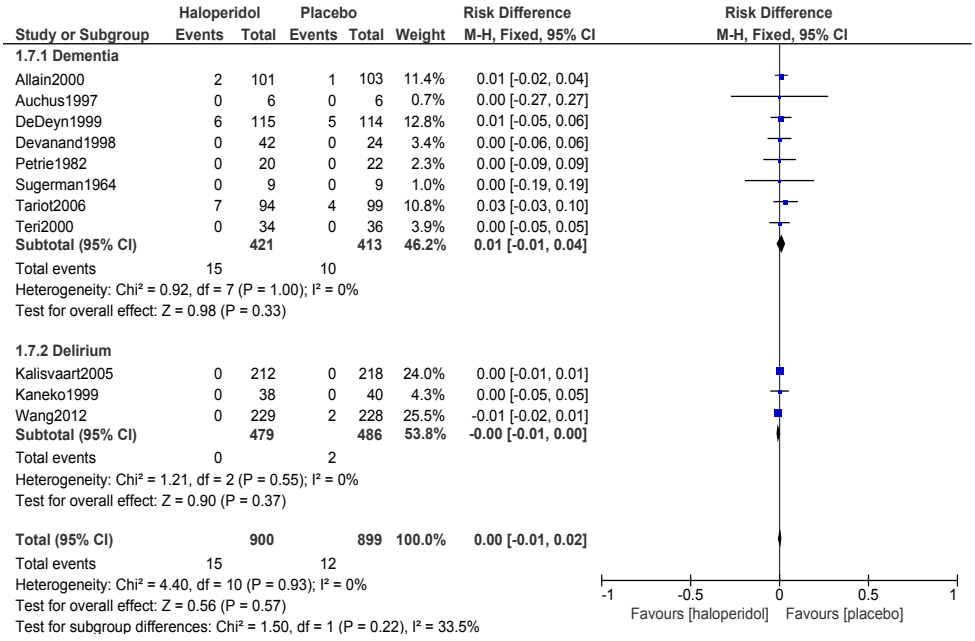
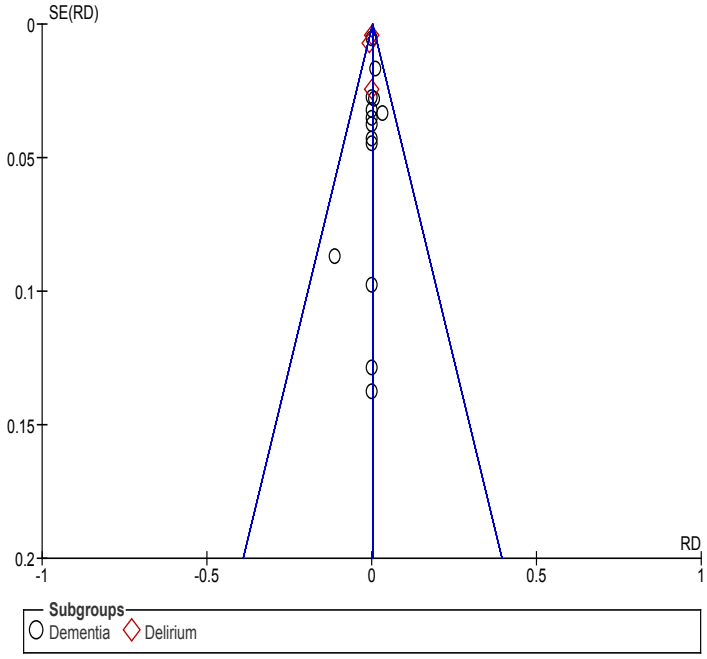


Figure 1. The mortality risk of haloperidol versus placebo

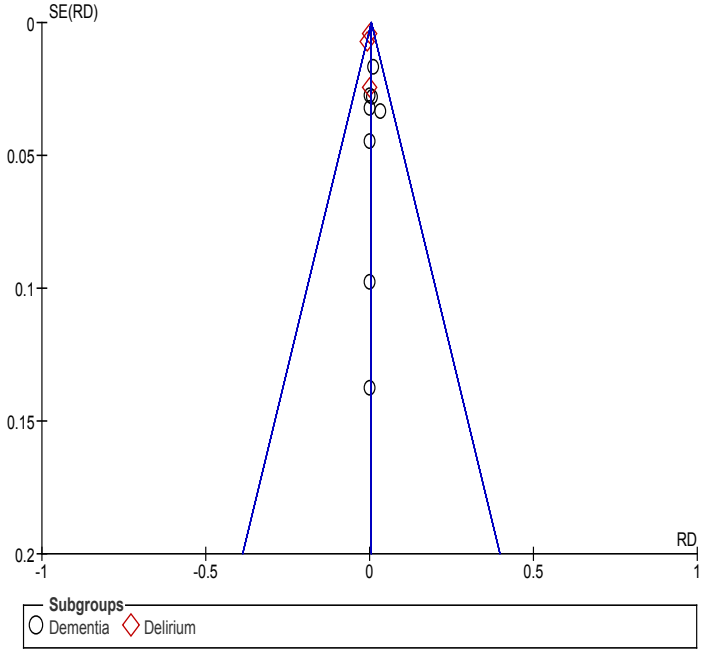
Table 1. Results of sensitivity analyses haloperidol trials only

Selected trials	*RCTs included*	*RD, %*	*95% CI, %*	*P-value*
1.1 Dementia trials	21–23,25–27,30,31	1.3%	[-1.3%;3.8%]	0.33
1.2 Delirium prevention trials	24,28,29	-0.4%	[-1.3%;0.5%]	0.37
1.3 No baseline differences	23,26,28,31	0.0%	[-1.4%;1.4%]	1.00
1.4 Adequate blinding	25,26,28–30	0.3%	[-1.1%;1.7%]	0.66
1.5 Drop-out low and similar across groups	21,28,29	-0.4%	[-1.4%;0.5%]	0.37

RD: risk difference



eFigure 2 Funnel plot of all trials



eFigure 3 Funnel plot of haloperidol trials only

Chapter 3.

Efficacy of antipsychotics in dementia depended on the definition of patients and outcomes: a meta-epidemiological study

J Clin Epidemiol. 2018 Sep;101:17-27

Smeets, C.H.W.
Zuidema, S.U.
Hulshof, T.A.
Smalbrugge, M.
Gerritsen, D.L.
Koopmans, R.T.C.M.
Luijendijk, H.J.

Abstract

Objective

Postulating that efficacy of antipsychotics for agitation and psychosis in dementia is best estimated in trials among patients with these symptoms and with symptom-specific outcomes, we investigated whether clinically broader definitions affected the pooled efficacy.

Study Design and Setting

Trials were searched in multiple databases and categorized according to patient population (agitated, psychotic, mixed) and outcome scale (agitation, psychosis, generic). Standardized mean differences with 95% confidence intervals were calculated for conventional and atypical antipsychotics separately.

Results

Thirty trials met our inclusion criteria. Conventional antipsychotics might have a small effect in agitated patients on agitation scales (-0.44; -0.88, 0.01), and in psychotic patients on psychosis scales (-0.31; -0.61, -0.02). There was no effect on generic scales. Efficacy of atypical antipsychotics was not established in agitated patients on agitation scales (-0.15; -0.43, 0.13), and in psychotic patients on psychosis scales (-0.11; -0.20, -0.03), but was small in mixed patients on agitation scales (-0.29; -0.40, -0.18).

Conclusion

Pooled efficacy of antipsychotics for agitation and psychosis in dementia is biased when based on trials that included patients without these target symptoms, or on results measured with generic scales. This finding is important for reviewers and guideline developers who select trials for reviews.

Introduction

Systematic reviews and guidelines are key information sources for clinicians who wish to practice evidence-based medicine. To ensure the validity of review results, reviewers usually adhere to internationally accepted methods, such as those described in the Cochrane Handbook and GRADE recommendations.[1,2] Both methods advise to define the research question in terms of the Patients, Intervention of interest, Comparison intervention and Outcome (PICO) a-priori.[3] Subsequently, only those trials that meet this PICO should be included in the review.

Whereas the definition of the intervention of interest and the comparison intervention seem straightforward, the patient population and outcome may deserve more attention. The Cochrane Handbook and GRADE recommendations emphasize that they need to be determined meticulously. Patients should be defined 'sufficiently broad' but 'sufficiently narrow' to include the most important characteristics.[1] If efficacy is pooled across different patient populations in which it cannot be expected to be similar, there is a risk that results of a review are not meaningful or even misleading.[1,4] With respect to defining the outcome, it is advised to focus on outcomes that are likely to be clinically relevant, and to exclude those that are 'trivial or meaningless'.[1] Pooled results based on irrelevant or intermediate outcomes might be deceptive, and may be a reason to rate down the quality of evidence.[1,4]

A problem with defining the patients and outcome appears to exist in reviews about the efficacy of antipsychotics for agitation and psychosis in dementia. Those reviews have included not only trials among patients with agitation or psychosis, but also trials among patients with neuropsychiatric symptoms (NPS) in general.[5-8] NPS can consist of agitation and psychosis, but also of depression, anxiety, night-time behavior or appetite change. As a result, those reviews were based on patients who did not necessarily all have the target symptom agitation or psychosis. For example, they may have included also patients with only depression.

Furthermore, reviews on the efficacy of antipsychotics for agitation and psychosis in dementia have pooled results that were not exclusively based on agitation- and psychosis-specific outcome scales.[5-8] Results based on generic outcome scales such as the Neuropsychiatric Inventory (NPI) and Behavioral Pathology in Alzheimer's Disease Scale (BEHAVE-AD) were included as well.[9,10] These scales cover not only agitation and psychosis, but also other NPS. Yet, a treatment effect established with a generic scale does not represent the effect on agitation or psychosis specifically, and may reflect a change in any other symptom profile. Such a change could therefore be regarded as less important or indirect to start with.

Current guidelines are based on meta-analyses of trials among patients with any kind of NPS and include treatment effects measured with generic outcome scales. These guidelines support the use of antipsychotic drugs for severe agitation and for psychosis in dementia.[11-15] Usually, they differentiate between conventional and atypical antipsychotics for their pharmacological properties, presumed mechanisms of effect, and side effect profiles. Some guidelines recommend the atypical antipsychotic risperdone as drug of first choice, or alternatively the conventional antipsychotic haloperidol.[11, 13-15]

We postulate that the best estimate for efficacy of antipsychotics in patients with dementia and agitation, respectively psychosis, is assessed in patients with the target symptom (i.e. indication) and measured with a target-specific outcome scale. We investigated whether a broad definition of patients and outcome, differs clinically from a target-specific definition, for the pooled efficacy of antipsychotics for agitation and psychosis in dementia.

The aim of this study was to assess:

1. the efficacy of conventional and atypical antipsychotics measured in patients with dementia and agitation or psychosis, and measured with agitation- or psychosis-specific outcome scales,
2. the efficacy of antipsychotics in patients with dementia and any type of NPS, and measured with agitation- or psychosis-specific outcome scales; and
3. the efficacy of antipsychotics in patients with dementia and agitation or psychosis, measured with generic outcome scales for NPS.

Methods

Search

Two researchers (TAH and HJL) searched Pubmed, Embase, Cinahl, and the Cochrane Library through August 2017 for reported trials. In addition, references of systematic reviews and meta-analyses were hand-searched for relevant trials. For unpublished trials, we searched 17 trial registration websites and the databases of the Dutch Medicines Evaluation Board and the U.S. Food and Drug Administration. Search terms included individual generic drug names in the group N05A of the World Health Organization Anatomical Therapeutic Chemical classification , 'dementia', and 'trial'.[16]

We screened title and abstracts of the hits, followed by full text review of potentially eligible studies. We included trials that met the following criteria according to two

independent reviewers (CS, HJL): 1) a randomized trial, 2) testing efficacy of oral antipsychotics against placebo, 3) in patients with Alzheimer's, vascular and/or mixed dementia, and 4) who had agitation, psychosis, or NPS in general. We used no restrictions with regard to duration, language or publication date.

Data extraction

A pair of reviewers (TAH, CS, or HJL) independently extracted the following descriptive data per trial: type of the antipsychotic drug, type of dementia, exclusion criteria with regard to psychiatric disorders including substance abuse, number of patients randomized per arm, setting, country, publication year, and trial duration. Based on the eligibility criteria of every trial, we (CS and HJL) categorized each trial into three types of patient populations: 1) dementia and (at least) agitation, 2) dementia and (at least) psychosis, or 3) dementia and any type of NPS.

We (CS and HJL) extracted the trial results in terms of the reduction in agitation, psychosis and generic NPS in the active treatment and placebo group independently, i.e. the mean change from baseline to end point with standard deviations (SDs) as measured with an agitation-specific (sub)scale, psychosis-specific (sub)scale, and generic scale respectively. For studies that used more than one scale for one outcome, we used the scale that was the reported primary outcome, and otherwise the most frequently used scale across trials. If no specific instrument for agitation or psychosis was used, we extracted the reported relevant subscale of the generic instrument (e.g. NPI-psychosis of the NPI). If only subscales of agitation- or psychosis-specific scales (e.g. the subscale Physically Non-Aggressive Behavior of the Cohen-Mansfield Agitation Inventory (CMAI)) were reported, we did not extract these data for risk of selective reporting. For trials with multiple atypical antipsychotic groups or groups with different doses, we calculated average changes and SDs for the combined groups.

We used the standard error, p-value, t-value or confidence interval (CI) to calculate missing SDs. If this information was missing as well, we imputed the SD with that from another trial or cohort study with the same patient population and outcome scale:[17] we used the SDs of Tariot 2006 to impute the SD in the trials of Barnes 1982, Petrie 1982, and Devanand 1998, and the SDs of Finkel 1995 in the trial of Auchus 1997.[18-23] Discrepancies in study selection and data extraction were discussed until consensus was reached (CS and HJL).

Statistical analysis

For each combination of patients and outcome, we pooled trial results using standard meta-analysis. Because different scales were used for one outcome, we calculated

standardized mean differences (SMDs) with a 95% confidence intervals (CIs). Analyses for conventional and atypical antipsychotic trials were performed separately. We applied a fixed effects model when heterogeneity between the trials was low (I-squared below 40% and p-value of standard chi-square statistic above 0.05), and otherwise a random-effects model.[24] All analyses were performed using Stata statistical software version 13.1.17 (StataCorp, College Station, Texas). We applied the traditional SMDs cut-offs to compare the pooled results: we interpreted -0.2 or lower as a negligible treatment effect, -0.2 to -0.5 as a small treatment effect (noticeably smaller than medium but not so small as to be trivial), -0.5 to -0.8 as a medium effect (likely to be visible to the naked eye of a careful observer), and above -0.8 as a large effect.[25]

Sensitivity analyses

We conducted three sensitivity analyses: (1) including only trials with haloperidol and risperidone, since they are the most frequently studied and used conventional respectively atypical antipsychotic drug; (2) including only trials among patients with agitation but without psychosis; (3) including only trials that did not require imputation of missing data.

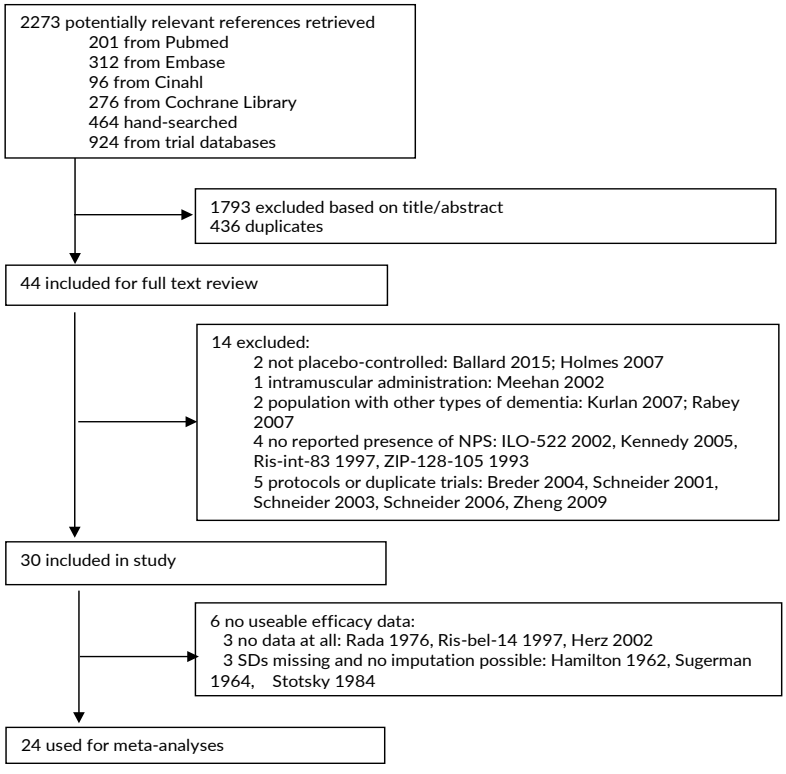


Figure 1. Flow diagram of literature search and study selection

Results

Our search strategy yielded 2363 hits, of which 44 underwent full text review.[18-23, 26-63] Thirty trials met the inclusion criteria.[18-23, 26-48, 63] All trials were written in English. Six of these did not provide data that could be pooled, but we describe these results narratively as part of our review.[28-31,34,63] Figure 1 presents our search including reasons for exclusion.

Table 1 summarizes the general study characteristics of the 30 trials. The trial populations consisted of patients with Alzheimer's, vascular, or mixed dementia that resided in nursing homes, hospitals, or the community. The trials included 12 to 652 patients, lasted 2.5 to 36 weeks, and were published between 1962 and 2008. Eleven trials assessed conventional antipsychotics (haloperidol, thioridazine, thiothixene, trifluoperazine, loxapine, perphenazine), 16 atypical antipsychotics (risperidone, olanzapine, quetiapine, aripiprazole, tiapride), and three assessed both classes of drugs in a three-armed trial.

In eight trials, the investigated patients had agitation, in nine trials psychosis, and in 13 any NPS. Patients with agitation had been defined as eligible for a trial if they had shown aggression, inappropriate verbal or motor activity, hostility, tension, uncooperativeness, excitement, or poor impulse control. Patients with psychosis had been included in a trial if they had had delusions, hallucinations, conceptual disorganization, suspiciousness, or unusual thought content. The category 'any NPS' encompassed trials that had included patients with any NPS. In 21 trials, eligibility had been determined with an assessment instrument, in the other nine trials with the clinical observation of the target symptom by a health care professional or caregiver. Many trials excluded patients with a history of psychiatric disorders such as schizophrenia and depression. For details, see Appendix A.

Outcomes had been measured with agitation-specific scales in 14 trials, with psychosis-specific scales in 13 trials and with generic scales that measure NPS in 26 trials.

Table 1 presents the trials we included for each combination of patients and outcome scale. The number of trials that could be included in the meta-analyses varied. For example, in the meta-analysis on efficacy of conventional antipsychotics among patients with agitation and measured with agitation-specific outcomes, we included four trials.[22,23,26,27] For conventional antipsychotics, five of six meta-analyses were based on one trial; the analysis in patients with agitation on agitation outcomes included four trials. For atypical antipsychotics, the meta-analyses were based on two to eight trials.

Table 1. Study characteristics of antipsychotic trials for NPS in dementia

Publication	Drug	Eligibility criterion	Setting
Conventional antipsychotics			
Patients with agitation			
Finkel 1995	thiothixene	CMAI	NH
Auchus 1997	haloperidol	CMAI	OUT
Allain 2000[a, c]	haloperidol	MOSES	NH/HOS
Teri 2000[c]	haloperidol	clinically	OUT
Patients with psychosis			
Hamilton 1962	trifluoperazine	clinically	HOS
Tariot 2006[a]	haloperidol	BPRS + NPI	NH
Patients with any NPS			
Sugerman 1964	haloperidol	clinically	HOS
Rada 1976	thiothixene	clinically	HOS
Barnes 1982	thioridazine, loxapine	clinically	NH
Petrie 1982	loxapine	clinically	HOS
Stotsky 1984	thioridazine	clinically	NH/HOS
Devanand 1998	haloperidol	(SADS + BPRS) / BSSD	OUT
De Deyn 1999[a]	haloperidol	BEHAVE-AD	NH
Pollock 2002	perphenazine	NRS	NH
Atypical antipsychotics			
Patients with agitation			
Allain 2000[a, c]	tiapride	MOSES	NH/HOS
Herz 2002	risperidone, olanzapine	CGS + ADAS/BPRS	UNK
Brodaty 2003	risperidone	CMAI	NH
Ballard 2005	quetiapine	CMAI + NPI	NH
Zhong 2007	quetiapine	PANSS-EC	NH
Patients with psychosis			
Satterlee 1995	olanzapine	BEHAVE-AD	UNK
De Deyn 2004	olanzapine	clinically	NH/HOS
Deberdt 2005	risperidone, olanzapine	NPI-NH	NH/OUT
De Deyn 2005	aripiprazole	NPI	OUT
Mintzer 2006	risperidone	BEHAVE-AD	NH
Tariot 2006[a]	quetiapine	BPRS	NH
Mintzer 2007	aripiprazole	NPI-NH	NH
Streim 2008	aripiprazole	NPI-NH	NH
Patients with any NPS			
Ris-bel-14 1997	risperidone	clinically	UNK
De Deyn 1999[a]	risperidone	BEHAVE-AD	NH
Katz 1999	risperidone	BEHAVE-AD	NH
Street 2000	olanzapine	NPI-NH	NH
Sultzer 2008	risperidone, olanzapine, quetiapine	BPRS/NPI	OUT
Paleacu 2008	quetiapine	NPI	UNK

[a] Trials that investigated both conventional and atypical antipsychotics.

[b] Although measured, no useable data

[c] Patients with psychosis excluded

Country	N	Duration, weeks	Outcome Agitation	Psychosis	Generic
USA	35	11	CMAI	-	-
USA	12	6	CMAI	-	-
Europe	204	3	MOSES I/A	-	-
USA	70	16	CMAI	-	BRSD
USA	27	8	-	-	MACC[b]
USA	193	10	-	NPI-NH P	BPRS
USA	18	6	-	-	PSC[b]
USA	42	4	-	-	NR[b]
USA	53	8	-	-	BPRS
USA	61	8	-	-	BPRS
USA	358	4	HS A[b]	-	HS[b]
USA	66	6	-	-	BPRS
Europe, Canada	229	12	BEHAVE-AD A	BEHAVE-AD P	BEHAVE-AD
USA	54	2.5	NRS A[b]	NRS P[b]	NRS
Europe	205	3	MOSES I/A	-	-
USA	29	6	CMAI[b]	-	BPRS[b]
Australasia	345	12	CMAI[b]	-	BEHAVE-AD
UK	62	6	CMAI	-	-
USA	333	10	CMAI	-	NPI-NH
UNK	238	8	-	-	BEHAVE-AD
worldwide	652	10	-	NPI-NH P	NPI-NH
USA	494	10	-	NPI-NH P	NPI
USA	208	10	-	NPI-NH P	NPI
USA	473	8	-	BEHAVE-AD P	BEHAVE-AD
USA	190	10	-	NPI-NH P	BPRS
worldwide	487	10	-	NPI-NH P	NPI-NH
USA	256	10	-	NPI-NH P	NPI-NH
UNK	39	4	-	-	NR[b]
Europe, Canada	229	12	BEHAVE-AD A	BEHAVE-AD P	BEHAVE-AD
USA	625	12	BEHAVE-AD A	BEHAVE-AD P	BEHAVE-AD
USA	206	6	NPI-NH A	NPI-NH P	NPI-NH
USA	421	36	BPRS A	BPRS P	BPRS
Israel	40	6	NPI A	NPI P	NPI

A: agitation subscore; ADAS: Alzheimer's Disease Assessment Scale;[64] BEHAVE-AD: Behavior Pathology in Alzheimer's Disease Rating Scale;[10] BPRS: Brief Psychiatric Rating Scale;[65] BRSD: Behavioral Rating Scale for Dementia;[66] BSSD: Behavioral Syndromes Scale for Dementia;[67] CGS: abbreviation not written in full;[34] CMAI: Cohen-Mansfield Agitation Inventory;[68] HOS: hospitalized; HS: modified Hamilton Anxiety Scale (no reference reported); I/A: irritability/aggressiveness subscale; MACC: Motility Affect Cooperation Communication behavioral adjustment scale;[69] MOSES: Multidimensional Observation Scale for Elderly Subjects;[70] N: number of patients randomized; NH: nursing home; NPI: Neuropsychiatric Inventory;[9] NPI-NH: Neuropsychiatric Inventory – Nursing Home version;[71] NR: used instrument was not reported; NRS: Neurobehavioral Rating Scale;[72] OUT: outpatient; P: psychosis subscore; PANSS-EC: Positive and Negative Syndrome Scale - Excitement Component;[73] PSC: Psychiatric Symptom Checklist (no reference reported); SADS: Schedule for Affective Disorders and Schizophrenia;[74] UNK: unknown

Table 2 summarizes the pooled efficacy of conventional antipsychotics and atypical antipsychotics by patients and outcome scale. Results of the six trials without poolable data are described narratively in the footnote; these generally confirmed the pooled results. The appendices B, C, and D present the forest plots.

Efficacy of conventional antipsychotics

Conventional antipsychotics had a small treatment effect in patients with agitation on agitation scales (SMD -0.44; 95% CI -0.88 to 0.01), and in patients with psychosis on psychosis scales (SMD -0.31; 95% CI -0.61 to -0.02). Both results included the possibility of a negligible and a large effect. In studies among patients with any kind of NPS, the effect was again small when assessed with agitation scales (SMD -0.28; 95% CI -0.54 to -0.02) and psychosis scales (SMD -0.23; 95% CI -0.49 to 0.04), and confidence intervals were wide again. Among studies in which the effect was assessed with generic NPS scales, the point estimates did not indicate an effect in patients with agitation (SMD -0.00; 95% CI -0.47 to 0.47) or psychosis (SMD -0.04; 95% CI -0.33 to 0.26).

Efficacy of atypical antipsychotics

Atypical antipsychotics had a negligible effect with a wide confidence interval in patients with agitation measured with agitation outcome scales (SMD -0.15; 95% CI -0.43 to 0.13). The treatment effect in patients with psychosis on psychosis outcome scales was negligible as well (SMD -0.11; 95% CI -0.20 to -0.03). When assessed in patients with any NPS, a small treatment effect on agitation outcomes was found (SMD -0.29; 95% CI -0.40 to -0.18), and a negligible effect on psychosis outcomes (SMD -0.13; 95% CI -0.24 to -0.02).

Table 2. Efficacy of antipsychotic drugs according to patients and outcome

	Outcomes		
	Agitation	**Psychosis**	**Generic**
	SMD (95% CI) n/N	SMD (95% CI) n/N	SMD (95% CI) n/N
Conventional antipsychotics			
Patients with agitation	-0.44‡ (-0.88, 0.01) 4/4	NA	-0.00 (-0.47, 0.47) 1/1
Patients with psychosis	NA	-0.31 (-0.61, -0.02) 1/1	-0.04 (-0.33, 0.25) 1/2[a]
Patients with any NPS	-0.28 (-0.54, -0.02) 1/3[b]	-0.23 (-0.49, 0.03) 1/2[c]	NA
Atypical antipsychotics			
Patients with agitation	-0.15§ (-0.43, 0.13) 3/5[d]	NA	-0.22† (-0.55, 0.11) 2/3[e]
Patients with psychosis	NA	-0.11 (-0.20, -0.03) 7/7	-0.10 (-0.19, -0.02) 8/8
Patients with any NPS	-0.29 (-0.40, -0.18) 5/5	-0.13 (-0.24, -0.02) 5/5	NA

‡ random effects analysis; heterogeneity chi^2 = 7.42, df = 3 (p = 0.060); I^2 = 59.6%; Tau^2 = 0.1121

§ random effects analysis; heterogeneity chi^2 = 4.42, df = 2 (p = 0.110); I^2 = 54.7%; Tau^2 = 0.0320

† random effects analysis; heterogeneity chi^2 = 3.92, df = 1 (p = 0.048); I^2 = 74.5%; Tau^2 = 0.0419

n/N: number of trials included in the meta-analysis per number of trials that measured this specific outcome and specific patients, NPS: neuropsychiatric symptoms

a: No data from one negative trial (n = 27).[28]

b: Excluding one positive trial that reported 0.9 improvement in the intervention group versus 0.2 in the placebo group (p < 0.001) on the agitation item of the modified Hamilton Anxiety Scale ranging from 1 to 5 (n = 358);[31] and one negative trial that reported no significant difference between intervention and placebo (n = 54).[33]

c: Excluding one negative trial that reported no significant difference between intervention and placebo (n =54).[33]

d: Excluding two trials that reported only results of CMAI subscales: one positive trial that reported 7.5 point improvement on the CMAI aggression subscale in the intervention group versus 3.1 point in the placebo group (p < 0.001), and 7.3 point improvement on the CMAI non-aggression subscale in the intervention group versus 2.8 point improvement in the placebo group (p = 0.002) (n = 345);[35] and one trial that reported non-statistical difference on one CMAI item (n = 29).[34]

e: Excluding one trial that reported only the results on some BPRS items (n = 29).[34]

Table 3. Efficacy of haloperidol and risperidone according to patients and outcome

	Outcomes		
	Agitation SMD (95% CI) n/N	**Psychosis** SMD (95% CI) n/N	**Generic** SMD (95% CI) n/N
Haloperidol			
Patients with agitation	-0.30 [-0.53, -0.06] 3/3	NA	-0.00 [-0.47, 0.47] 1/1
Patients with psychosis		-0.31 [-0.61, -0.02] 1/1	-0.04 [-0.33, 0.26] 1/1
Patients with any NPS	-0.28 [-0.54, -0.02] 1/1	-0.23 [-0.49, 0.04] 1/1	NA
Risperidone			
Patients with agitation	- - 0/2[a]	NA	-0.38 [-0.61, -0.16] 1/2[b]
Patients with psychosis	NA	-0.05‡ [-0.30, 0.20] 2/2	0.06 [-0.10, 0.21] 2/2
Patients with any NPS	-0.27 [-0.40, -0.14] 3/3	-0.18 [-0.31, -0.04] 3/3	NA

‡ random effects analysis; heterogeneity chi^2 =2.55, df = 1 (p = 0.110) I^2 = 60.8%; Tau^2 = 0.0202
n/N: number of trials included in the meta-analysis per number of trials that measured this specific outcome and specific patients, NPS: neuropsychiatric symptoms

a: Excluding two trials that reported only results of CMAI subscales or items: one positive trial that reported 7.5 point improvement on the CMAI aggression subscale in the intervention group versus 3.1 point in the placebo group (p < 0.001), and 7.3 point improvement on the CMAI non-aggression subscale in the intervention group versus 2.8 point improvement in the placebo group (p = 0.002) (n = 345);[35] and one trial that reported non-statistical difference on one CMAI item (n = 29).[34]

b: Excluding one trial that did not report the results on the BPRS (n = 29).[34]

In patients with agitation, measurements with generic NPS scales yielded a small effect with a wide confidence interval (SMD -0.22; 95% CI -0.55 to 0.11), and a negligible effect (SMD -0.11; 95% CI-0.19 to -0.02) in patients with psychosis.

Sensitivity analyses

The sensitivity analyses for trials with haloperidol and risperidone showed results that were clinically similar to those from all conventional, respectively all atypical antipsychotics (see table 3). One exception was the meta-analysis of risperdone in

patients with agitation on agitation outcome scales, for which no data were available. The sensitivity analysis that included trials among patients with agitation and no psychosis, showed clinically similar results for conventional antipsychotics.[26,27] For atypical antipsychotics however, including the additional trial yielded a small effect on agitation outcomes (SMD -0.39; 95% CI -0.67 to -0.11), in contrast to when this trial was excluded (SMD -0.15; 95% CI -0.43 to 0.13).[26] The sensitivity analysis excluding trials with imputed data gave similar effect sizes to the analyses including these trials.[19-21, 23]

Discussion

Our meta-epidemiological study shows that the effect of conventional antipsychotics on agitation and psychosis might be underestimated when assessed with generic outcome scales compared to symptom-specific scales. By contrast, efficacy of atypical antipsychotics on agitation is conceivably overestimated when assessed in patients with diverse NPS and with a generic outcome scale. This implies that the precise definition of patients and choice of outcome scales affects the reported pooled efficacy of antipsychotics on agitation and psychosis in dementia. It is important to consider the potential impact of an accurate definition of the target symptom when defining trial selection criteria for a review.

Efficacy of antipsychotics in other reviews

We found that conventional antipsychotics had a small but statistically not significant treatment effect on agitation in patients with dementia and agitation (SMD -0.44; 95% CI -0.88 to 0.01), and a small treatment effect on psychosis in patients with dementia and psychosis (SMD -0.31; 95% CI -0.61 to -0.02). One prior review assessed the effect of conventional antipsychotics, that is haloperidol, on agitation in dementia.[6] This review, that included two trials in patients with any NPS, reported a negligible effect on agitation (SMD -0.12; 95% CI -0.31 to 0.08), but a small effect on aggression (SMD -0.31; 95% CI-0.49 to -0.13).[21,32] Including studies in patients with any NPS instead of with agitation specifically, might have diluted the effect of conventional antipsychotics on agitation. We found no published meta-analysis of conventional antipsychotics on psychosis outcomes to compare with our results.

For atypical antipsychotics, our meta-analysis yielded a negligible and statistically nonsignificant effect on agitation (SMD -0.15; 95% CI -0.43 to 0.13), and a negligible significant effect on psychosis (SMD -0.11; 95% CI -0.20 to -0.03). It is difficult to compare these findings with those from prior reviews that differentiated between individual antipsychotics and doses and partly reported weighted mean differences.[7,8] Nevertheless, those reviews reported modest effects on aggression

(with or without agitation) and on psychosis. Around half of the trials included in those reviews had been performed in patients with any kind of NPS. Our results indicate that the reviews' selection of trials among patients with diverse NPS might have led to overestimated efficacy of atypical antipsychotics.

Strengths and limitations

To the best of our knowledge, we are the first to study how the definition of patients and outcomes in reviews has affected the pooled efficacy of antipsychotics in dementia. We investigated conventional and atypical antipsychotics, and also the most widely used antipsychotics haloperidol and risperidone in particular. The main limitation of our study is the uncertainty around some point estimates due to the small number of trials or patients. This was especially the case for the meta-analyses of trials about conventional antipsychotics among patients with target symptoms and symptom-specific outcome scales. In some trials, outcomes of interest had been measured but not reported, or not reported in full.[30,33-35,63] For instance, we chose not to include results measured with subscales or items if results on total scales were not available, because these results may have been biased by selective reporting.

Bias

We postulated that the best estimate for efficacy of antipsychotics for agitation and psychosis in dementia, is obtained from target-specific patients and outcomes. Differences between 'target-specific' and 'non-target-specific' results of meta-analyses, indicate the presence of bias. Such bias can occur as a result of different interpretations on how to define the patients, or the outcome.

Bias due to the imprecise definition of patients

Efficacy of atypical antipsychotics on agitation appears to be higher when assessed among patients with any kind of NPS. There are a number of possible explanations for this finding. First, atypical antipsychotics may reduce other NPS that are related to agitation. Second, the efficacy of individual antipsychotics may differ. For instance, in the meta-analysis of agitation scales in trials among patients with agitation, the tiapride trial showed small efficacy for agitation, whereas the two quetiapine trials showed none (see forest plots in Appendix B.3).[26,36,37] An unequal distribution of individual drugs between meta-analyses may therefore cause bias. Third, there might be an association between publication year and type of patients enrolled in the trials. The three trials in patients with diverse NPS and the highest reported efficacy were published in or before 2000.[26,32,45,46] An unequal distribution of old and new trials might therefore also cause bias. Fourth, the aim to investigate efficacy on a broad range of NPS, and the reporting of results on items or subscales enholds the risk of positive findings by chance.

Bias due to the imprecise definition of outcome scale

Efficacy might be underestimated for conventional antipsychotics on agitation and psychosis, and overestimated for atypical antipsychotics on agitation when assessed with generic outcome scales. This bias due to definition of the outcome may be caused by an effect of the drugs on neuropsychiatric side effects, such as sedation, if these are included in a generic scale. Sedation, which is linked to increased levels of apathy, might counterbalance a decrease of agitation on generic scales. Bias could also be caused by efficacy on other NPS, such as co-existing agitation in the treatment of psychosis, or treatment of underlying psychosis when reduction of agitation was aimed for. Although generic outcomes may be of added value to symptom-specific outcomes, it is crucial to specifically interpret those that are clinically relevant. Furthermore, trials that assessed but did not report symptom-specific outcomes, need to be included in reviews because the missing outcomes can be considered a potential source of selective reporting.

Implications

Our study implies that trial selection criteria and extracted data should reflect a review's PICO in detail including the target symptom and outcome for the treatment of interest. The Cochrane Handbook and GRADE instructions address that the definition of patients and outcome as part of the research question can be challenging.[4] For the definition of the patients, there is evidently a balance between including sufficiently narrow, but not excluding relevant trials. Nevertheless, Cochrane describes a list of factors to consider for defining the patients, among which 'What are the most important characteristics?' Our results show that for efficacy of antipsychotics on agitation or psychosis in dementia, it is crucial that the target symptoms agitation, respectively psychosis, are considered an important characteristic for the PICO and selection of trials. Our results also demonstrate that it is important to interpret the pooled results on generic outcome scales with caution.

Conclusion

Our study shows that reviewers and guideline developers should define PICOs that represent the symptom of interest, and select trials accordingly. Trials among patients without these specific symptoms may give inaccurate estimates, as will trial results based on non-specific outcome scales. We conclude that the pooled efficacy of conventional and atypical antipsychotics is biased when based on trials that included patients without these target symptoms, and when generic outcome scales are used.

References

1. Higgins, J. and S. Green, Cochrane Handbook for Systematic Reviews of Interventions. The Cochrane Collaboration.
2. Guyatt, G., et al., GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol, 2011. 64(4): p. 383-94.
3. Counsell, C., Formulating questions and locating primary studies for inclusion in systematic reviews. Ann Intern Med, 1997. 127(5): p. 380-7.
4. Guyatt, G.H., et al., GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol, 2011. 64(4): p. 395-400.
5. Maglione, M., et al., Off-Label Use of Atypical Antipsychotics: An Update. 2011: Rockville (MD).
6. Lonergan, E., et al., Haloperidol for agitation in dementia. Cochrane Database Syst Rev, 2010(12).
7. Ballard, C. and J. Waite, The effectiveness of atypical antipsychotics for the treatment of aggression and psychosis in Alzheimer's disease. Cochrane Database Syst Rev, 2006(1): p. CD003476.
8. Ma, H., et al., The efficacy and safety of atypical antipsychotics for the treatment of dementia: a meta-analysis of randomized placebo-controlled trials. J Alzheimers Dis, 2014. 42(3): p. 915-37.
9. Cummings, J.L., et al., The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia. Neurology, 1994. 44(12): p. 2308-14.
10. Reisberg, B., S.R. Auer, and I.M. Monteiro, Behavioral pathology in Alzheimer's disease (BEHAVE-AD) rating scale. Int Psychogeriatr, 1996. 8 Suppl 3: p. 301-8; discussion 351-4.
11. Nederlandse Vereniging voor Klinische Geriatrie, Richtlijn Diagnostiek en Behandeling van dementie. 2014.
12. National Institute for Health and Clinical Excellence, Dementia: supporting people with dementia and their carers in health and social care. 'NICE Clinical Guideline 42', 2006.
13. American Psychiatric Association, The American Psychiatric Association Practice Guideline on the Use of Antipsychotics to Treat Agitation or Psychosis in Patients With Dementia. 2016.
14. Guideline Adaptation Committee, Clinical Practice Guidelines and Principles of Care for People with Dementia. Sydney. 2016.
15. Smalbrugge, M., et al., Guideline problem behavior, in 2008: Utrecht.
16. Nordic Council on Medicines, Guidelines for ATC classification. 1990, WHO Collaborating Center for Drugs Statistics Methodology: Oslo.
17. Furukawa, T.A., et al., Imputing missing standard deviations in meta-analyses can provide accurate results. J Clin Epidemiol, 2006. 59(1): p. 7-10.
18. Tariot, P.N., et al., Quetiapine treatment of psychosis associated with dementia: a double-blind, randomized, placebo-controlled clinical trial. Am J Geriatr Psychiatry, 2006(1064-7481

(Print)).
19. Barnes, R., et al., Efficacy of antipsychotic medications in behaviorally disturbed dementia patients. Am J Psychiatry, 1982. 139(9): p. 1170-4.
20. Petrie, W.M., et al., Loxapine in psychogeriatrics: a placebo- and standard-controlled clinical investigation. J Clin Psychopharmacol, 1982. 2(2): p. 122-6.
21. Devanand, D.P., et al., A randomized, placebo-controlled dose-comparison trial of haloperidol for psychosis and disruptive behaviors in Alzheimer's disease. Am J Psychiatry, 1998. 155(11): p. 1512-20.
22. Finkel, S.I., et al., A randomized, placebo-controlled trial of thiothixene in agitated, demented nursing home patients'. International Journal of Geriatric Psychiatry, 1995. 10(2): p. 129-136.
23. Auchus, A.P. and C. Bissey-Black, Pilot study of haloperidol, fluoxetine, and placebo for agitation in Alzheimer's disease. J Neuropsychiatry Clin Neurosci, 1997. 9(4): p. 591-3.
24. Deeks, J., J. Higgins, and D. Altman, Analysing data and undertaking meta-analyses., in Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0, J. Higgins and S. Green, Editors. 2011.
25. Cohen, J., A power primer. Psychol Bull, 1992. 112(1): p. 155-9.
26. Allain, H., et al., Double blind study of tiapride versus haloperidol and placebo in agitation and aggressiveness in elderly patients with cognitive impairment. Psychopharmacology (Berl), 2000. 148(4): p. 361-6.
27. Teri, L., et al., Treatment of agitation in AD: a randomized, placebo-controlled clinical trial. Neurology, 2000. 55(9): p. 1271-8.
28. Hamilton, L.D. and J.L. Bennett, The use of trifluoperazine in geriatric patients with chronic brain syndrome. J Am Geriatr Soc, 1962. 10: p. 140-7.
29. Sugerman, A.A., B.H. Williams, and A.M. Adlerstein, Haloperidol in the psychiatric disorders of old age. Am J Psychiatry, 1964. 120: p. 1190-2
30. Rada, R.T. and R. Kellner, Thiothixene in the treatment of geriatric patients with chronic organic brain syndrome. J Am Geriatr Soc, 1976. 24(3): p. 105-7.
31. Stotsky, B., Multicenter study comparing thioridazine with diazepam and placebo in elderly, nonpsychotic patients with emotional and behavioral disorders. Clin Ther, 1984. 6(4): p. 546-59.
32. De Deyn, P.P., et al., A randomized trial of risperidone, placebo, and haloperidol for behavioral symptoms of dementia. Neurology, 1999. 53(5): p. 946-55.
33. Pollock, B.G., et al., Comparison of citalopram, perphenazine, and placebo for the acute treatment of psychosis and behavioral disturbances in hospitalized, demented patients. Am J Psychiatry, 2002. 159(3): p. 460-5.
34. Herz, L., et al., A 6-Week, Double-Blind Comparison of Olanzapine, Risperidone, and Placebo for Behavioral Disturbance in Alzheimer's Disease. J Clin Psychiatry, 2002. 63:11.
35. Brodaty, H., et al., A randomized pla-

cebo-controlled trial of risperidone for the treatment of aggression, agitation, and psychosis of dementia. J Clin Psychiatry, 2003. 64(2): p. 134-43.
36. Ballard, C., et al., Quetiapine and rivastigmine and cognitive decline in Alzheimer's disease: randomised double blind placebo controlled trial. BMJ, 2005. 330(7496): p. 874.
37. Zhong, K.X., et al., Quetiapine to treat agitation in dementia: a randomized, double-blind, placebo-controlled study. Curr Alzheimer Res, 2007. 4(1): p. 81-93.
38. Satterlee, W., et al., A clinical update on olanzapine treatment in schizophrenia and in elderly Alzheimer's disease patients. Psychopharmacol Bull, 1995. 31(3): p. 534.
39. De Deyn, P.P., et al., Olanzapine versus placebo in the treatment of psychosis with or without associated behavioral disturbances in patients with Alzheimer's disease. Int J Geriatr Psychiatry, 2004. 19(2): p. 115-26.
40. Deberdt, W.G., et al., Comparison of olanzapine and risperidone in the treatment of psychosis and associated behavioral disturbances in patients with dementia. Am J Geriatr Psychiatry, 2005. 13(8): p. 722-30.
41. De Deyn, P., et al., Aripiprazole for the treatment of psychosis in patients with Alzheimer's disease: a randomized, placebo-controlled study. J Clin Psychopharmacol, 2005. 25(5): p. 463-7.
42. Mintzer, J., et al., Risperidone in the treatment of psychosis of Alzheimer disease: results from a prospective clinical trial. Am J Geriatr Psychiatry, 2006. 14(3): p. 280-91.
43. Mintzer, J.E., et al., Aripiprazole for the treatment of psychoses in institutionalized patients with Alzheimer dementia: a multicenter, randomized, double-blind, placebo-controlled assessment of three fixed doses. Am J Geriatr Psychiatry, 2007. 15(11): p. 918-31.
44. Streim, J.E., et al., A randomized, double-blind, placebo-controlled study of aripiprazole for the treatment of psychosis in nursing home patients with Alzheimer disease. Am J Geriatr Psychiatry, 2008. 16(7): p. 537-50.
45. Katz, I.R., et al., Comparison of risperidone and placebo for psychosis and behavioral disturbances associated with dementia: a randomized, double-blind trial. Risperidone Study Group. J Clin Psychiatry, 1999. 60(2): p. 107-15.
46. Street, J.S., et al., Olanzapine treatment of psychotic and behavioral symptoms in patients with Alzheimer disease in nursing care facilities: a double-blind, randomized, placebo-controlled trial. The HGEU Study Group. Arch Gen Psychiatry, 2000. 57(10): p. 968-76.
47. Sultzer, D.L., et al., Clinical symptom responses to atypical antipsychotic medications in Alzheimer's disease: phase 1 outcomes from the CATIE-AD effectiveness trial. Am J Psychiatry, 2008. 165(7): p. 844-54.
48. Paleacu, D., et al., Quetiapine treatment for behavioural and psychological symptoms of dementia in Alzheimer's disease patients: a 6-week, double-blind, placebo-controlled study.

Int J Geriatr Psychiatry, 2008. 23(4): p. 393-400.
49. Ballard, C., et al., A double-blind randomized placebo-controlled withdrawal trial comparing memantine and antipsychotics for the long-term treatment of function and neuropsychiatric symptoms in people with Alzheimer's disease (MAIN-AD). J Am Med Dir Assoc, 2015. 16(4): p. 316-22.
50. Holmes, C., et al., Risperidone and rivastigmine and agitated behaviour in severe Alzheimer's disease: a randomised double blind placebo controlled study. Int J Geriatr Psychiatry, 2007. 22(4): p. 380-1.
51. Meehan, K.M., et al., Comparison of rapidly acting intramuscular olanzapine, lorazepam, and placebo: a double-blind, randomized study in acutely agitated patients with dementia. Neuropsychopharmacology, 2002. 26(4): p. 494-504.
52. Kurlan, R., et al., Quetiapine for agitation or psychosis in patients with dementia and parkinsonism. Neurology, 2007. 68(17): p. 1356-63.
53. Rabey, J.M., et al., Effect of quetiapine in psychotic Parkinson's disease patients: a double-blind labeled study of 3 months' duration. Mov Disord, 2007. 22(3): p. 313-8.
54. Kennedy, J., et al., Olanzapine does not enhance cognition in non-agitated and non-psychotic patients with mild to moderate Alzheimer's dementia. Int J Geriatr Psychiatry, 2005. 20(11): p. 1020-7.
55. Breder, C., R. Swanink, and R. Marcus. Dose ranging study of aripiprazole in patients with Alzheimer's dementia [poster]. in 157th American Association Meeting. 2004.
56. Schneider, L.S., et al., National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE): Alzheimer disease trial methodology. Am J Geriatr Psychiatry, 2001. 9(4): p. 346-60.
57. Schneider, L.S., et al., Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE): Alzheimer's disease trial. Schizophr Bull, 2003. 29(1): p. 57-72.
58. Schneider, L.S., et al., Effectiveness of atypical antipsychotic drugs in patients with Alzheimer's disease. N Engl J Med, 2006. 355(15): p. 1525-38.
59. Zheng, L., et al., Metabolic changes associated with second-generation antipsychotic use in Alzheimer's disease patients: the CATIE-AD study. Am J Psychiatry, 2009. 166(5): p. 583-90.
60. ILO-522 [unpublished]. 2002.
61. Ris-int-83 [unpublished]. 1997.
62. ZIP-128-105 [unpublished]. 1993.
63. Ris-bel-14 [unpublished]. 1997.
64. Rosen, W.G., R.C. Mohs, and K.L. Davis, A new rating scale for Alzheimer's disease. The American Journal of Psychiatry, 1984. 141(11): p. 1356-1364.
65. Overall, J.E. and H.M. Rhoades, Clinician-rated scales for multidimensional assessment of psychopathology in the elderly. Psychopharmacol Bull, 1988. 24(4): p. 587-94.
66. Mack, J.L., M.B. Patterson, and P.N. Tariot, Behavior Rating Scale for Dementia: development of test scales and presentation of data for 555 individuals with Alzheimer's disease. J

Geriatr Psychiatry Neurol, 1999. 12(4): p. 211-23.
67. Devanand, D.P., et al., Behavioral syndromes in Alzheimer's disease. Int Psychogeriatr, 1992. 4 Suppl 2: p. 161-84.
68. Cohen-Mansfield, J., M.S. Marx, and A.S. Rosenthal, A description of agitation in a nursing home. J Gerontol, 1989. 44(3): p. M77-84.
69. Ellsworth, The MACC behavioral adjustment scale, Form II 1962, Western Psychological Services: Beverly Hills
70. Helmes, E., Multidimensional Observation Scale for Elderly Subjects (MOSES). Psychopharmacol Bull, 1988. 24(4): p. 733-45.
71. Wood, S., et al., The use of the neuropsychiatric inventory in nursing home residents. Characterization and measurement. Am J Geriatr Psychiatry, 2000. 8(1): p. 75-83.
72. Levin HS, H.W., Goethe KE, Sisson RA, Overall JE, Rhoades HM, Eisenberg HM, Kalisky Z, Gary HE., The neurobehavioural rating scale: assessment of the behavioural sequelae of head injury by the clinician. J Neurol Neurosurg Psychiatry, 1987. 50(2): p. 183-93.
73. Lindenmayer, J.P., et al., An excitement subscale of the Positive and Negative Syndrome Scale. Schizophr Res, 2004. 68(2-3): p. 331-7.
74. Endicott, J. and R.L. Spitzer, A diagnostic interview: the schedule for affective disorders and schizophrenia. Arch Gen Psychiatry, 1978. 35(7): p. 837-44.

Appendix

Appendix A. Exclusion criteria with regard to psychiatric disorders including substance abuse

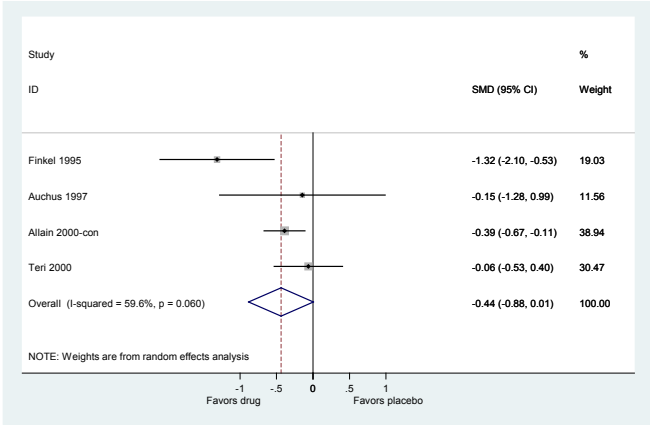
Publication	Exclusion criteria with regard to psychiatric disorders including substance abuse
Conventional antipsychotics	
Patients with agitation	
Finkel 1995	schizophrenia, major depression
Auchus 1997	schizophrenia, schizoaffective disorder, DSM-III-R criteria for major depressive episode or manic episode
Allain 2000	psychiatric disorders such as psychosis, depression
Teri 2000	major psychiatric disorders, delirium, alcohol or drug abuse
Patients with psychosis	
Hamilton 1962	not reported
Tariot 2006	concurrent other Axis I DSM–IV diagnosis
Patients with any NPS	
Sugerman 1964	not reported
Rada 1976	epilepsy, schizophrenia, depressive pseudodementia, pseudosenility
Barnes 1982	schizophrenia
Petrie 1982	schizophrenia
Stotsky 1984	psychosis
Devanand 1998	drug or alcohol dependence
De Deyn 1999	psychiatric disorders
Pollock 2002	delirium, schizophrenia, bipolar disorder, major depressive disorder
Atypical antipsychotics	
Patients with agitation	
Allain 2000	psychiatric disorders such as psychosis, depression
Herz 2002	not reported
Brodaty 2003	major depression, psychiatric disorders that could have accounted for psychotic disturbances
Ballard 2005	not reported
Zhong 2007	schizophrenia, schizoaffective disorder, bipolar disorder, agitation not related to dementia
Patients with psychosis	
Satterlee 1995	not reported
De Deyn 2004	primary mood disorder, other Axis I disorder e.g. schizophrenia, bipolar disorder, delusional disorder
Deberdt 2005	not reported
De Deyn 2005	Axis I DSM–IV diagnosis of delirium, amnesic disorders, bipolar disorder, schizophrenia, schizoaffective disorder; mood disorder with psychotic features; psychotic symptoms due to general medical condition or physiologic substance effects
Mintzer 2006	psychiatric disorders that produce psychotic symptoms; epilepsy
Tariot 2006	Axis I DSM–IV diagnosis

table continues

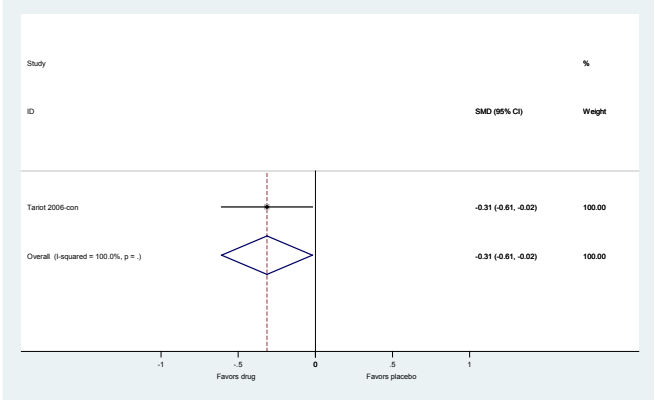
Publication	Exclusion criteria with regard to psychiatric disorders including substance abuse
Mintzer 2007	Axis I diagnosis of delirium, amnestic disorder, bipolar disorder, schizophrenia, schizoaffective disorder, mood disorder with psychotic features; major depressive episode with psychotic symptoms; seizure disorders, suicidal ideation or history
Streim 2008	Axis I diagnosis of delirium or schizophrenia; a schizoaffective, mood, bipolar, or amnestic disorder; continuous symptoms of psychosis before onset of dementia; major depression with symptoms of psychosis; at risk of suicide; substance use disorder according to DSM–IV criteria
Patients with any NPS	
Ris-bel-14 1997	psychiatric diagnosis
De Deyn 1999	psychiatric disorders
Katz 1999	delirium or amnestic disorder, psychiatric diagnosis causing psychotic disturbances
Street 2000	Axis I DSM-IV disorder (e.g. schizophrenia, bipolar disorder, depression), non-dementia related psychosis
Sultzer 2008	schizophrenia, schizoaffective disorder, delusional disorder, mood disorder with psychotic features, delirium, in need of psychiatric admission, suicidal
Paleacu 2008	alcohol or drug abuse

DSM: Diagnostic and Statistical Manual of Mental Disorder

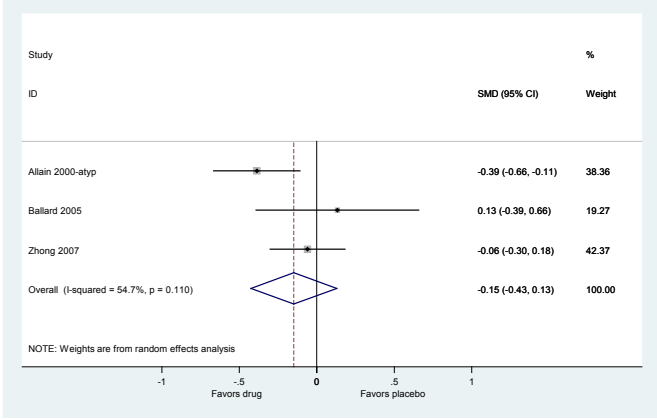
Appendix B. Efficacy of conventional and atypical antipsychotics in patients with agitation on agitation outcomes, and in patients with psychosis on psychosis outcomes.



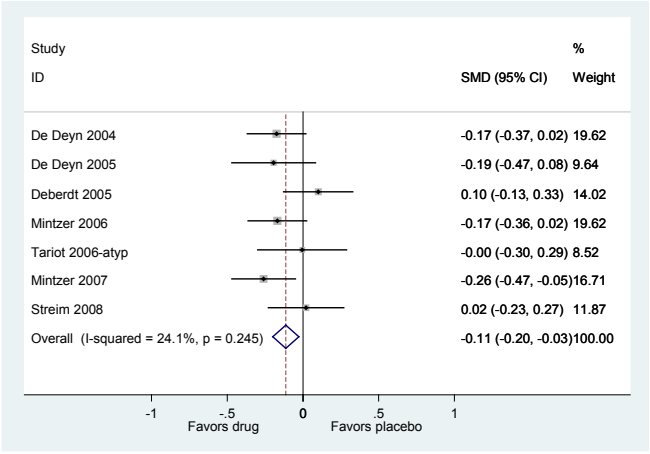
Conventional antipsychotics in patients with agitation on agitation outcomes



Conventional antipsychotics in patients with psychosis on psychosis outcomes

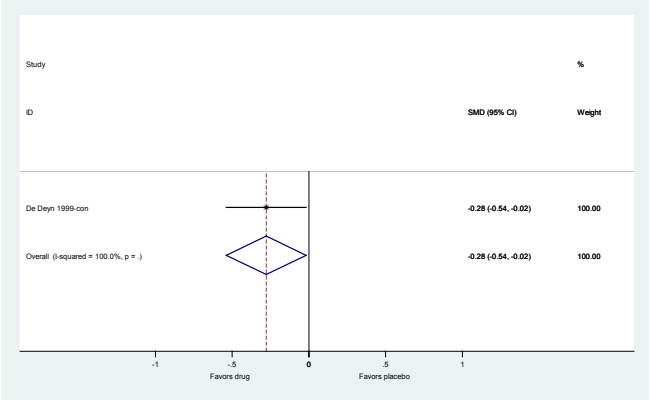


Atypical antipsychotics in patients with agitation on agitation outcomes

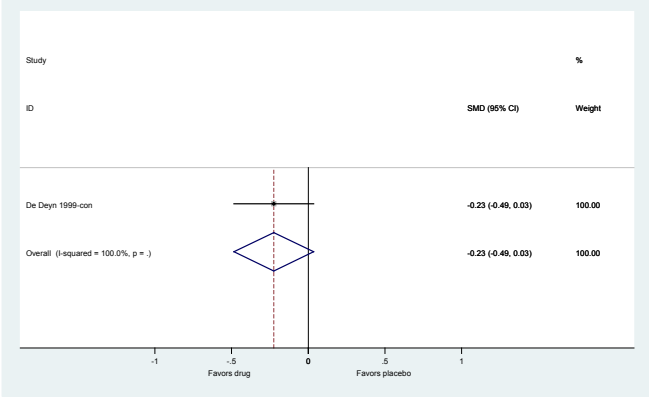


Atypical antipsychotics in patients with psychosis on psychosis outcomes

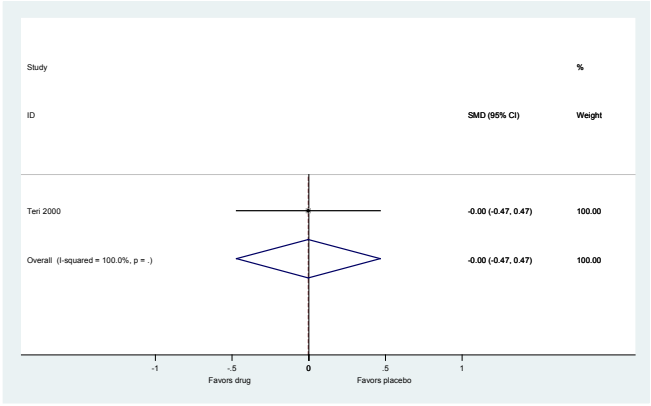
Appendix C. Efficacy of conventional antipsychotics when assessed within patients with any type of NPS, and with generic outcomes.



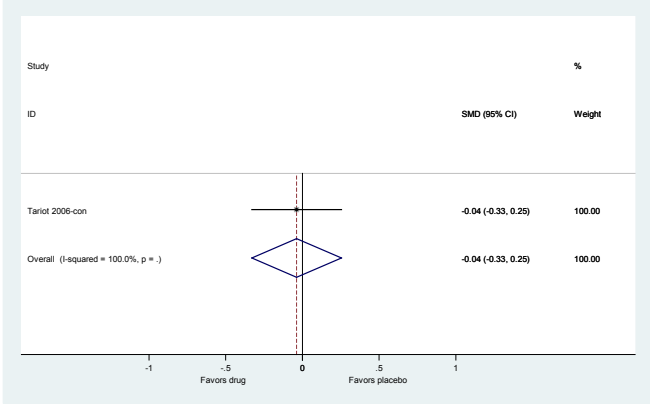
Conventional antipsychotics in patients with any type of NPS on agitation outcomes



Conventional antipsychotics in patients with any type of NPS on psychosis outcomes

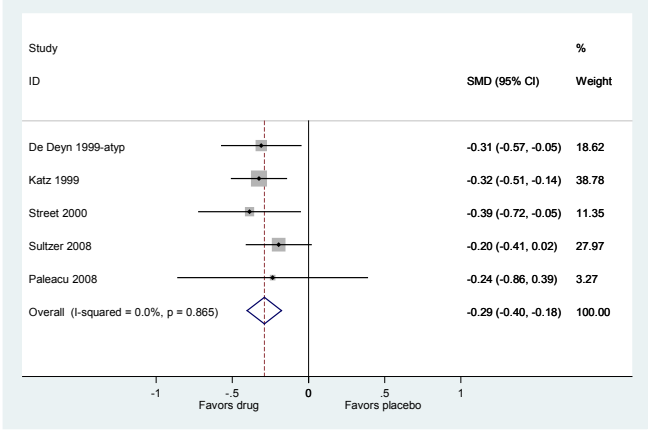


Conventional antipsychotics in patients with agitation on generic outcomes

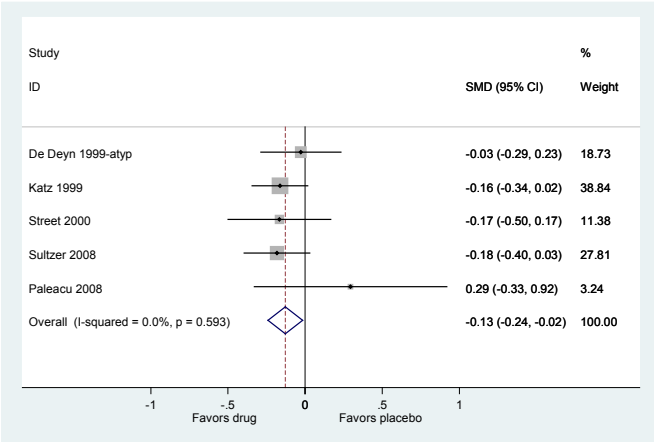


Conventional antipsychotics in patients with psychosis on generic outcomes

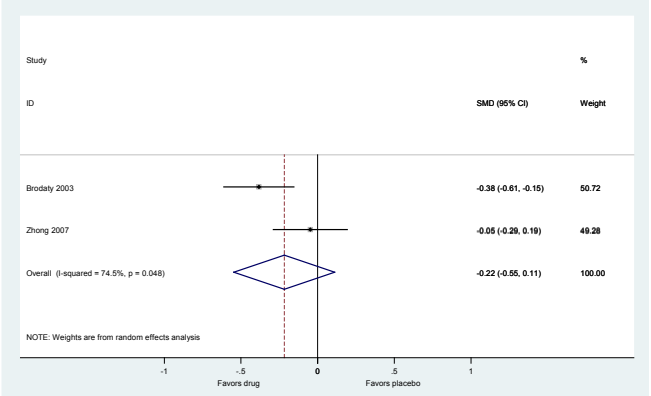
Appendix D. Efficacy of atypical antipsychotics when assessed within patients with any type of NPS, and with generic outcomes.



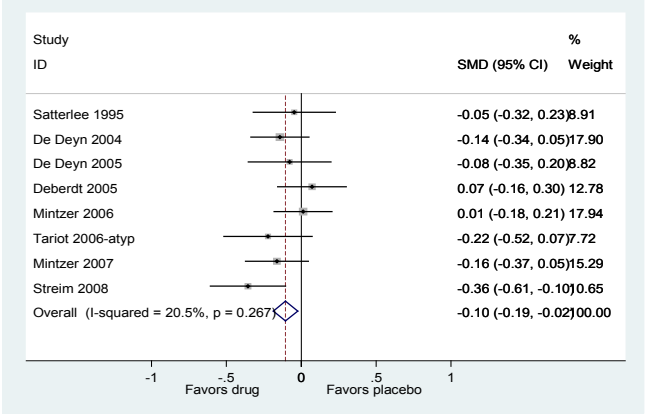
Atypical antipsychotics in patients with any type of NPS on agitation outcomes



Atypical antipsychotics in patients with any type of NPS on psychosis outcomes



Atypical antipsychotics in patients with agitation on generic outcomes



Atypical antipsychotics in patients with psychosis on generic outcomes

Chapter 4.

Large sample size fallacy in trials about antipsychotics for neuropsychiatric symptoms in dementia

Front Pharmacol. 2020 Feb 21;10:1701

Hulshof, T.A.
Zuidema, S.U.
Janus, S.I.M.
Luijendijk, H.J.

Abstract

Background

Atypical antipsychotics for neuropsychiatric symptoms in dementia have been tested in much larger trials than the older conventional drugs. The advantage of larger sample sizes is that negative findings become less likely and the effect estimates more precise. However, as sample sizes increase, the trials also get more expensive and time consuming while exposing more patients to drugs with unknown safety profiles. Moreover, a large sample size might yield a statistically significant effect that is not necessarily clinically relevant.

Objective

To assess (1) the variation in sample size and sample size calculations of antipsychotic trials in dementia, (2) the size of reported treatment effects and related statistical significance, and (3) general study characteristics that might be related to sample size.

Study Design and Setting

We performed a meta-epidemiological study of randomized trials that tested antipsychotics for neuropsychiatric symptoms in dementia. The trials compared conventional or atypical antipsychotics with placebo or another antipsychotic. Two reviewers independently extracted sample size, sample size calculations, reported treatment effects with p-values, and general study characteristics (drug type, trial duration, type of funding). We calculated a reference sample size of 83 and 433 per study group for the placebo-controlled and head-to-head trials respectively.

Results

We identified 33 placebo-controlled trials, and 18 head-to-head trials. Only 14 (42%) and 2 (11%) respectively reported a sample size calculation. The average sample size per arm was 34 (range 6-179) in placebo-controlled trials testing conventional drugs, 107 (8-237) in such trials testing atypical drugs, and 104 (95-115) in such trials testing both drug types; it was 31 (10-88) in head-to-head trials. Thirteen out of 18 trials with sample sizes larger than required (72%) reported a statistically significant treatment effect, of which two (15%) were clinically relevant. None of the head-to-head trials reported a statistically significant treatment effect, even though some suggested non-inferiority. In placebo-controlled trials of atypical drugs, longer trial duration (>6 weeks) and commercial funding were associated with higher sample size.

Conclusion

Sample size calculations were poorly reported in antipsychotic trials for dementia. Placebo-controlled trials of atypical antipsychotics showed large sample size fallacy while head-to-head trials were massively underpowered.

Introduction

Over the years the sample sizes of antipsychotic trials in dementia have increased from as low as 18 in the 1960s to as high as 652 in the 1990s.[1–3] The increase in sample sizes is generally viewed as a favourable development. Larger sample sizes provide more power to identify a treatment effect that is really present. In addition, the effect is estimated more precisely (smaller confidence intervals). Larger trials are also a natural consequence of head-to-head trials because the difference between two active drugs is generally expected to be small, and therefore, the required sample size needs to be relatively high.

However, larger sample sizes also make trials expensive and time consuming.[3] This can be barrier for non-commercial investigators to perform a trial. Moreover, it can be ethically questionable to ask more patients to participate, especially when the safety of the tested drug has not yet been established.[4] Another disadvantage of (very) large sample size is that a difference in outcomes between the groups will become (very) statistically significant, no matter how small or clinically meaningless it is.[5] If such results are nevertheless interpreted as clinically relevant, the 'large sample size fallacy' occurs.[6]

Sample size calculations for trials are based on four parameters if the response rate is the outcome. These are alpha, beta, the expected response rate in the active treatment groups, and the expected response rate in the comparison group (e.g. placebo).[7] Alpha is the probability of identifying a treatment effect that is not really present, which is usually set at 5%. Beta is the risk of not identifying a treatment effect that is really present, and is usually set at 20%. Sample size calculations for trials with continuous outcomes, such as the reduction of neuropsychiatric symptoms (NPS), are based on alpha, beta, the expected (difference between) means in the active and comparison group, and the population variance around the mean. Furthermore, the expected number of participants dropping out should be taken into account when determining the final target sample size of a trial.

A different expected treatment effect might explain why the sample sizes of antipsychotic trials increased over time. Perhaps, atypical antipsychotics were expected to be less effective than conventional antipsychotics, even before it was shown in systematic reviews that they did not affect psychotic symptoms compared to placebo.[8,9] Alternatively, drop-out could have increased because recent trials lasted longer and participants have become more assertive.

On the other hand, general study characteristics, which are not directly related to sample size calculation might have contributed to the increase in trial sample

sizes over the years. Large sample size is generally considered a sign of high trial quality and this increases the probability of publication and citation.[10] In addition, pharmaceutical companies will have more resources to fund larger trials than non-commercial organizations. Therefore, the aim of this meta-epidemiological study was to assess (1) the variation in sample size and sample size calculations of antipsychotic trials in dementia, (2) the size of the reported treatment effects and related statistical significance, and (3) general study characteristics that might be related to sample size.

Methods

Search strategy

Two reviewers (TAH, HJL) used a list of conventional and atypical antipsychotics from the websites of the World Health Organization, Food and Drug Administration and Wikipedia to search the literature.[11–13] First, we searched for studies in the electronic databases PubMed, Cinahl, Embase and Cochrane library with the string 'generic name of atypical/conventional antipsychotic' and trial and dementia (see appendix). We restricted the position of the drug name to title and abstract. Subsequently, we manually searched the references of published systematic reviews, which were identified with the same electronic databases. Titles and abstracts of potentially eligible studies were retrieved from PubMed. In addition, we sought trials in trial registration websites with the abovementioned search terms if possible; otherwise we used only the term dementia. These three searches were last re-run in June 2019. Finally, we had used the databases of the Dutch Medicines Evaluation Board and the FDA to find unpublished trials as part of a previous search performed in 2015.[14]

Study selection

We screened the title and abstract of the hits. Full texts of potentially eligible published studies and online protocols for unpublished studies were retrieved. Two reviewers used the full texts to determine definitive eligibility (TAH, HJL). The selected trials had to have been randomized and double-blind. They should have tested the efficacy of antipsychotics on neuropsychiatric symptoms in persons diagnosed with Alzheimer or vascular dementia. The trial had to compare conventional or atypical antipsychotics with placebo or another antipsychotic (head-to-head trial). We excluded studies with multiple drugs in a single intervention arm, studies that were stopped early and thus did not reach the targeted sample size, and studies with a cross-over design as other than standard sample size calculations need to be applied for this design. There were no restrictions with respect to publication date, language and duration of the study.

Data extraction

Two reviewers (TAH or SIMJ and HJL) independently extracted the following general study characteristics besides the sample size from the included studies: placebo-controlled or head-to-head trial, type of dementia (Alzheimer's disease, vascular dementia, mixed, unspecified), type of NPS (agitation, psychosis, diverse), setting (nursing home, hospital, outpatient clinic), active drug tested (conventional, atypical, or both), trial duration, type of funding (not-for-profit or commercial), and whether a sample size calculation was reported.

If the sample size calculation was reported, we extracted the input for sample size calculations: alpha, beta, expected treatment effects in the comparison groups (response rate, or mean symptom reduction with population variance at endpoint), and the expected drop-out rates. For trials that had been published in an abstract or online trial registration only, this data-extraction was considered inapplicable.

In addition, we extracted the reported treatment effects and related statistical significance. The primary outcome of trials that test antipsychotics for NPS in dementia is most often the difference in response rate or difference in reduction of target symptoms between the treatment groups. We extracted both for each trial with the related p-value. For the response rate, we extracted the number of patients with a clinically relevant improvement as defined by the authors. For reduction in symptoms, we extracted the difference in mean change from baseline to endpoint as measured with a symptom scale, such as the Cohen-Mansfield Agitation Inventory (CMAI) for agitation and Neuropsychiatric Inventory-Nursing Home (NPI-NH) for mixed symptoms. Initially, we also set out to extract standard deviations to calculate standardized mean differences, so that we could compare trial results. However, as many SDs turned out to be missing, we decided to extract the mean on the symptom scale at baseline as a reference instead (see data-analysis).

The primary source of extracted data was the published main results article. If that was not available, then conference abstracts or online published results were used. We received the individual patient data of two trials,[8,15] and additional meta-data of two others for use in another study.[14,16,17]

Data analyses

First, we described the variation in sample sizes for the different types of trials by plotting the mean number of participants per comparison group against the publication year of the trial. We present these data for the conventional and atypical placebo-controlled trials and head-to head trials separately.

To assess the adequacy of the reported sample sizes, we calculated reference sample sizes for trials with the response rate as outcome. For the placebo-controlled trials, we used an alpha of 0.05, beta of 0.20, a treatment response rate in the antipsychotic group of 55% and in the placebo group of 30%, and an expected drop-out of 30%.[18] A treatment effect of 25% (NNT =4) and drop-out rate of 30% is in line with previous literature and the reported response rates in antipsychotic trials in dementia.[8,19,20] We used a conservative drop-out rate of 30% (it was 26% on average in the included trials), so that the reference sample size would not be an underestimation. The required sample size per study group was 58 without loss to drop-out, and 83 with loss.

For the head-to-head trials (no placebo group), we used a treatment effect of 55% for the drug of interest and 45% for the control antipsychotic drug, because a 10% difference seems the upper limit of no difference. The expected drop-out rate was set at 10%, which is in line with the average drop-out rate in the included head-to-head trials. The required sample size was 389 per group without loss, and 433 with loss. We used the ssi command in Stata version 15.0 to calculate the reference sample sizes.[21]

To calculate reference sample sizes based on the outcome mean symptom reduction, the minimal clinically important difference (MCID) is required. However, the MCID is not known for most symptom scales used in this field.[22] The exception is the NPI, which was found to have an MCID of at least 8.0.[23,24] Nine of the included placebo-controlled trials in our study used this instrument, and we used the reported data to check our calculated reference sample size based on response rates. The reported mean reduction in symptoms was 19 (SD 14) for the placebo group (see appendix table 1), and hence, assuming an MCID of 8.0, 27 (SD 16) for the antipsychotic group. We calculated a required sample size of 80 based on these data, and this finding confirms the reference sample size of 83 based on response rates. In addition, the MCID of 8.0 reflects an SMD of 0.500 given the SD of 16 reported in the included trials. This is in line with the lower limit for a visible (medium) treatment effect suggested by Cohen.[25]

The next step was to assess whether studies with larger sample size reported statistically significant treatment effects that were not clinically relevant (difference in response rate <25%; difference in symptom reduction < MCID or SMD <0.5), which would suggest the presence of large sample size fallacy. Treatment effects in terms of reported response rates can be compared between trials with varying sample sizes. However, it was not possible to use MCIDs or SMDs to compare reported reductions in symptoms across different symptoms scales. Therefore, we calculated the relative symptom reduction as the ratio of the difference in symptom reduction between the study groups relative to the baseline mean in the groups. This approach has been used before.[26] Moreover, the MCID of 8.0 on the NPI and a mean baseline

of 39 (see appendix table 1) would translate into a relative symptom reduction of 21%. Hence, a relative symptom reduction of >=20% seems appropriate.

Finally, we analysed the association between other general study characteristics and mean sample size per group. The characteristics were type of drug tested (category: conventional, atypical, or both), trial duration (<= 6 weeks, > 6 weeks), and type of funding (non-for-profit, commercial). We calculated mean sample sizes of comparison groups per category, and used the two-sample t-test to determine whether the means differed between the first (reference) category and other categories. The analyses were performed for the placebo-controlled and head-to-head trials separately. All analyses were carried out with Stata version 15.0.[21]

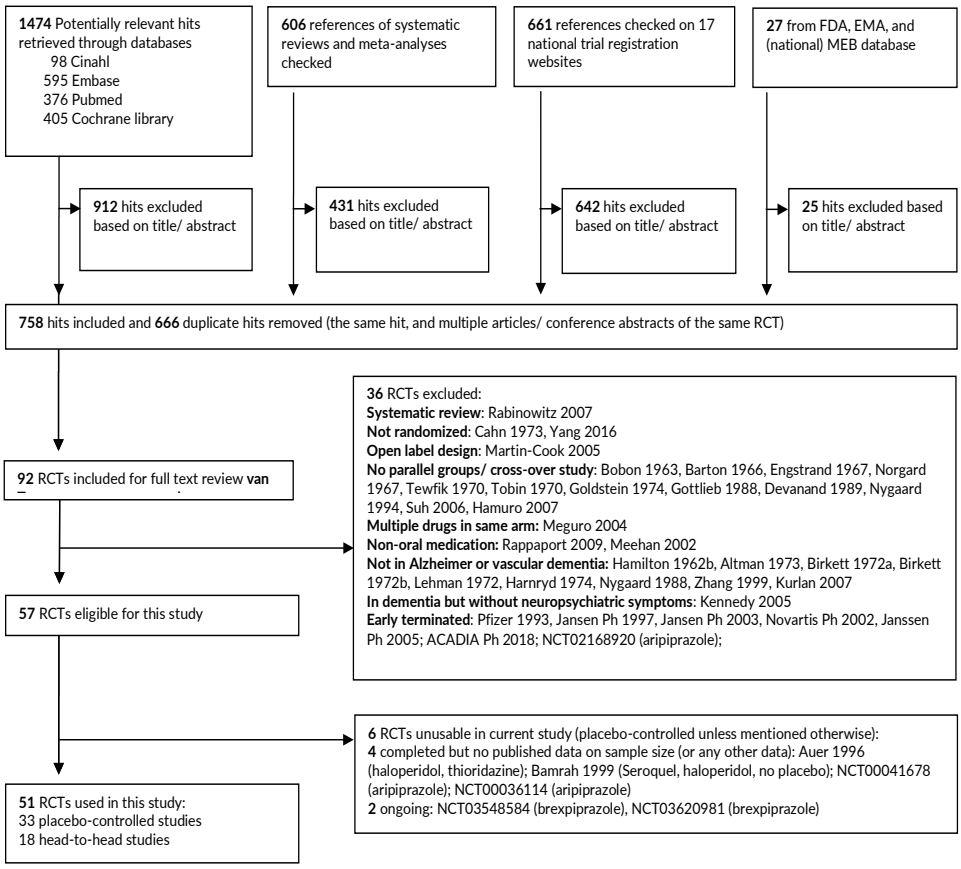
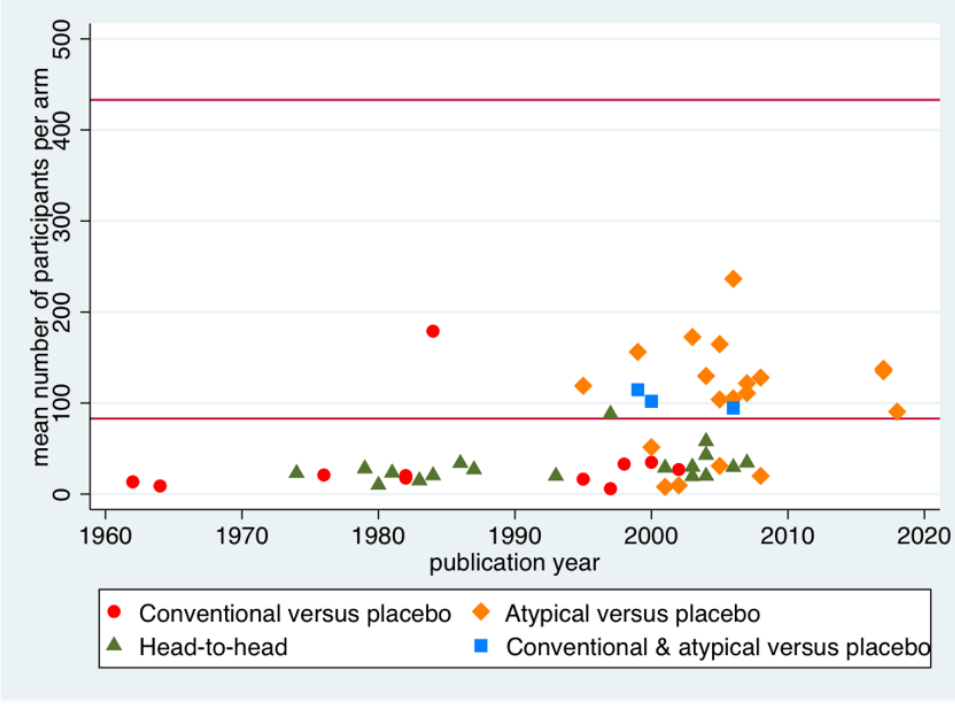


Figure 1. Flow diagram of literature and study selection

Results

Our search yielded 2768 potentially relevant hits (Figure 1). We obtained the reports of 92 studies for full text review. We considered 57 studies eligible but 6 had no useable data at the time of assessment. Hence, we used 51 studies in the current study.[15–17,26–66] Online or other clinical trial reports of the following studies were used: NCT00287742, NCT01862640, NCT01922258, NCT02992132, ZIP-128-105, RIS-BEL-14, RIS-INT-83.



* Red lines indicate calculated reference sample sizes per arm for the placebo-controlled trials (N=83) and head-to-head trials (N=433).

Figure 2. Scatter plot of sample size per arm over the years per treatment group.

Table 1 shows the general study characteristics. Eleven trials compared conventional antipsychotics to placebo and 19 trials atypical antipsychotics to placebo. Six of the latter 19 trials tested multiple doses of one atypical drug, so they had more than one drug group (range 2-4). Three placebo-controlled trials tested both conventional and atypical antipsychotics. Eighteen trials compared an antipsychotic drug with another antipsychotic drug. The studies were performed in outpatients, nursing homes or hospitals. The target symptom for treatment consisted of agitation, psychosis, or diverse neuropsychiatric symptoms.

Sample size variation and calculations

Figure 2 shows the mean number of participants per comparison group in each trial against publication year. The symbols indicate the type of drug tested (conventional, atypical, or both) and type of study (placebo-controlled or head-to-head). In the conventional antipsychotic placebo-controlled studies, the mean number per group was 34 patients (range 6-179), while those comparing atypical antipsychotics to placebo included on average 107 patients per group (range 8-237). The three trials that included both conventional and atypical antipsychotics and compared these to placebo included 104 patients per group (range 95-115). Head-to-head trials included a mean number of 31 patients per group (range 10-88). The increase in sample size over time seems to be related to type of drug tested.

We calculated a reference sample size of 83 patients per group for the placebo-controlled trials and 433 patients for the head-to-head trials, as explained above. The group sample size was lower than the reference sample size in 10 placebo-controlled trials of conventional antipsychotics (small sample size) and higher in one such trial (large sample size), whereas 5 of the 19 atypical antipsychotic trials and none of the 3 trials including both conventional and atypical antipsychotics had small sample sizes. At least 4 of the 5 atypical underpowered antipsychotic trials were investigator initiated, although one was performed with commercially acquired funds. All head-to-head trials had a small sample size that was lower than the reference sample size of 433.

Sixteen of 47 articles (excluding 2 abstracts and 2 reports on online trial registers) reported a sample size calculation (34%), which was often called a power analysis (table 1). Fourteen were placebo-controlled trials and two head-to-head trials (table 2). Table 2 shows, which input for these sample size calculations was reported. There were only four studies that reported sufficient information.[8,61,,67,68] Two studies reported an alpha that differed from 5% (2.5% and 7%). Eight studies reported a beta that differed from 20% and it varied between 1% and 15%. Except for the alpha of 2.5%, this input will yield higher sample sizes. Expected drop-out rates were reported in seven studies and varied between 10% and 30%.

There were seven placebo-controlled trials that postulated an expected treatment effect in terms of symptom reduction, four of which reflected a relative symptom reduction below 20%. The expected differences in relation to baseline means (relative symptom reduction) were: 10%;[61] 11%;[42] 12%;[59] 14%;[55] 20%;[67] 31%;[60] 31%.[68] For a head-to-head trial, the expected relative risk reduction was 16%.[27]

Table 1 Characteristics of randomized placebo-controlled and head-to-head trials of antipsychotics in patients with dementia

Study	Drug(s) studied	Type of dementia	Type of NPS (at least)
Antipsychotic versus placebo (33)			
Auchus and Cheryl Bissey-Black, 1997	Haloperidol	AD	Agitation
Howanitz and Wisotzek, 2001	Olanzapine	VAS	Diverse NPS
Sugerman et al., 1964	Haloperidol	CBS	Psychosis
Herz et al., 2002°	Risperidone, Olanzapine	AD	Agitation
Hamilton and Bennet, 1962	Trifluoperazine	CBS	Psychosis
Finkel et al., 1995	Thiothixene	NR	Agitation
Barnes et al., 1982	Loxapine, Thioridazine	NR	Diverse NPS
Petrie et al., 1982	Loxapine, Haloperidol	NR	Diverse NPS
Paleacu et al., 2008	Quetiapine	AD	Diverse NPS
Rada and Kellner, 1976	Thiothixene	CBS	Diverse NPS
Devanand et al., 1998	Haloperidol	AD	Diverse NPS
Ballard et al., 2005	Quetiapine	AD	Agitation
Pollock et al., 2002	Perphenazine	AD, VAS and MIX	Diverse NPS
Teri et al., 2000	Haloperidol	AD	Agitation
Street et al., 2000	Olanzapine	AD	Diverse NPS
Ballard et al., 2018	Pimavanserin	AD	Psychosis
Tariot et al., 2006	Quetiapine, Haloperidol	AD	Psychosis
Allain et al., 2000	Tiapride, Haloperidol	AD, VAS and MIX	Agitation
De Deyn et al., 2005	Aripiprazole	AD	Psychosis
Zhong et al., 2007	Quetiapine	AD and VAS	Agitation
Schneider et al., 2006	Olanzapine, Quetiapine, Risperidone	AD	Diverse NPS
De Deyn et al., 1999	Risperidone, Haloperidol	AD, VAS and MIX	Diverse NPS
Satterlee et al., 1995°	Olanzapine	AD	Diverse NPS
Mintzer et al., 2007	Aripiprazole	AD	Psychosis
Streim et al., 2008	Aripiprazole	AD	Psychosis
De Deyn et al., 2004	Olanzapine	AD	Psychosis
Otsuka Ph, 2017a†	Brexpiprazole	AD	Agitation
Otsuka Ph, 2017b	Brexpiprazole	AD	Agitation
Deberdt et al., 2005	Olanzapine, Risperidone	AD, VAS and MIX	Psychosis
Katz et al., 1999	Risperidone	AD, VAS and MIX	Diverse NPS
Brodaty et al., 2003	Risperidone	AD, VAS and MIX	Aggression
Stotsky, 1984	Thioridazine	NR	Diverse NPS
Mintzer et al., 2006	Risperidone	AD	Psychosis
Head-to-head trials (18)			
Vergara et al., 1980	Clomacran vs thioridazine	CBS	Diverse NPS
Spagnolo et al., 1983	Clomacran, thioridazine	VAS	Diverse NPS

Setting	N, total randomized	Duration, weeks	Sample size calculation reported	Commercial funding (drug of sponsor)
OUTP	12	6	-	- (non-commercial)
NR	16	6	- (abstract)	NR
HOS	18	6	-	+ (haloperidol)
NR	29	6	- (abstract)	NR
HOS	27	8	-	NR
NH	35	11	-	+ (thiothixene)
NH	60	8	-	+ (loxapine)
HOS	63	8	-	+ (loxapine)
NR	40	6	+	+ (quetiapine)
HOS	63	4	-	NR
OUTP	66	6	-	- (non-commercial)
NH	62	6	+	+ (commercial)#
NH	54	2,5	-	- (non-commercial)
HOS	70	16	+	+ (trazodone)
NH	206	6	+	+ (olanzapine)
NH	181	12*	+	+ (pimvaserin)
NH	284	10	+	+ (quetiapine)
NH-HOS	306	3	+	+ (tiapride)
OUTP	208	10	-	+ (aripiprazole)
NH	333	10	+	+ (quetiapine)
OUTP	421	12^	+	+ (olanzapine, quetiapine, risperidone)
NH	344	12	+	+ (risperidone)
NR	238	8	-	+ (olanzapine)
NH	487	10	-	+ (aripiprazole)
NH	265	10	-	+ (aripiprazole)
NH-HOS	652	10	+	+ (olanzapine)
NH	413	12	- (online)	+ (brexpiprazole)
NH-OUTP	270	12	- (online)	+ (brexpiprazole)
NH-OUTP	494	10	-	+ (olanzapine)
NH	625	12	+	+ (risperidone)
NH	345	12	+	+ (risperidone)
NH-HOS	358	4	-	NR
NH	473	8	+	+ (risperidone)
HOS	20	12	-	+ (clomacran)
HOS	30	3	-	NR

table continues

Study	Drug(s) studied	Type of dementia	Type of NPS (at least)
Fontaine et al., 2003	Etoperidone, thioridazine	NR	Agitation
Carlyle et al., 1993	Olanzapine, risperidone	AD, VAS and MIX	Aggression
Gareri et al., 2004	Loxapine, haloperidol	AD, VAS and MIX	Diverse NPS
Morris and Rickels, 1984	Risperidone, olanzapine, promazine	NR	Diverse NPS
Rosen, 1979	Loxapine, thioridazine	Organic cerebral disease#	Diverse NPS
Smith et al., 1974	Haloperidol, thioridazine	CBS	Psychosis
Götestam et al., 1981	Haloperidol, thioridazine	(Pre)senile and VAS	Diverse NPS
Lovett et al., 1987	Cis(Z)-clopenthixol, haloperidol	CBS	Psychosis
Chan et al., 2001	Trifluoperazine, haloperidol	AD, VAS and MIX	Diverse NPS
Verhey et al., 2006	Risperidone, haloperidol	NR	Agitation
Ather et al., 1986	Olanzapine, haloperidol	NR	Diverse NPS
Sheng et al., 2004	Chlormethiazole, thioridazine	AD and VAS	Diverse NPS
Rainer et al., 2007	Risperidone, haloperidol	AD, VAS, MIX, FTD	Diverse NPS
Mulsant et al., 2004	Quetiapine, risperidone	AD, VAS and MIX	Diverse NPS
Sun et al., 2004	Risperidone, olanzapine	Not Lewy Body	Diverse NPS
Gutzmann et al., 1997	Risperidone, haloperidol	NR	Restlessness

AD: Alzheimer's disease; CBS: chronic brain syndrome; HOS: hospital; MIX: mixed dementia (Alzheimer/ Vascular); NH: nursing home; NPS: neuropsychiatric symptoms; OUTP: outpatients; Ph: Pharmaceutical company; NR: not reported; VAS: vascular dementia.

° abstract only; * reduction in NPI Psychosis items at 12 weeks was the original primary outcome (clinicaltrials.gov); ^ Discontinuation rate at week 36 was the primary outcome, but as it is incomparable to other trials, we used response rate and reduction of symptoms at 12 weeks (see table 3); † results of 0.5mg group (n=20) were not reported; # the term senile brain disease was also used.

Setting	N, total randomized	Duration, weeks	Sample size calculation reported	Commercial funding (drug of sponsor)
NH	39	2	-	+ (olanzapine)
HOS	40	4	-	NR
NR	60	8	-	- (non-commercial)
NH	41	8	-	+ (loxapine)
OUTP	56	6	-	+ (haloperidol)
NH	46	6	-	NR
HOS	47	8	-	NR
NH	54	6	-	+ (trifluoperazine)
OUTP-HOS	58	12	-	+ (risperidone)
OUTP-NH	59	5	+	NR
NR	68	4	-	+ (chlormethiazole)
NR	60	8	-	+ (risperidone)
OUTP	68	8	+	+ (quetiapine)
NH	86	6	-	+ (risperidone)
HOS-OUTP	116	8	-	+ (risperidone)
HOS	176	4	-	+ (tiapride)

Reported treatment effects in relation to sample size

Table 3 presents the reported treatment effects in order of sample size per study group. A positive difference in response rate and negative difference in symptom reduction means that the investigated drug performed better than the control group. Six trials did not report what the effect of treatment on the primary outcome was: four studies were old, published between 1974-1983, but two were relatively new, published after 2000.[26,32,34,45,58,65] Five placebo-controlled studies reported only p-values without effect sizes in the abstract.[37,39,54,59,69]

Thirteen of 18 overpowered trials (72%) versus seven of 15 underpowered placebo-controlled trials (47%) yielded a statistically significant difference between the study groups in either response rate or symptom reduction. Two of 13 (15%) and four of seven (57%) of these treatment effects respectively were clinically relevant (difference in response rate >=25%, or relative symptom reduction >=20%). The statistically significant response rates were 10-22% and reported by studies with large sample sizes. The two studies with a difference in response rate of >=25%, which is the difference deemed clinically relevant,[25] were underpowered and did not report a statistically significant result. In addition, large sample size trials reported statistically significant relative symptom reductions between 10% and 23%, and small sample size trials reported statistically significant relative symptom reductions varying between 17% and 55%.

Many placebo-controlled trials had more than one intervention group, adding up to a total of 54 individual comparisons. Thirteen of the 33 overpowered comparisons (39%) from 18 trials yielded a statistically significant treatment effect on either response rate or symptom reduction, versus seven of the 21 underpowered comparisons (33%) from 15 trials.

Five of 18 head-to-head trials reported a difference in response rate of 10%, the lower limit that we set for non-inferiority in our reference sample size calculation, and four a relative symptom reduction of 10%. Yet, none of these results were statistically significant.

The reported treatment effect was lower than the expected treatment effect in the 14 studies that presented an expected treatment effect in a sample size calculation, except in two studies.[55,59] The reported drop-out rates varied between 6% and 37% (not shown), which was higher than the expected drop-out rate in most studies.

*Table 2 Input for sample size calculations**

Study	Alpha, %	Beta, %	Response rate or mean symptom change in drug group	Response rate or mean symptom change in control group	Difference in rates or means (SD) between groups¶	Expected dropout, %
Placebo-controlled trials						
Teri et al., 2000†	5	20	70%	30%	40%	NR
Katz et al., 1999	5	20	50%	30%	20%	NR
Street et al., 2000	5	20	NA	NA	-2.0 pts (NR)	NR
Brodaty et al., 2003	5	20	NA	NA	-4.15 pts (NR)	30
De Deyn et al., 2004	5	15	NA	NA	-3.0 pts (NR)	NR
Ballard et al., 2005	5	10	NA	NA	-6.0 pts (6)	25
Schneider et al., 2006	5	1^	27%#	60%#	-33%#	NA#
Mintzer et al., 2006	5	5	45%	25%	20%	20
Zhong et al., 2007	2.5	20	NR	NR	NR	10
Paleacu et al., 2008	7	10	NR	NR	-25% pts (NR)	NR
Ballard et al., 2018	5	10	NA	NA	-3.0 pts (6)	20
De Deyn et al., 1999	5	20	NR	NR	20%	20
Allain et al., 2000	5	20	55%	30%	25%	NR
Tariot et al., 2006	5	10	NA	NA	-4.5 pts (9)	NR
Head-to-head trials						
Verhey et al., 2006	5	10$	-14 pts	-2.8 pts	< -11.2 pts (NR)	25
Rainer et al., 2007	5	20	NR	NR	NR	NR

NA: not applicable; NR: not reported; pts: points (on instrument used to measure neuropsychiatric symptoms);

*this table presents the 16 studies that reported a sample size calculation ('power analysis') were included in this table; ¶ a difference in means needs to be accompanied by the population variance to calculate a sample size; †except for Teri 2000, all calculations were based on the comparison of the atypical antipsychotic group versus placebo; ^beta was reported to be 20% for a difference in rates of -20%; #discontinuation (not response) was the outcome; $text also mentions 20%.

Table 3 Results of randomized trials in order of group sample size

Study	Comparison groups	N per group	Reported effect in terms of response rate
			Definition/ measurement (bold if primary outcome)
Antipsychotic versus placebo (33)			
Auchus and Cheryl Bissey-Black, 1997	Haloperidol vs placebo	6 – 6	-
Howanitz and Wisotzek, 2001	Olanzapine vs placebo	8 – 8	-
Sugerman et al., 1964	Haloperidol vs placebo	9 – 9	**improvement on psychiatric observation**
Herz et al., 2002°	Risperidone vs placebo	14 – 8	-
	Olanzapine vs placebo	7 – 8	-
Hamilton and Bennet, 1962	Trifluoperazine vs placebo	18 – 9	**improvement on psychiatric observation**
Finkel et al., 1995	Thiothixene vs placebo	17 – 18	>5 points on CMAI
Barnes et al., 1982	Loxapine vs placebo	19 – 17	improvement on CGI
	Thioridazine vs placebo	17 – 17	
Petrie et al., 1982	Loxapine vs placebo	19 – 22	**>= moderate improvement on CGI**
	Haloperidol vs placebo	20 – 22	
Paleacu et al., 2008	Quetiapine vs placebo	20 – 20	Improved on CGIC
Rada and Kellner, 1976	Thiothixene vs placebo	22 – 20	improved on global rating
Devanand et al., 1998	Haloperidol 0.5-0.75mg vs placebo	21 – 24	>=25% reduction BPRS Psychosis items
	Haloperidol 2-3mg vs placebo	21 – 24	
Ballard et al., 2005	Quetiapine vs placebo	31 – 31	-
Pollock et al., 2002	Perphenazine vs placebo	33 – 21	-
Teri et al., 2000	Haloperidol vs placebo	34 – 36	**improvement on ADCS-CGIC**$
Street et al., 2000	Olanzapine 5mg vs placebo	56 – 47	-
	Olanzapine 10mg vs placebo	50 – 47	-
	Olanzapine 15mg vs placebo	53 – 47	-
Ballard et al., 2018¶	Pimavanserin vs placebo	90 – 91	>=30% decrease on NPI-NH Psychosis items
Tariot et al., 2006	Quetiapine vs placebo	91 – 99	>=30% decrease on BPRS
	Haloperidol vs placebo	94 – 99	
Allain et al., 2000	Tiapride vs placebo	102 – 103	**>=25% decrease on MOSES irritability/ aggression items**)
	Haloperidol vs placebo	101 – 103	
De Deyn et al., 2005	Aripiprazole vs placebo	106 – 102	improvement on CGI-I

Reported effect in terms of symptom reduction				
Difference between groups	p-value	Symptom scale (bold if primary outcome)	Difference between groups (baseline mean); relative symptom reduction	p-value
-	-	**CMAI**	-1.0 (35.2); 5%	.82
-	-	-	-	-
22%	nr	'symptom checklist'	-2.5 (nr); nr	nr
-	-	**BPRS Excitement**	Nr (nr); nr	ns
-	-		Nr (nr); nr	.0001
22%	nr	MACC	-0.7 (31.4); 2%	ns
51%	nr	**CMAI**	-9.0 (30.5); 55%	<.001
17%	ns	**BPRS**	-2.9 (45.8); 6%	ns
12%	ns		0.0 (45.8); 0%	ns
23%	nr	BPRS	-9.5 (47.9); 20%	<.05
26%	nr		-9.3 (47.9); 19%	<.05
-5%	ns	**NPI-NH**	-5.2 (41.0); 13%	ns
4%	ns	**BPRS**	Nr (nr;); nr	ns
0%	nr	**BPRS Psychosis**	0.0 (6.8); 0%	ns
30%	<0.06		-1.2 (6.8); 18%	<.03
-	-	**CMAI**	3.5 (57.7); 8%	.30
-	-	**NRS**	-4.9 (57.6); 9%	.14
1%	0.81	CMAI	-1.3 (49.2*); 3%	>.25
-	-	**NPI-NH Agitation +**	-3.9 (14.2); 27%	<.001
-	-	**Psychosis**	-2.4 (14.2); 17%	.006
-	-		-1.2 (14.2); 8%	.24
nr	nr	**NPI-NH Psychosis**	-0.5 (9.8); 5%	.561
11%	.265	BPRS	-2.3 (39.5); 6%	.217
7%	nr		-0.4 (39.5); 1%	.354
14%	.04	MOSES irritability/	-1.9 (20.3); 9%	.009
20%	.004	aggression	-2.1 (20.3); 10%	.005
8%	.18	**NPI Psychosis**	-1.03 (12.4); 8%	.017

table continues

Study	Comparison groups	N per group	Reported effect in terms of response rate
Zhong et al., 2007	Quetiapine 100mg vs placebo	124 – 92	moderate & marked
	Quetiapine 200mg vs placebo	117 – 92	improvement on CGI-C
Schneider et al., 2006	Olanzapine vs placebo	100 – 142	improvement on
	Quetiapine vs placebo	94 – 142	CGIC†
	Risperidone vs placebo	85 – 142	
De Deyn et al., 1999	Risperidone vs placebo	115 – 114	**>=30% decrease on**
	Haloperidol vs placebo	115 – 114	**BEHAVE-AD**
Satterlee et al., 1995°	Olanzapine vs placebo	120 – 118	-
Mintzer et al., 2007	Aripiprazole 2mg vs placebo	118 – 121	>=50% decrease
	Aripiprazole 5mg vs placebo	122 – 121	NPI-NH Psychosis
	Aripiprazole 10mg vs placebo	126 – 121	
Streim et al., 2008	Aripiprazole vs placebo	131 – 125	>=50% decr NPI-NH
De Deyn et al., 2004	Olanzapine 1mg vs placebo	129 – 129	-
	Olanzapine 2.5mg vs placebo	134 – 129	(CGI-C was
	Olanzapine 5mg vs placebo	125 – 129	administered)
	Olanzapine 7.5mg vs placebo	132 – 129	
Otsuka Ph, 2017a^	Brexpiprazole 1mg vs placebo	137 – 136	-
	Brexpiprazole 2mg vs placebo	140 – 136	
Otsuka Ph, 2017b	Brexpiprazole vs placebo	133 – 137	-
Deberdt et al., 2005	Olanzapine vs placebo	204 – 94	>=30% decr NPI-NH Psychosis
	Risperidone vs placebo	196 – 94	
Katz et al., 1999	Risperidone 0.5mg vs placebo	149 – 163	>= 50% reduction
	Risperidone 1mg vs placebo	148 – 163	on BEHAVE-AD
	Risperidone 2mg vs placebo	165 – 163	
Brodaty et al., 2003	Risperidone vs placebo	173 – 172	improvement on CGI-I
Stotsky, 1984	Thioridazine vs placebo	183 – 175	-
Mintzer et al., 2006	Risperidone vs placebo	235 – 238	improvement on CGI-C
Head-to-head trials (18)			
Vergara et al., 1980	Clomacran vs thioridazine	20 total	Improvement on CGI
Spagnolo et al., 1983	Etoperidone vs thioridazine	15 – 15	clinical judgement
Fontaine et al., 2003	Olanzapine vs risperidone	20 – 19	- (CGI-C was administered)
Carlyle et al., 1993	Loxapine vs haloperidol	20 – 20	Any decrease in weekly # of aggressive acts
Gareri et al., 2004	Risperidone vs promazine	20 – 20	**>=50% decrease**
	Olanzapine vs promazine	20 – 20	**on NPI**
Morris and Rickels, 1984	Loxapine vs thioridazine	21 – 20	global improvement

		Reported effect in terms of symptom reduction		
8%	ns	**PANSS Excitement**	-0.8 (23.0); 3%	.457
22%	.002		-2.7 (23.0); 12%	.014
11%	.05	NPI	-5.0 (36.9); 14%	nr
5%	.37		-7.6 (36.9); 21%	nr
8%	.21		-7.4 (36.9); 20%	nr
11%	.13	BEHAVE-AD	-2.4 (16.5); 15%	.05
8%	.25	BEHAVE-AD	-1.3 (16.5); 8%	nr
-	-	BEHAVE-AD	-0.4 (19.8); 2%	ns
5%	Ns	**NPI-NH Psychosis**	-0.5 (11.6**);** 4%	Ns
13%	ns		-1.2 (11.6**);** 10%	ns
15%	.019		-1.8 (11.6**);** 16%	.013
18%	.006	**NPI-NH Psychosis**	+0.1 (10.6); 1%	ns
-	-	**NPI-NH Psychosis**	-1.0 (9.7); 10%	.171
-	-		-0.8 (9.7); 8%	.089
-	-		-0.6 (9.7); 6%	.274
-	-		-1.2 (9.7); 12%	.032
-	-	**CMAI**	+0.2 (nr); nr	.902
			-3.8 (nr); nr	.040
-	-	**CMAI**	-2.4 (nr); nr	.145
-4%	Ns	**NPI Psychosis**	-0.7 (11.3); 6%	0.421
-3%	ns		-0.5 (11.3); 4%	0.585
nr	nr	**BEHAVE-AD**	-1.2 (15.8); 8%	.13
12%	.02		-2.2 (15.8); 14%	.02
17%	.002		-3.3 (15.8); 21%	<.001
22%	<.001	**CMAI aggression**	-4.4 (33.5); 23%	<.001
-	-	**Modified HAS**	-4.3 (nr); nr	<.001
10%	.019	**BEHAVE-AD Psychosis**	-0.6 (7.9); 8%	. 118
0%	na	**VTSRS**	nr (nr); nr	ns
0%	nr	**SHGRS**	nr (nr); nr	nr
Nr	ns	**NPI**	+8 (51.8); 15%	ns
15%	nr	**weekly # of aggressive acts**	-1.1 (6.9); 16%	ns
5%	nr		-	-
15%	nr		-	-
nr	nr	**BPRS**	+1.7 (63.6); 3%	ns

table continues

Study	Comparison groups	N per group	Reported effect in terms of response rate
Rosen, 1979	Haloperidol vs thioridazine	24 – 18	-
Smith et al., 1974	Haloperidol vs thioridazine	23 – 23	CGI
Götestam et al., 1981	Cis(Z)-clopenthixol vs haloperidol	25 – 22	improvement on CGI
Lovett et al., 1987	Trifluoperazine vs haloperidol	26 – 28	improvement on CGI
Chan et al., 2001	Risperidone vs haloperidol	29 – 29	-
Verhey et al., 2006	Olanzapine vs haloperidol	30 – 28	- (CGI was administered)
Ather et al., 1986	Chlormethiazole vs thioridazine	30 – 30	-
Sheng et al., 2004	Risperidone vs haloperidol	30 – 30	**improvement on CGI**
Rainer et al., 2007	Quetiapine vs risperidone	36 – 32	improvement on CGI
Mulsant et al., 2004	Risperidone vs olanzapine	42 – 43	-
Sun et al., 2004	Risperidone vs haloperidol	57 – 59	**>=30% decrease on BEHAVE-AD**
Gutzmann et al., 1997	Tiapride vs melperone	88 – 87	**improvement on CGI**

nr: not reported, ns: the effect was reported as not statistically significant but no p-value was given; ADCS-CGIC: Alzheimer's Disease Cooperative Study Clinical Global Impression of Change; BEHAVE-AD: Behavioural pathology in Alzheimer's disease scale; BPRS: Brief Psychiatric Rating Scale; CMAI: Cohen-Mansfield Agitation Inventory; CGBRS: Crichton Geriatric Behavioral Rating Scale; GCGRS: Gottfires-Cronholm Geriatric Rating Scale; MACC: Motility affect communication cooperation behavioral adjustment scale; NPI(-NH): Neuropsychiatric Inventory (-Nursing Home version); NRS: Neurobehavioral Rating Scale; PANSS: Positive and Negative Syndrome Scale; SHGRS: Stuard Hospital Geriatric Rating Scale; VTSRS: Verdun Target Symptom Rating Scale.
°abstract only; *49.2 is the weighted mean of baseline mean of all studies with CMAI total; ¶reduction in NPI Psychosis at 12 weeks was originally the primary outcome (clinicaltrials.gov); †Discontinuation rate at week 36 is primary outcome of trial, but as it is incomparable to other trials, we used response rate and reduction of symptoms at 12 weeks; ^results of 0.5mg group (n=20) were not reported.

		Reported effect in terms of symptom reduction		
-	-	**Modified BPRS**	+0.1 (3.2); 3%	ns
22%	nr	**BPRS**	nr (nr); 11%	.01
-6%	nr	**GCGRS**	-4.1 (26.9); 15%	<.05
18%	ns	**BPRS**	-1.2 (50.4); 2%	ns
-	-	**CMAI**	+2.0 (47.7); 4%	ns
-	-	**CMAI**	+6.5 (70); 9%	0.338
-	-	**CGBRS**	-1.9 (37.1); 5%	nr
10%	>.05	**BEHAVE-AD**	0 (15); 0%	>.05
-3.4%	nr	**NPI**	+2.2 (57.9); 4%	ns
-	-	**NPI**	nr (nr); nr	ns
1%	nr	**BEHAVE-AD**	+0.1 (17.5); 1%	ns
1%	.675	restlessness	-1.4 (56.2); 2%	ns

Table 4 Mean sample size by study characteristic

Study characteristic		Placebo-controlled trials		Head-to-head trials	
		n	Mean (SD)	n	Mean (SD)
Type of drug	Conventional antipsychotic (ref)	11	34.4 (48.8)	9	22.3 (7.1)
	Atypical antipsychotic	19	107.0 (60.5)^	4	46.3 (29.5)^
	Conventional and atypical antipsychotic	3	103.8 (94.7)^	3	33.3 (14.4)
Trial duration	=<6 weeks (ref)	11	28.9 (4.4)	10	32.7 (21.0)
	>6 weeks	22	109.2 (12.6)*	8	28.2 (14.2)
Type of funding	Non-commercial (ref)	7	18.1 (9.6)	5	21.6 (5.4)
	Commercial	24	100.3 (57.5)*	13	34.2 (20.0)

^ p <0.05 compared to reference group; * p < .001 compared to reference group.

Study characteristics and sample size

Table 4 shows the mean sample size per comparison group by type of drug tested, trial duration, and type of funding. The mean sample size per study group was statistically significantly higher in placebo-controlled trials that tested an atypical antipsychotic drug (107.0) or both a conventional and an atypical drug (103.8) in comparison to placebo-controlled trials of conventional antipsychotics (34.4; p<.05). The mean sample size per study group was also statistically significantly higher in trials that lasted more than 6 weeks (109.2) compared to less than 6 weeks (28.9; p < .001), and that were commercially (100.3) versus non-commercially (18.1; p < .001) funded. Head-to-head-trials that tested atypical drugs only had a significantly larger mean sample size (46.3) than trials that tested conventional drugs (22.3; p <.05). Trial duration and commercial funding did not seem to be related to the sample size of head-to-head trials.

Discussion

We assessed the presence of large sample size fallacy in 51 antipsychotic trials in dementia. Most placebo-controlled trials of conventional antipsychotics had small sample size, i.e. smaller than the calculated reference sample size, but most trials of atypical antipsychotics had large sample sizes. All head-to-head trials had very small sample sizes. Only one third of trials reported a sample size calculation. Thirteen of 18 trials with large sample sizes (72%) reported a statistically significant treatment effect, of which two (15%) were clinically relevant. In contrast, seven of 15 placebo-controlled trials with small sample sizes (47%) yielded a statistically significant treatment effect, and four were clinically relevant (57%). None of the head-to-head trials reported a statistically significant treatment effect, even though some suggested non-inferiority.

Large sample size fallacy

Sample sizes need to be large enough to guarantee a minimum level of discriminative power to detect a real treatment effect. Moreover, precision of an estimate increases with sample size. Studies based on small sample size may yield a non-statistically significant but clinically relevant treatment effect. On the other hand, studies based on large sample size – larger than necessary – may yield statistically significant but clinically insignificant treatment effects.[70,71] Large sample size fallacy occurs when such results are interpreted as relevant for medical practice.[6,72] Nevertheless, pharmaceutical companies and academic scholars benefit from statistically significant treatment results being interpreted as clinically relevant.[10] The emphasis on statistical significance was confirmed by six trials in our review that did not report effect sizes, and five trials that reported just p-values in the abstract.

The sample sizes of trials testing atypical antipsychotics versus placebo, whether or not simultaneously with a conventional antipsychotic, were generally larger than necessary. These trials were commercially funded by the manufacturer of the atypical antipsychotic drugs. Only investigator-initiated trials were too small. The majority of large trials reported a statistically significant treatment effect, despite lack of clinical relevance, which confirms the presence of large sample size fallacy. The mean sample size was also higher when the study lasted longer than 6 weeks and was commercially funded, but this might be explained by the fact that placebo-controlled trials of atypical antipsychotics were generally longer and often industry-initiated. The chance of statistically significant findings was further enhanced by the use of multiple comparisons per study and multiple measurement scales per outcome in a number of the larger trials.

Many placebo-controlled trials of conventional antipsychotics had small sample sizes. Most were relatively old (published before 1990) and seemed to be investigator-initiated. Some of these trials reported clinically relevant results, but most were not statistically significant. That small placebo-controlled trials yielded statistically significant and clinically relevant effects relatively often might reflect publication bias.

Head-to-head trials had sample sizes that were (much) smaller than required, and these studies yielded non-statistically significant results that sometimes suggested a substantial effect. Even if we had set the limit for non-inferiority at 15%, the required sample would have been a lot higher than the sample sizes of the included studies were (346 without loss, and 385 with loss). It is unclear why these trials were so clearly underpowered. Perhaps, industry has little to gain from properly testing their own product against that of competitors. Non-commercial funds might not

be interested in a trial with at least 2 x 433 patients to show that the tested drugs are non-inferior, even if patients might be quite willing to participate in a study that ensure treatment with an active drug.

Sample size requirements

It is generally agreed that a trial protocol and report should report a sample size calculation.[73] Nevertheless, only a third of trials in our review reported a sample size calculation and just three were complete. Although some trials can be considered old, most were published in the 1990s or later when it had become common to report trial methods in detail. Sample size calculations are often not (completely) reported in randomized trials in other fields of research was well.[70,74] One review found that articles about newer randomized controlled trials included sample size calculations more often, and showed positive results more often (76%) than older studies (55%).[75]

Some studies in our review reported a lower alpha (2.5%) or beta (5%) than is usual in sample size calculations (5% and 20% respectively). In addition, the MCID proposed in the sample size calculations seemed rather small: difference in response rates <25% in 3/6 trials, and in relative risk reduction of <20% in 4/7 trials. The lower the alpha, beta and MCID, the higher the calculated sample size will be and hence the power to detect a statistically significant but not clinically relevant treatment effect. Moreover, even if the expected difference is equal to the MCID, a proportion of the patients will not have a clinically relevant effect on the individual level. On the other hand, the expected drop-out rate in the sample size calculations was mostly lower than the (mean) reported drop-out, and this would have led to a spuriously smaller calculated sample size. Real drop-out might have been high because trial duration was long on average. Most trials lasted more than a month, even though in clinical practice, antipsychotics usually show an effect within 2 weeks, 4 at the most. It has been estimated that up to 64% of trials with continuous outcomes are underpowered or overpowered because of imprecise input.[76]

Strengths and limitations

To our knowledge determinants of sample size in trials testing antipsychotics for neuropsychiatric symptoms in dementia have not been studied previously. Our study showed that sample size calculations in the reports of these trials were missing on a large scale as was the correct interpretation of effect size. A limitation of our study is its focus on antipsychotic trials in dementia, which might be perceived as a small field of research. In addition, the interpretation of our results is limited by the possible presence of multiple testing. Many trials used multiple comparisons of either different drugs, different dosages, multiple outcomes, and sometimes multiple

measurement instruments per outcome. Such multiple testing might reinforce the large sample size fallacy.

With our study, we do not want to suggest that large sample sizes should be avoided. It is important for clinical practice that study results are precise. Moreover, large sample sizes are very useful for identification of adverse effects. Small trials should not be avoided either, as long as they are published irrespective of results and available for pooling in meta-analyses.

The implication of our study is that researchers need to be encouraged to report and consider effect sizes in line with p-values to avoid the large sample size fallacy. Journals should probably mention this in their author instructions.

Conclusion

Placebo-controlled trials that tested atypical antipsychotics showed large sample size fallacy. Placebo-controlled trials of conventional antipsychotics and head-to-head trials had insufficient power to detect a real difference between the treatment groups. Sample size calculations in antipsychotic trials for dementia need to be reported adequately.

References

1. Hulshof T, Zuidema S, Ostelo R, Luijendijk H. The Mortality Risk of Conventional Antipsychotics in Elderly Patients: A Systematic Review and Meta-analysis of Randomized Placebo-Controlled Trials. J Am Med Dir Assoc. 2015;16(10):817-824.
2. Schneider L, Dagerman K, Insel P. Risk of Death With Atypical Antipsychotic Drug Treatment for Dementia. J Am Med Assoc. 2005;294(15):1934-1943.
3. Cox Gad S. Trial Design. John Wiley. Hoboken, New Jersey; 2009.
4. Schipper I, Weyzig F. SOMO briefing paper on ethics in clinical trials: Examples of unethical trials. Somo Brief Pap ethics clin9cal Res. 2008;2008(February):1-16. doi:10.1002/micr
5. Sullivan G, Feinn R. Using Effect Size—or Why the P Value Is Not Enough. J Grad Med Educ. 2012;(September):279-282.
6. Lantz B. The large sample size fallacy. Scandanavian J Caring Sci. 2013;27(1):487-492.
7. Noordzij M, Tripepi G, Dekker FW, Zoccali C, Tanck MW, Jager KJ. Sample size calculations: Basic principles and common pitfalls. Nephrol Dial Transplant. 2010;25(5):1388-1393. doi:10.1093/ndt/gfp732
8. Schneider L, Dagerman K, Insel P. Efficacy and adverse effects of atypical antipsychotics for dementia: meta-analysis of randomized, placebo-controlled trials. Am J Geriatr Psychiatry. 2006;14(March):191-210.
9. Smeets CHW, Zuidema SU, Hulshof TA, et al. Efficacy of antipsychotics in dementia depended on the definition of patients and outcomes: a meta-epidemiological study. J Clin Epidemiol. 2018;101:17-27. doi:10.1016/j.jclinepi.2018.05.004
10. Dickersin K, Min YI, Meinert CL. Factors Influencing Publication of Research Results: Follow-up of Applications Submitted to Two Institutional Review Boards. J Am Med Assoc. 1992;267(3):374-378. doi:10.1001/jama.1992.03480030052036
11. World Health Organization. WHO Collaborating Centre for Drug Statistics Methodology. http://www.whocc.no/atc_ddd_index/?code=N05A. Published 2013.
12. US Food and Drug Administration. Atypical Antipsychotic Drugs Information. http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm094303.htm. Published 2013.
13. Wikipedia. Typical antipsychotic. Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Typical_antipsychotic. Published 2015. Accessed January 8, 2015.
14. Hulshof TA, Zuidema SU, van Meer PJK, Gispen-de Wied CC, Luijendijk HJ. Baseline imbalances and clinical outcomes of atypical antipsychotics in dementia: A meta-epidemiological study of randomized trials. Int J Methods Psychiatr Res. 2019;28(1):1-10.

doi:10.1002/mpr.1757
15. Paleacu D, Barak Y, Mirecky I, Mazeh D. Quetiapine treatmdnt for behavioural and psychological symptoms of dementia in alzheimer's disease patients: A 6-week, double-blind, placebo-controlled study. Int J Geriatr Psychiatry. 2008;23(4):393-400. doi:10.1002/gps.1892
16. De Deyn PP, Rabheru K, Rasmussen A, et al. A randomized trial of risperidone, placebo, and haloperidol for behavioral symptoms of dementia. Neurology. 1999;53(5):946-955. doi:10.1212/WNL.53.5.946
17. De Deyn P, Jeste D, Swanink R, Kostic D, Breder C. Aripiprazole for the Treatment of Psychosis in Patients With Alzheimer's Disease. J Clin Psychopharmacol. 2005;25(5):463-467. http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00004714-200510000-00010. Accessed April 17, 2014.
18. Brant R. Inference for Proportions: Comparing Two Independent Samples. http://www.stat.ubc.ca/~rollin/stats/ssize/b2.html. Published 2016. Accessed March 17, 2016.
19. Jeste D, Blazer D, Casey D, et al. ACNP White Paper: Update on Use of Antipsychotic Drugs in Elderly Persons with Dementia. Neuropsychopharmacology. 2008;33(5):957-970.
20. Drouillard N, Mithani A, Chan P. Therapeutic approaches in the management of behavioral and psychological symptoms of dementia in the elderly. BC Med J. 2013;55(2):90-95.
21. StataCorp. Stata Statistical Software: Release 15. 2017.
22. Shabbir SH, Sanders AE. Clinical significance in dementia research: A review of the literature. Am J Alzheimers Dis Other Demen. 2014;29(6):492-497. doi:10.1177/1533317514522539
23. Howard R, Phillips P, Johnson T, et al. Determining the minimum clinically important differences for outcomes in the DOMINO trial. Int J Geriatr Psychiatry. 2011;26(8):812-817. doi:10.1002/gps.2607
24. Zuidema SU, Buursema AL, Gerritsen MGJM, et al. Assessing neuropsychiatric symptoms in nursing home patients with dementia: Reliability and Reliable ChangeIndex of the Neuropsychiatric Inventory and the Cohen-Mansfield Agitation Inventory. Int J Geriatr Psychiatry. 2011;26(2):127-134. doi:10.1002/gps.2499
25. Cohen J. A Power Primer. Tutor Quant Methods Psychol. 2007;112. doi:10.1037/0033-2909.112.1.155
26. Smith GR, Taylor CW, Linkous P. Haloperidol Versus Thioridazine For The Treatment Of Psychogeriatric Patients: A Double-Blind Clinical Trial. Psychosomatics. 1974;15(3):134-138. doi:10.1016/S0033-3182(74)71262-2
27. Verhey FRJ, Verkaaik M, Lousberg R. Olanzapine versus haloperidol in the treatment of agitation in elderly patients with dementia: Results of a randomized controlled double-blind trial. Dement Geriatr Cogn Disord. 2005;21(1):1-8. doi:10.1159/000089136
28. Rainer M, Haushofer M, Pfolz H, Stru-

hal C, Wick W. Quetiapine versus risperidone in elderly patients with behavioural and psychological symptoms of dementia: efficacy, safety and cognitive function. Eur psychiatry. 2007;22(6):395-403.

29. Lovett WC, Stokes DK, Taylor LB, Young ML, Free SM, Phelan DG. Management of behavioral symptoms in disturbed elderly patients: Comparison of trifluoperazine and haloperidol. J Clin Psychiatry. 1987;48(6):234-236.
30. Auer SR, Monteiro IM, Reisberg B. Behavioral symptoms in dementia: Community-based research. Int Psychogeriatrics. 1996;8(SUPPL. 3):363-366. doi:10.1017/S1041610297003633
31. Rosen HJ. Double-blind comparison of haloperidol and thioridazine in geriatric outpatients. J Clin Psychiatry. 1979.
32. Vergara L, Amin M, Ban T. Systematic clinical studies with clomacran III. A standard controlled clinical trial in geropsychiatric organic brain syndrome patients. Curr Ther Res. 1980;27(1).
33. Götestam KG, Ljunghall S, Olsson B. A double-blind comparison of the effects of haloperidol and cis (Z)-clopenthixol in senile dementia. Acta Psychiatr Scand. 1981;64:46-53.
34. Spagnolo C, Dall Asta D, Iannnuccelli M, Cucinotta D, Passeri M. A controlled double blind trial comparing etoperidone with thioridazine in the management of severe senile dementia. Drugs Exptl Clin Res. 1983;4(12):873.
35. Morris R, Rickels K. Loxapine In Geriatric Patients With Chronic Brain Syndrome. Curr Ther Res. 1984;35(4):519-531.
36. Ather SA, Shaw SH, Stoker MJ. A comparison of chlormethiazole and thioridazine in agitated confusional states of the elderly. Acta Psychiatr Scand. 1986;73(10):81-91. doi:10.1111/j.1600-0447.1986.tb10541.x
37. Deberdt WG, Dysken MW, Rappaport S a, et al. Comparison of olanzapine and risperidone in the treatment of psychosis and associated behavioral disturbances in patients with dementia. Am J Geriatr Psychiatry. 2005;13(8):722-730. doi:10.1176/appi.ajgp.13.8.722
38. Mintzer JE, Tune LE, Breder CD, et al. Aripiprazole for the Treatment of Psychoses in Institutionalized Patients With Alzheimer Dementia: A Multicenter, Randomized, Double-Blind, Placebo-Controlled Assessment of Three Fixed Doses.; 2007.
39. Zhong K, Tariot P, Mintzer J, Minkwitz M, Devine N. Quetiapine to Treat Agitation in Dementia : A Randomized , Double-Blind , Placebo-Controlled Study. Curr Alzheimer Res. 2007;4(1):81-93.
40. Streim JE, Porsteinsson AP, Breder CD, et al. From the Section on Geriatric Psychiatry. Vol 16. Philadelphia Veteran Affairs Medical Center; 2008.
41. Allain H, Dautzenberg P, Maurer K, Schuck S, Bonhomme D, Gerard D. Double blind study of tiapride versus haloperidol and placebo in agitation and aggressiveness in elderly patients with cognitive impairment. Psycho-

pharmacology (Berl). 2000;(148):361-366.
42. Tariot PN, Schneider L, Katz IR, Mintzer JE. Quetiapine Treatment of Psychosis Associated With Dementia: A Double-Blind, Randomized, Placebo-Controlled Clinical Trial. Am J Geriatr Psychiatry. 2006;14(9):767-776.
43. Hamilton L, Bennet J. The use of trifluoperazine in geriatric patients with chronic organic brain syndrome. J Am Geriatr SOC. 1962;10(140).
44. Sugerman A, Williams B, Adlerstein A. Haloperidol in the psychiatric disorders of old age. Am J Psychiatry. 1964;120(12):1190-1192.
45. Rada RT, Kellner R. Thiothixene in the treatment of geriatric patients with chronic organic brain syndrome. J Am Geriatr Soc. 1976;24(3):105-107.
46. Barnes R, Veith R, Okimoto J, Raskind M, Gumbrecht G. Efficacy of antipsychotic medications in behaviorally disturbed dementia patients. Am J Psychiatry. 1982.
47. Petrie W, Ban T, Berney S, et al. Loxapine in Psychogeriatrics: A Placebo- and Standard-Controlled Clinical Investigation. J Clin Psychopharmacol. 1982;2(2):122-126.
48. Stotsky B. Multicenter Study Comparing Thioridazine with Diazepam and Placebo in Elderly, Nonpsychotic Patients with Emotional and Behavioral Disorders. Clin Ther. 1984;6(4):546-559.
49. Finkel SI, Lyons JS, Anderson RL, et al. A randomized, placebo-controlled trial of thiothixene in agitated, demented nursing home patients'. Int J Geriatr Psychiatry. 1995;10(2):129-136. doi:10.1002/gps.930100208
50. Auchus AP, Cheryl Bissey-Black RNC. CLINICAL Pilot Study of Haloperidol, Fluoxetine, and Placebo for Agitation in Alzheimer's Disease.; 1997.
51. Devanand DP, Marder K, Michaels KS, et al. A radomized, placebo-controlled dose-comparison trial of haloperidol for psychosis and disruptive behaviors in Alzheimer's disease. Am J Psychiatry. 1998;155(11):1512-1520. doi:10.1176/ajp.155.11.1512
52. Teri L, Logsdon RG, Peskind E, et al. Treatment of agitation in AD: A randomized, placebo-controlled clinical trial. Neurology. 2011;55:1271.
53. Pollock BG, Mulsant BH, Rosen J, et al. Article Comparison of Citalopram, Perphenazine, and Placebo for the Acute Treatment of Psychosis and Behavioral Disturbances in Hospitalized, Demented Patients. Vol 159.; 2002.
54. Katz IR, Jeste V, Mintzer JE, Clyde C, Napolitano J, Brecher M. Comparison of Risperidone and Placebo for Psychosis and Behavioral Disturbances Associated With Dementia: A Randomized, Double-Blind Trial. Vol 60.; 1999.
55. Street JS, Clark W, Gannon KS, Cummings JI, et al. Olanzapine Treatment of Psychotic and Behavioral Symptoms in Patients With Alzheimer Disease in Nursing Care Facilities. Arch Gen Psychiatry. 2000;57(10):968. doi:10.1001/archpsyc.57.10.968
56. Howanitz E, Wisotzek I. Olanzapine versus placebo in the treatment of behavioral disturbances associated with

vascular dementia. 2001.
57. Carlyle W, Ancill R, Sheldon L. Aggression in the demented patient: a double-blind study of loxapine versus haloperidol. Int Clin Psychopharmacol. 1993;8:103-108.
58. Herz L, Volicer L, Frankenburg F, Colon S, S K. A 6 week, double blind comparison of olanzapine, risperidone, and placebo for behavioural disturbance in Alzheimers disease. In: The International College of Geriatric Psychoneuropharmacology. ; 2002.
59. Brodaty H, Ames D, Snowdon J, et al. A randomized placebo-controlled trial of risperidone for the treatment of aggression, agitation, and psychosis of dementia. J Clin Psychiatry. 2003;64(2):134-143. doi:http://dx.doi.org/10.4088/JCP.v64n0205
60. De Deyn PP, Carrasco MM, Deberdt W, et al. Olanzapine versus placebo in the treatment of psychosis with or without associated behavioral disturbances in patients with Alzheimer's disease. Int J Geriatr Psychiatry. 2004;19(2):115-126. doi:10.1002/gps.1032
61. Ballard C, Margallo-Lana M, Juszczak E, et al. Quetiapine and rivastigmine and cognitive decline in Alzheimer's disease: randomised double blind placebo controlled trial. BMJ. 2005;330(7496):874. doi:10.1136/bmj.38369.459988.8F
62. Fontaine CS, Koch K, Martin-Cook K, Svetlik D, Weiner MF, Hynan LS. A double-blind comparison of olanzapine versus risperidone in the acute treatment of dementia-related behavioral disturbances in extended care facilities. J Clin Psychiatry. 2003;64(6):726-730. doi:10.4088/JCP.v64n0617
63. Sheng S, Gao Z, Chen M, et al. Risperidone vs haloperidol in treatment of behavioral and psychological symptoms of dementia. A randomized, double blind trial. Chinese J New Drugs Clin Remedies. 2004.
64. Gareri P, Cotroneo A, Lacava R, et al. Comparison of The Efficacy of New and Conventional Antipsychotic Drugs in the Treatment of Behavioral and Psychological Symptoms of Dementia (BPSD). Arch gerontol geriatr Suppl. 2004;9:207-215.
65. Mulsant BH, Pollock BG, Gharabawi GM, et al. Correlates of anticholinergic activity in patients with dementia and psychosis treated with risperidone or olanzapine. J Clin Psychiatry. 2004;65(12):1708-1714. doi:10.4088/JCP.v65n1217
66. Sun X, Gao Z, Feng F. A randomized double-blind trial of haloperidol and risperidone for behavioral and psychological symptoms of dementia. Chinese J Psychiatry. 2004;37(3):156-159.
67. Mintzer J, Greenspan A, Caers I, et al. Risperidone in the Treatment of Psychosis of Alzheimer Disease: Results From a Prospective Clinical Trial. Am J Geriatr Psychiatry. 2006;(March):280-291.
68. Ballard C, Banister C, Khan Z, et al. Evaluation of the safety, tolerability, and efficacy of pimavanserin versus placebo in patients with Alzheimer's disease psychosis: a phase 2, randomised, placebo-controlled, dou-

ble-blind study. LANCET Neurol. 2018;17(3):213-222.
69. Mintzer J, Greenspan A. Risperidone in the Treatment of Psychosis of Alzheimer Disease: Results From a Prospective Clinical Trial. Vol 14.; 2006.
70. Chan A, Hrobjartsson A, Jorgensen K, Gotzsche P, Altman D. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. BMJ. 2008;337:1-8. doi:10.1136/bmj.a2299
71. Roggla G, Fortunat S. Are cancer trials frequently overpowered? Br Med J. 2004;328:1463.
72. Lin M, Henry C. Lucas J, Galit Shmueli. Research Commentary—Too Big to Fail: Large Samples and the p-Value Problem. Inf Syst Res. 2013;24(4):906-917.
73. CONSORT Group. Consolidated Standards of Reporting Trials.
74. Charles P, Giraudeau B, Dechartres A, Baron G, Ravaud P. Reporting of sample size calculation in randomised controlled trials: review. BMJ. 2009;338(b1732).
75. Latif L, Eduardo J, Amadera D, Pimentel D, Pimentel T, Fregni F. Sample Size Calculation in Physical Medicine and Rehabilitation: A Systematic Review of Reporting, Characteristics, and Results in Randomized Controlled Trials. Arch Phys Med Rehabil. 2011;92(2):306-315.
76. Tavernier E, Giraudeau B. Sample size calculation: Inaccurate A priori assumptions for nuisance parameters can greatly affect the power of a randomized controlled trial. PLoS One. 2015;10(7):8-15. doi:10.1371/journal.pone.0132578

Appendix

Search string

In title/ abstract

chlorpromazine OR levomepromazine OR promazine OR acepromazine OR triflupromazine OR cyamemazine OR chlorproethazine OR dixyrazine OR fluphenazine OR perphenazine OR prochlorperazine OR thiopropazate OR trifluoperazine OR acetophenazine OR thioproperazine OR butaperazine OR perazine OR periciazine OR thioridazine OR mesoridazine OR pipotiazine OR haloperidol OR trifluperidol OR melperone OR moperone OR pipamperone OR bromperidol OR benperidol OR droperidol OR fluanisone OR oxypertine OR molindone OR sertindole OR ziprasidone OR lurasidone OR flupentixol OR clopenthixol OR chlorprothixene OR tiotixene OR zuclopenthixol OR fluspirilene OR pimozide OR penfluridol OR loxapine OR clozapine OR olanzapine OR quetiapine OR asenapine OR clotiapine OR sulpiride OR sultopride OR tiapride OR remoxipride OR amisulpride OR veralipride OR levosulpiride OR prothipendyl OR risperidone OR mosapramine OR zotepine OR aripiprazole OR paliperidone OR iloperidone OR cariprazine OR brexpiprazole OR pimavanserin

AND

Anywhere: trial

AND

Anywhere: dementia

Table 1. NPI-NH in included studies that used this instrument

Study	Baseline NPI-NH	Change in placebo group
Street 2000	43	28
DeDeyn 2004	33	20
Deberdt 2005	42	22
DeDeyn 2005	40	19
Schneider 2006	37	9
Mintzer 2007	41	16
Zhong 2007	36	23
Paleacu 2008	41	19
Streim 2008	38	19
Mean	**39**	**19**

Part II.

Quantifying risk of bias

Chapter 5.

Run-in periods and clinical outcomes of antipsychotics in dementia: A meta-epidemiological study of placebo-controlled trials

Pharmacoepidemiol Drug Saf. 2016 Feb;25(2):113-22

Hulshof, T.A.
Zuidema, S.U.
Gispen-de Wied, C.C.
Luijendijk, H.J.

Abstract

Purpose

Run-in periods are used to identify placebo-responders and washout. Our aim was to assess the association of run-in periods with clinical outcomes of antipsychotics in dementia.

Methods

We searched randomized placebo-controlled trials of conventional and atypical antipsychotics for neuropsychiatric symptoms (NPS) in dementia in electronic sources and references of selected articles. We extracted (1) presence of a run-in period, use of placebo/ investigated drug during run-in (versus washout only), and run-in duration (one week or more), and (2) reduction in NPS, number of participants with somnolence, extrapyramidal symptoms (EPS) and deaths per treatment group. We pooled clinical outcomes comparing antipsychotic and placebo groups in trials with and without run-in.

Results

We identified 35 trials. Twenty-nine trials used run-in. The pooled standardized mean difference in reduction of NPS was -0.170 (95%CI: -0.227 to -0.112) in trials with, and -0.142 (95%CI: -0.331 to 0.047) in trials without run-in. The pooled odds ratio for somnolence was 2.8 (95%CI: 2.3 to 3.5) in trials with run-in and 3.5 (95%CI: 1.2 to 10.7) in trials without run-in; for EPS these ORs were 1.8 (95%CI: 1.4 to 2.2) and 2.0 (95%CI: 1.3 to 3.1) respectively, and for mortality 1.4 (95%CI: 1.0 to 2.0) and 1.6 (95%CI: 0.7 to 3.4). Use of placebo/investigated drug during run-in and run-in duration did not affect the estimates in a consistent way.

Conclusions

Use of run-in in trials might have led to overestimated efficacy and especially underestimated risks of side effects of antipsychotics compared to placebo for NPS in dementia.

Introduction

Results of randomized controlled trials are important for regulatory and clinical decisions. Researchers have therefore sought to optimize treatment effects and identify patients that will benefit most from treatment. One way of enhancing trial design is using a run-in period between screening for eligibility and before randomization.[1,2] During this period of usually one to two weeks, drugs that the eligible patients already use are washed out. In some trials, the drugs are replaced by placebo to blind the participants for the change in treatment. Drug-naïve patients can also be given placebo, or the active drug of interest. Patients with high placebo response, poor compliance, low treatment response or intolerance for the drug can thus be identified.[3,4] At the end of the run-in phase, the researchers select the participants that are definitively included in the study. It is assumed that a run-in period will decrease placebo response and dropout during the trial, and consequently increase the effect size and the power of a trial.[5,6]

A small number of reviews have studied the effect that a run-in period can have on trial outcomes of psychopharmacological drugs. Antidepressants in children were 15% more effective in trials with a run-in period than in trials without a run-in period and above the threshold for a small effect size (standardized mean difference 0.26 versus 0.17 respectively; cut-off for small effect is 0.20).[7] Another meta-analysis of antidepressant trials in depressed outpatients showed that a placebo run-in period was associated with higher efficacy and more power.[8] On the other hand, run-in periods were not associated with greater efficacy in trials of antidepressants for major depression, benzodiazepines for anxiety, and naltrexon for alcohol addiction.[9–13]

To our knowledge, the effect of a run-in period on efficacy and side effects of antipsychotics has not been investigated before. This is notable, because high placebo response rates, high dropout rates, and decreasing effectiveness over the years are a major problem in antipsychotic trials.[14] An association between use of a run-in period and drug safety is not unlikely, because run-in periods can lead to exclusion of persons not tolerating the drugs and of noncompliant subjects.[15] Moreover, atypical antipsychotics have been marketed with the claim of a more favorable side effects profile compared to conventional antipsychotics, i.e. lower rates of somnolence and extrapyramidal symptoms (EPS).[16]

Antipsychotics are often prescribed for neuropsychiatric symptoms in dementia. Trials that tested the efficacy of antipsychotics for this indication commonly used run-in periods. The aim of this study was to assess the association of run-in periods in trials of conventional and atypical antipsychotics in dementia with clinical outcomes and also dropout.

Methods

We performed a meta-epidemiological study. We wrote a research proposal for the sponsor in advance and it can be requested from the corresponding author.

Search strategy

Four sources were used to identify trials. Two reviewers (TAH, HJL) first searched the electronic databases Cinahl, Embase, Pubmed, and Cochrane library with the strings [*'generic name atypical/conventional antipsychotic'* AND *trial* AND *dementia* (see appendix). We composed a list of all conventional and atypical antipsychotics from the websites of the World Health Organization, Food and Drug Administration and Wikipedia to enable this search.[17–19] Secondly, we hand-searched the references of published systematic reviews, which were identified with the same electronic databases. Titles and abstracts of potentially eligible studies were retrieved in Pubmed. Thirdly, we sought RCTs in trial registration websites with the same keywords where possible. Finally, we searched the databases of the Dutch Medicines Evaluation Board and the FDA for unpublished trials of atypical antipsychotics.

Study selection

Randomized placebo-controlled trials that tested the efficacy of orally administered conventional or atypical antipsychotics for neuropsychiatric symptoms in dementia were included. If studies seemed potentially eligible given title and abstract, full articles were retrieved as well as online protocols of unpublished studies. Two reviewers (TAH, HJL) reviewed these articles for definitive eligibility. Studies with no information on the use of a run-in period, and with multiple drugs in one intervention arm were excluded. There were no restrictions with respect to publication date, language, flexible or fixed dosing of the active treatment, and duration of the study. The search was last rerun in June 2019.

Data extraction

Two reviewers (TAH, HJL) independently extracted data from the included trials. First, we extracted general study characteristics: publication year, type of antipsychotic groups, setting, type of neuropsychiatric symptoms for which the antipsychotics were tested, and total number of randomized patients. Next, we extracted characteristics of the run-in period: presence/ absence, replacement drug (no drug, placebo or active treatment), and duration, as well as the percentage of patients excluded at the end of the run-in phase.

We then extracted the clinical outcomes for the drug and placebo groups. Efficacy of antipsychotics in dementia can be measured with a generic instrument that covers

various neuropsychiatric symptoms (e.g. NPI, BEHAVE-AD), or an instrument for specific symptoms such as agitation (e.g. CMAI). We used the results measured with the instrument that matched the symptoms at enrollment. E.g. if patients had to have agitation to enter a trial, we used the result reported with an agitation scale. We extracted the mean change from baseline to end point. When multiple dosages or multiple drug groups were included in a trial, an average change was calculated. We also extracted the standard deviation (SD) of the difference between the groups in mean change. If the SD was not reported, it was calculated with the p-value, range, or confidence interval reported for the difference in mean change. Otherwise, the SD was imputed with the average of the reported SD of all trials with the same indication and instrument. For two trials that did not report the data we needed, we obtained the IPD and calculated the mean changes and SDs.[20,21] In addition, we abstracted the number of patients with somnolence (sedation, drowsiness), and with EPS, and the number that died.

Finally, we extracted the total number of patients that dropped out (total drop-out), and the drop-out in the groups (selective drop-out). Run-in is often used to decrease drop-out and enhance power. Drop-out is also considered to represent the balance between efficacy and side-effects.[20]

The published main results article of a trial was our primary source of information. When the article did not report the data that we needed, secondary publications, trial reports and meta-analyses published online by industry were our secondary source. We contacted the authors of eligible trials to provide missing data, or individual patient data, and received such data of four studies.[20–23] The reviewers discussed differences in the extracted data until consensus was reached.

Data analyses

First, we assessed the relationship between the presence of a run-in period with the four clinical outcomes and selective dropout. We performed meta-analyses to pool efficacy, and risk of somnolence, EPS and mortality of the antipsychotic versus placebo groups in trials with and without run-in periods. For efficacy in terms of reduction in NPS, we calculated standardized mean differences (SMDs) to take into account the use of different instruments in the trials. SMDs were calculated with a 95% confidence interval (CI). For risk of somnolence, EPS, mortality, and selective dropout we calculated odds ratios (ORs) with 95% CIs. Heterogeneity, presented as I^2, was calculated for all meta-analyses. A fixed-effects model was applied when I^2 was below 40%, otherwise a random-effects model.[24]

Table 1. Randomized placebo-controlled trials of antipsychotics in dementia with neuropsychiatric symptoms related to dementia

Author, year	Antipsychotic	Setting	Neuropsychiatric symptom
Conventional antipsychotic trials (11)			
Hamilton, 1962	Trifluoperazine	Hospital	Diverse
Sugerman, 1964	Haloperidol	Hospital	Diverse
Rada, 1976	Thiothixene	Hospital	Diverse
Barnes, 1982	Thioridazine Loxapine	Nursing home	Diverse
Petrie, 1982	Haloperidol Loxapine	Hospital	Diverse
Stotsky, 1984	Thioridazine	Nursing home & Hospital	Diverse
Finkel, 1995	Thiothixene	Nursing home	Agitation
Auchus, 1997	Haloperidol	Outpatients	Agitation
Devanand, 1998	Haloperidol	Outpatients	Diverse
Teri, 2000	Haloperidol	Hospital	Agitation
Pollock, 2002	Perphenazine	Nursing home	Diverse
Atypical antipsychotic trials (21)			
Satterlee, 1995	Olanzapine	Nursing home	Psychosis†
Janssen Ph, 1997	Risperidone	NR	Diverse
Katz, 1999	Risperidone	Nursing home	Diverse
Street, 2000	Olanzapine	Nursing home	Diverse
Brodaty, 2003	Risperidone	Nursing home	Aggression
Janssen Ph, 2003	Risperidone	Nursing home	Psychosis
De Deyn, 2004	Olanzapine	Nursing home	Psychosis
Ballard, 2005	Quetiapine	Nursing home	Agitation
De Deyn, 2005	Aripiprazole	Outpatients	Psychosis
Deberdt, 2005	Risperidone Olanzapine	Nursing home & Outpatients	Psychosis
Janssen Ph, 2005	Risperidone	NR	Psychosis
Mintzer, 2006	Risperidone	Nursing home	Psychosis
Schneider, 2006	Risperidone Olanzapine Quetiapine	Outpatients	Diverse
Mintzer, 2007	Aripiprazole	Nursing home	Psychosis
Zhong, 2007	Quetiapine	Nursing home	Agitation
Paleacu, 2008	Quetiapine	Not reported	Diverse
Streim, 2008	Aripiprazole	Nursing home	Psychosis
Otsuka Ph, 2017a	Brexpiprazole	Nursing home	Agitation
Otsuka Ph, 2017b	Brexpiprazole	Nursing home & Outpatients	Agitation
ACADIA, 2018	Pimavanserin	Nursing home & Outpatients	Agitation
Ballard, 2018	Pimavanserin	Nursing home	Psychosis
Trials with conventional and atypical antipsychotic drug group (3)			
Allain, 2000	Tiapride Haloperidol	Nursing home & Hospital	Agitation
De Deyn, 1999	Risperidone Haloperidol	Nursing home	Diverse
Tariot, 2006	Quetiapine Haloperidol	Nursing home	Psychosis

NR: Not Reported; NA: Not Applicable; Ph: Pharmaceuticals; † reduction measured with generic instrument (in all other studies indication and outcome scale were congruent); ‡ at least 2 days.

Run-in period (duration in weeks)	Patients excluded after run-in, n (%)	Randomized patients, n
No	NA	27
No	NA	18
Yes, with placebo (1)	0 (0)	42
Yes, with placebo (2)	7 (11.7)	53
Yes, with placebo (2)	0 (0)	61
Yes, washout only (2)	NR	358
Yes, washout only (1)	0 (0)	33
Yes, washout only (2)	0 (0)	12
Yes, with placebo (1)	5 (7.0)	66
Yes, washout only (2)	0 (0)	70
Yes, washout only (<1)	NR	54
Yes, washout only	51 (17.7)	238
Yes, with placebo (1)	NR	39
Yes, with placebo (1)	104 (14.3)	625
Yes, with placebo (2)	82 (28.5)	206
Yes, with placebo (1)	39 (10.2)	345
Yes, with placebo (1)	NR	18
Yes, with placebo (2)	NR	652
No	NA	62
Yes, washout only (1)	NR	208
Yes, with placebo (2)	NR	494
Yes, with risperidone (1)	NR	33
Yes, with placebo (1)	87 (15.5)	473
No	NA	421
Yes, washout only (1)	NR	487
No	NA	333
Yes, with quetiapine (2)	NR	40
Yes, washout only (1)	NR	256
Yes, washout only (6)	NR	413
Yes, washout only (6)	NR	270
Yes, washout only (4)	NR	111
Yes, washout only (3)	25 (12.1)	181
No	NA	306
Yes, with placebo (1)	27 (7.3)	344
Yes, washout only (1)‡	123 (24.6)	284

We then investigated the relationship of the characteristics of the run-in periods with the clinical outcomes and selective dropout. The outcomes of trials that had run-in periods with and without a placebo were pooled (we did not find trials with active drugs in the run-in period). We also pooled outcomes of trials with a run-in period up to one week versus those with a longer run-in period (8 to 14 days). Finally, we tested the association between run-in characteristics and total drop-out with meta-regression.

We ran the above analyses for all antipsychotics combined first, and then for the conventional antipsychotics and atypical antipsychotics separately. We wanted to perform an a-priori sensitivity-analysis for atypical antipsychotics without quetiapine, because its efficacy and side-effect profile is considered to differ. A post-hoc analysis was run in which trials that had tested both an atypical drug and haloperidol were excluded from the analysis. All analyses were performed with Stata statistical software version 15.0.[25]

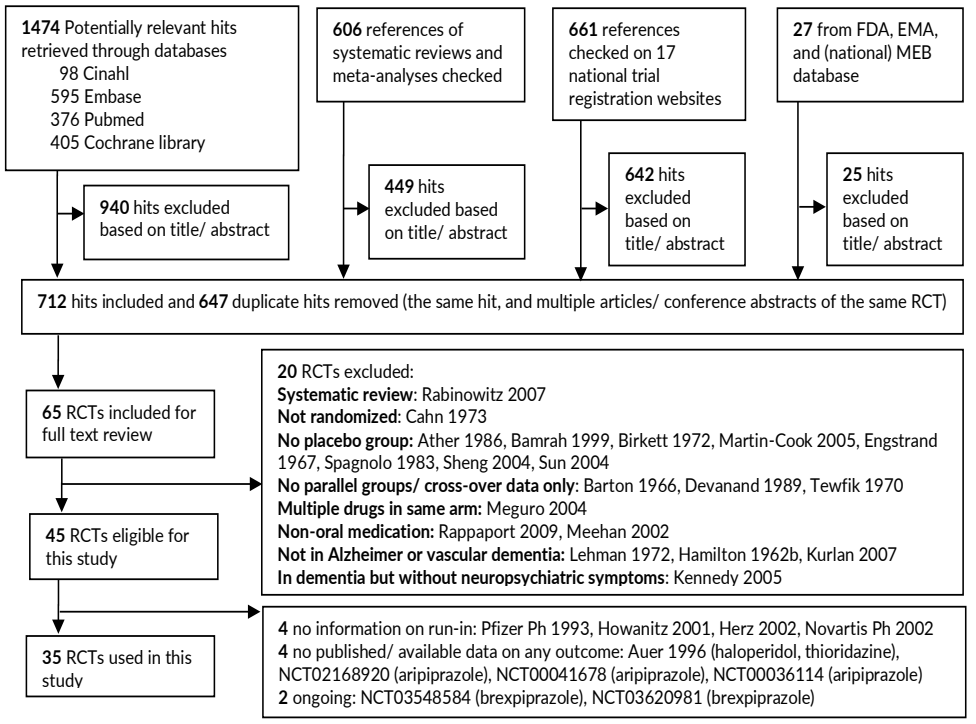


Figure 1. Flow diagram of literature search and study selection.

Table 2. Use of run-in periods and clinical outcomes of antipsychotics versus placebo in randomized trials

	Efficacy SMD (95% CI)	N	**Somnolence** OR (95% CI)	N	**EPS** OR (95% CI)	N	**Mortality** OR (95% CI)	N
Conventional and atypical antipsychotics								
No run-in	-0.142 (-0.331; 0.047)[a]	4	3.5 (1.2; 10.7)[a]	4	2.0 (1.3; 3.1)	5	1.6 (0.7; 3.4)	6
With run-in	-0.170 (-0.227; -0.112)	25	2.8 (2.3; 3.5)	16	1.8 (1.4; 2.2)	14	1.4 (1.0; 2.0)	28
- washout only	-0.146 (-0.267; -0.024)[a]	12	2.6 (1.4; 4.9)[a]	8	1.7 (1.1; 2.7)	5	1.2 (0.7; 2.1)	14
- with placebo/ drug	-0.190 (-0.267; -0.112)	13	2.7 (2.1; 3.6)	8	1.8 (1.4; 2.4)	9	1.5 (1.0; 2.4)	14
- duration =< 1 week	-0.214 (-0.289; -0.138)	13	3.0 (2.1; 4.5)[a]	9	1.8 (1.4; 2.4)	10	1.4 (0.9; 2.0)	14
- duration > 1 week	-0.108 (-0.197; -0.019)	12	2.5 (1.7; 3.9)	7	1.8 (0.8; 4.1)[a]	4	1.5 (0.7; 3.0)	13
Conventional antipsychotics								
No run-in	-0.389 (-0.669; -0.110)	1	4.3 (0.2; 110.2)[a]	2	2.7 (1.4; 5.1)	2	1.4 (0.3; 6.9)	3
With run-in	-0.345 (-0.492; -0.199)	9	5.4 (3.2; 9.3)	4	3.0 (1.9; 4.8)	4	1.2 (0.6; 2.3)	11
Atypical antipsychotics								
No run-in	-0.133 (-0.266; 0.001)	4	2.8 (0.9; 8.3)[a]	3	1.6 (0.7; 3.6)[a]	4	1.6 (0.7; 3.8)	4
With run-in	-0.141 (-0.202; -0.081)	18	2.6 (2.1; 3.3)	14	1.6 (1.2; 2.0)	12	1.4 (1.0; 2.1)	19

EPS: extrapyramidal symptoms; SMD: standardized mean difference; OR: odds ratio

[a] random effects model was used.

Table 3. Use of run-in periods and dropout in antipsychotic trials in dementia

	Selective dropout[a] OR (95% CI)	N	**Total dropout** Beta (95% CI)	N
Conventional and atypical antipsychotics				
No run-in	1.0 (0.7; 1.3)	5	ref	
With run-in	1.0 (0.9; 1.2)	26	0.3 (-15.1; 15.7)	33
- washout only	0.9 (0.7; 1.1)	13	ref	
- with placebo/ drug	1.2 (1.0; 1.4)	13	-4.7 (-14.9; 5.5)	27
- duration =< 1 week	0.9 (0.8; 1.1)	14	ref	
- duration > 1 week	1.2 (1.0; 1.5)	12	-5.0 (-14.7; 4.6)	26
Conventional antipsychotics				
No run-in	1.4 (0.7; 2.8)	2	ref	
With run-in	1.0 (0.7; 1.3)	9	16.0 (-7.5; 39.6)	13
Atypical antipsychotics				
No run-in	0.9 (0.5; 1.7)[b]	4	ref	
With run-in	1.0 (0.9; 1.2)	19	-7.7 (-25.1; 9.7)	23

OR: odds ratio, [a]drug group versus placebo, [b]random effects model was used

Results

Our search yielded 2768 potentially relevant RCTs (Figure 1). We obtained the reports of 65 RCTs for full text review. We identified 45 eligible RCTs, but four did not report whether a run-in period was used, four did not report any of the outcomes of interest, and two were ongoing. We used the other 35 studies in the current study.[20–23,26–56]

Table 1 presents the general study characteristics. Twenty-nine of the 35 studies had a run-in period: 9 of 11 conventional antipsychotic trials, 18 of 21 atypical antipsychotic trials, and 2 of 3 trials with both antipsychotics. Fourteen studies used placebo during the run-in period, two trials the investigated drug, and 19 studies no placebo or active treatment (washout only). The duration of the run-in periods varied between 2 days and 6 weeks. The percentage of patients excluded at the end of the run-in period varied from 0 to 29%. In five of seven relatively old conventional antipsychotic trials the percentage was 0%, while in the atypical trials it was at least 7%.

Run-in and clinical outcome

The analysis of efficacy encompassed 29 of 35 studies (table 2). The reduction in NPS in the drug versus placebo groups was somewhat higher in trials with a run-in period (SMD -0.170; 95% CI: -0.227 to -0.112) than in trials without a run-in period (SMD -0.142; 95%CI: -0.331 to 0.047). Efficacy was somewhat higher when placebo or active drug was used (SMD -0.190; 95% CI: -0.267 to - 0.112), and when run-in lasted 1 week at most (SMD -0.214; 95% CI: -0.289 to -0.138).

The number of participants with somnolence during the study period was reported in 20 of 35 studies. The pooled risk for somnolence was lower when a run-in period was present (OR 2.8; 95%CI: 2.3 to 3.5) versus when it was absent (OR 3.5; 95%CI 1.2 to 10.7). Use of placebo or active drug did not seem to affect the risk of somnolence further. The risk was lower for trials with a run-in period of one week at most (OR 2.5; 95%CI 1.7 to 3.9).

Nineteen of 35 studies reported the number of participants with EPS in the treatment groups. The risk of EPS in trials with a run-in period was slightly lower (OR 1.8; 95%CI: 1.4 to 2.2) than in trials without a run-in period (OR 2.0; 95%CI: 1.3 to 3.1). Use of placebo or duration of run-in did not seem to reduce the risk of EPS further. Data of all but one trial could be used for the analysis of mortality risk. The risk of mortality was 1.6 (95%CI 0.7 to 3.4) and 1.4 (95%CI 1.0 to 2.0) in trials without and with run-in respectively. The risk was slightly lower when no placebo was used, but run-in duration did not seem to affect it.

The sub-analyses in trials of atypical antipsychotics only yielded similar results for use of run-in versus no run-in on all outcomes (table 2). In trials of conventional antipsychotics, however, efficacy seemed lower and risk of somnolence and EPS higher in trials with run-in compared to trials without run-in, but the number of trials without run-in was 3 at most and confidence intervals were large.

Run-in and drop-out

Thirty-one of 35 studies reported dropout rates for the antipsychotic and placebo groups (table 3). The number of participants that dropped out was 1519 of 4963 in the antipsychotic groups (30.6%) and 783 of 2747 in the placebo groups (28.5%). The odds ratio of selective dropout was 1.0 (95%CI: 0.9 to 1.2) for trials with a run-in period and 1.0 (95%CI: 0.7 to 1.3) for those without. Total dropout in the studies varied between 0 and 81.7%. Use of a run-in period was not associated with a decreased total dropout (beta 0.3%; 95%CI -15.1 to 15.7) (Table 3).

Sensitivity analyses

There were not enough trials with and without run-in periods for the sensitivity-analysis of atypical antipsychotics without quetiapine.[57] The results of the analysis without trials that tested both haloperidol and an atypical drug confirmed the pattern of higher efficacy and lower risk of side effects for the conventional and atypical antipsychotic drugs in trials with versus without run-in (see table 1 in appendix).

Discussion

We assessed the association between use of a run-in period and clinical outcomes of 35 antipsychotic trials in dementia. The reduction in neuropsychiatric symptoms of antipsychotics versus placebo was somewhat higher in trials with a run-in period than in trials without a run-in period. Risk of somnolence, EPS, and mortality was lower in trials with than without run-in. Accordingly, risk of dropout in the antipsychotic compared to the placebo groups, which represents the balance between beneficial and harmful effects, was not affected when run-in was used. Use of run-in periods did not influence total dropout either.

Several reviews have reported an association between the use of run-in periods and the efficacy of psychotropics. Two meta-analyses of antidepressant trials showed that a placebo run-in period was associated with higher effectiveness and more power.[7,8] We found that the use of run-in periods was associated with a small increase in efficacy of antipsychotics in dementia. Additionally, we found an association between run-in periods and a decreased risk of side effects of antipsychotics in dementia. The

exclusion of placebo-responders and drug-intolerant patients after the run-in period might have led to increased efficacy and a more favorable side-effect profile. The effect of run-in periods on outcomes of trials has not been investigated often and remains an under-investigated and likely underestimated source of bias.

A common argument for use of run-in periods is the reduction of non-compliance and dropout. Our findings showed that a placebo run-in was not associated with lower between-group or total dropout rate.

Bias due to run-in periods

In the studies that we identified, patients that met the inclusion criteria could have been excluded from trial participation as a result of the outcomes during the run-in period. It is therefore not surprising that use of run-in period yielded higher efficacy estimates and lower risks of side-effects.[3,58] In observational studies, bias due to (de) selection of patients based on prior treatment and its outcomes, whether before or after the start of the study, is generally called selection bias.[59]

Bias due to run-in periods in trials is not commonly discussed in the literature. Most tools for risk of bias assessment in trials do not require consideration of run-in periods either. In 8 of 11 meta-analyses of antipsychotic trials in dementia, more than 80% of the included trials used a run-in period (see table 2 in appendix).[60–70] Risperidone and olanzapine, currently the two most popular antipsychotic drugs for use in dementia, have been tested in ten trials of which nine included a run-in period. The general assumption is that selection of patients before randomization is said to reduce only generalizability of study results, not the internal validity of the trial results.[1]

We propose a different view. The screening and selection of patients before randomization should be based on (contra-)indications that are applied in daily medical practice. The selected patients will then represent the patients of interest to doctors and the population of interest as defined in the PICOs of reviews. This selection needs to be distinguished from de-selection of eligible patients based on observed treatment effects during run-in (between screening and randomization).[2] The remaining randomized group does not represent the population of interest any more. Estimates of efficacy and risk of side effects will be biased for the population of interest. Therefore, run-in needs to be considered as a source of bias in trials and reviews of trials.

Another issue to consider is the ethics of entering patients who are doing (relatively) well on a certain antipsychotic drug into a trial of another or the same antipsychotic.

During washout, symptoms could return, and it is questionable whether the patient will respond as favorably to a new drug. Especially when it is difficult to convince patients to use antipsychotics and find an antipsychotic that has the desired effect, which is often the case in schizophrenia, switching to another drug for the sake of a trial is even more questionable. Including new instead of prevalent users in trials would be preferable, as is the recognized practice in observational epidemiology.

Strengths and limitations

To our knowledge, this is the first study that investigated the relationship of run-in periods with clinical outcomes of antipsychotic trials including side effects. Our pooled estimates of efficacy seemed low but are corroborated by previous reviews reporting SMDs between 0.12 and 0.21.[61,66,70] SMDs above the threshold of 0.200 for a small treatment effect were mainly found in meta-analyses that focused on aggression or agitation.[63,64,68,70]

A limitation of our research was that only six studies did not use a run-in phase. Most of these studies were performed in outpatients and with atypical antipsychotics, in particular quetiapine. As a result of this distribution, the higher efficacy in trials with run-in might be partly attributable to a higher efficacy of conventional antipsychotics. Nevertheless, our sensitivity analysis in atypical antipsychotic trials showed a higher efficacy for trials with a run-in period as well. Additionally, one would expect the risk of side effects to increase with run-in as well, but it did not.

Conclusion

Use of a run-in period is very common in antipsychotic trials for dementia. In these trials, efficacy was higher compared to trials without run-in, while the risk of side effects was lower. Therefore, use of a run-in period in trials might have led to overestimated efficacy and especially underestimated side effects of antipsychotics for neuropsychiatric symptoms in dementia. Meta-analyses should include sensitivity-analyses of trials with and without run-in periods.

References

1. Rothwell P. External validity of randomised controlled trials : To whom do the results of this trial apply? Lancet. 2005;365(1):82-93.
2. Leber P, Davis C. Threats to the Validity of Clinical Trials Sample Selection. Control Clin Trials. 1998;19:178-187.
3. Pablos-Méndez A, Barr R, Shea S. Run-in periods in randomized trials: implications for the application of results in clinical practice. JAMA. 1998;279(3):222-225.
4. Cipriani A, Gedds J. What is a run-in phase ? Epidemiol Psichiatr Soc. 2010;19(1):21-22.
5. Ulmer M, Robinaugh D, Friedberg J, Lipsitz S, Natarajan S. Usefulness of a run-in period to reduce drop-outs in a randomized controlled trial of a behavioral intervention. Contemp Clin Trials. 2008;29(5):705-710.
6. Brittain E, Wittes J. The run-in period in clinical trials. The effect of misclassification on efficiency. Control Clin Trials. 1990;11(5):327-338.
7. Bridge JA, Iyengar S, Salary CB. Clinical Response and Risk for Reported Suicidal Ideation and Suicide Attempts in Pediatric Antidepressant Treatment: A Meta-analysis of Randomized Controlled Trials. JAMA. 2007;297(15):1683-1696.
8. Khan A, Cohen S, Dager S, Avery D, Dunner D. Onset of response in relation to outcome in depressed outpatients with placebo and imipramine. J Affect Disord. 1989;17(1):33-38.
9. Trivedi M, Rush H. Does a placebo run-in or a placebo treatment cell affect the efficacy of antidepressant medications? Neuropsychopharmacology. 1994;11(1):33-43.
10. Lee S, Walker J, Jakul L, Sexton K. Does elimination of placebo responders in a placebo run-in increase the treatment effect in randomized clinical trials? A meta-analytic evaluation. Depress Anxiety. 2004;19(1):10-19.
11. Greenberg R, Fisher S, Riter J. Placebo washout is not a meaningful part of antidepressant drug trials. Percept Mot Ski. 1995;81(2):688-690.
12. Mitte K, Noack P, Steil R, Hautzinger M. A meta-analytic review of the efficacy of drug treatment in generalized anxiety disorder. J Clin Psychopharmacol. 2005;25(2):141-150.
13. Del Re A, Maisel N, Blodgett J, Finney J. The declining efficacy of naltrexone pharmacotherapy for alcohol use disorders over time: A multivariate meta-analysis. Alcohol Clin Exp Res. 2013;37(6):1064-1068.
14. Agid O, Siu C, Potkin S, Kapur S, Watsky E, Vanderburg D. Meta-regression analysis of placebo response in antipsychotic trials 1970-2010. Am J Psychiatry. 2013;170(11):1335-1344.
15. Schechtman K, Gordon M. A comprehensive algorithm for determining whether a run-in strategy will be a cost-effective design modification in a randomized clinical trial. Stat Med. 1993;12(2):111-128.
16. Arnt J, Skarsfeldt T. Do Novel Antipsychotics Have Similar Pharmacological

Characteristics? A Review of the Evidence. Neuropsychopharmacology. 1998;18(2):63-101.
17. World Health Organization. WHO Collaborating Centre for Drug Statistics Methodology. http://www.whocc.no/atc_ddd_index/?code=N05A. Published 2013.
18. US Food and Drug Administration. Atypical Antipsychotic Drugs Information. http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm094303.htm. Published 2013.
19. Wikipedia. Typical antipsychotic. Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Typical_antipsychotic. Published 2015. Accessed January 8, 2015.
20. Schneider L, Tariot P, Dagerman K, et al. Effectiveness of Atypical Antipsychotic Drugs in Patients with Alzheimer's Disease. N Engl J Med. 2006;355(15):1525-1538.
21. Paleacu D, Barak Y, Mirecky I, Mazeh D. Quetiapine treatment for behavioural and psychological symptoms of dementia in Alzheimer's disease patients: a 6-week, double-blind, placebo-controlled study. Int J Geriatr Psychiatry. 2008;23(September 2007):393-400.
22. De Deyn P, Rabheru K, Rasmussen A. A randomized trial of risperidone, placebo, and haloperidol for behavioral symptoms of dementia. Neurology. 1999;(53):945-956.
23. De Deyn P, Jeste D, Swanink R, Kostic D, Breder C. Aripiprazole for the Treatment of Psychosis in Patients With Alzheimer's Disease. J Clin Psychopharmacol. 2005;25(5):463-467. http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00004714-200510000-00010. Accessed April 17, 2014.
24. Deeks J, Higgins J, Altman D. Chapter 9: Analysing data and undertaking meta-analyses. In: Cochrane Handbook for Systematic Reviews of Interventions. Vol Version 5. Version 5. The Cochrane Collaboration; 2011.
25. StataCorp. Stata Statistical Software: Release 13. 2013.
26. Auchus A, Bissey-Black C. Pilot Study of Haloperidol, and Placebo Agitation in Alzheimer's Disease. Neuropsychiatry Clin Neurosci. 1997;9:591-593.
27. RIS-BEL-14. Janssen Pharmaceutical. 1997.
28. Devanand D, Sackeim HA, Marder K, et al. A Randomized, Placebo-Controlled Dose-Comparison Trial of Haloperidol for Psychosis and Disruptive Behaviors in Alzheimer's Disease. Am J Psychiatry. 1998;155(November):1512-1520.
29. Katz I, Jeste D, Mintzer J, Clyde C, Napolitano J, Brecher M. Comparison of Risperidone and Placebo for Psychosis and Behavioral Disturbances Associated With Dementia: A Randomized, Double-Blind Trial. J Clin Psychiatry. 1999;60(February):107-115.
30. Allain H, Dautzenberg P, Maurer K, Schuck S, Bonhomme D, Gerard D. Double blind study of tiapride versus haloperidol and placebo in agitation and aggressiveness in elderly patients with cognitive impairment. Psychopharmacology (Berl). 2000;(148):361-

366.

31. Street J, Clark S, Gannon K, Cummings J, Bymaster F. Olanzapine Treatment of Psychotic and Behavioral Symptoms in Patients With Alzheimer Disease in Nursing Care Facilities. Arch Gen Psychiatry. 2000;57:968-976.
32. Teri L, Logsdon RG, Peskind E. Treatment of agitation in AD: A randomized, placebo-controlled clinical trial. Neurology. 2000;55(November):1271-1277.
33. Pollock B, Mulsant B, Rosen J, et al. Comparison of Citalopram, Perphenazine, and Placebo for the Acute Treatment of Psychosis and Behavioral Disturbances in Hospitalized , Demented Patients. Am J Psychiatry. 2002;159(March):460-465.
34. Brodaty H, Ames D, Snowdon J, et al. A Randomized Placebo-Controlled Trial of Risperidone for the Treatment of Aggression, Agitation, and Psychosis of Dementia. J Clin Psychiatry. 2003;(February).
35. Hamilton L, Bennett J. The Use Of Trifluoperazine in Geriatric Patients With Chronic Brain Syndrome. J Am Geriatr Soc. 1962;Feb(10):140-147.
36. De Deyn P, Carrasco M, Deberdt W, et al. Olanzapine versus placebo in the treatment of psychosis with or without associated behavioral disturbances in patients with Alzheimer's disease. Int J Geriatr Psychiatry. 2004;19(2):115-126.
37. Ballard C, Margallo-Lana M, Juszczak E, et al. Quetiapine and rivastigmine and cognitive decline in Alzheimer's disease: randomised double blind placebo controlled trial. BMJ. 2005;330(7496):874.
38. Deberdt W, Dysken M, Rappaport S, Feldman P, Young C, Hay D. Comparison of Olanzapine and Risperidone in the Treatment of Psychosis and Associated Behavioral Disturbances in Patients With Dementia. Am J Geriatr Psychiatry. 2005;13(August):722-730.
39. Mintzer J, Greenspan A, Caers I, et al. Risperidone in the Treatment of Psychosis of Alzheimer Disease: Results From a Prospective Clinical Trial. Am J Geriatr Psychiatry. 2006;(March):280-291.
40. Tariot P, Schneider L, Katz I, et al. Quetiapine Treatment of Psychosis Associated With Dementia : A Double-Blind, Randomized, Placebo-Controlled Clinical Trial. Am J Geriatr Psychiatry. 2006;14(9):767-776.
41. Mintzer J, Tune L, Breder C, Swanink M, Marcus R. Aripiprazole for the Treatment of Psychoses in Institutionalized Patients With Alzheimer Dementia : A Multicenter, Randomized, of Three Fixed Doses. Am J Geriatr Psychiatry. 2007;(November):918-931.
42. Zhong K, Tariot P, Mintzer J, Minkwitz M, Devine N. Quetiapine to Treat Agitation in Dementia : A Randomized, Double-Blind, Placebo-Controlled Study. Curr Alzheimer Res. 2007;4(1):81-93.
43. Streim J, Porsteinsson A, Breder C, Swanink R, Marcus R. Placebo-Controlled Study of Aripiprazole for the Treatment of Psychosis in Nursing Home Patients with Alzheimer Disease. Am J Geriatr Psychiatry.

2008;(July):537-550.
44. RIS-INT-83. Janssen Pharmaceutical. 2003.
45. Sugerman A, Williams B, Alderstein A. Haloperidol in the psychiatric disorders of old age. Clin notes. 1964:1190-1192.
46. Janssen Pharmaceutical. Clinical study report synopsis. 2005;9–12.
47. Otsuka Pharmaceutical. 2017a. www.clinicaltrailsregister.eu. (EudraCT Number 2013-000504-41).
48. Otsuka Pharmaceutical. 2017b. www.clinicaltrailsregister.eu. (EudraCT Number 2013-000503-17).
49. Ballard C, Banister C, Khan Z, Cummings J, Demos G CB. Evaluation of the safety, tolerability, and efficacy of pimavanserin versus placebo in patients with Alzheimer's disease psychosis: a phase 2, randomised, placebo-controlled, double-blind study. Lancet Neurol. 2018;17(3):213-222.
50. ACADIA Pharmaceutical. 2018. www.clinicaltrails.gov. (NCT02992132).
51. Rada R, Kellner R. Thiothixene in the Treatment of Geriatric Patients With Chronic Organic Brain Syndrome. J Am Geriatr Soc. 1976;24(3):105-107.
52. Barnes R, Veith R, Okimoto J, Raskind M, Gumbrecht G. Efficacy of Antipsychotic Medications in Behaviorally Disturbed Dementia Patients. Am J Psychiatry. 1982;139(9):1170-1174.
53. Petrie W, Ban T, Berney S, et al. Loxapine in Psychogeriatrics: A Placebo- and Standard-Controlled Clinical Investigation. J Clin Psychopharmacol. 1982;2(2):122-126.
54. Stotsky B. Multicenter Study Comparing Thioridazine with Diazepam and Placebo in Elderly, Nonpsychotic Patients with Emotional and Behavioral Disorders. Clin Ther. 1984;6(4):546-559.
55. Finkel S, Lyons J, Anderson R, Sherrell K. A Randomized, Placebo-controlled Trial of Thiothixene in Agitated, Demented Nursing Home Patients. Int J Geriatr Psychiatry. 1995;10(June 1994):129-136.
56. Satterlee W. Olanzapine versus Placebo in the Treatment of Patients with Psychosis Associated with Dementia (F1D-MC-HGAO). 1995.
57. Borenstein M, Hedges L, Higgins J, Rothstein H. Subgroup Analyses. In: Introduction to Meta-Analysis. Chichester, UK: John Wiley & Sons, Ltd; 2009.
58. Berger V, Rezvani A, Makarewicz V. Direct effect on validity of response run-in selection in clinical trials. Control Clin Trials. 2003;24(2):156-166.
59. Hernán M, Hernández-Díaz S, Robins J. A Structural Approach to Selection Bias. Epidemiology. 2004;15(5):615-625. doi:10.1097/01.ede.0000135174.63482.43
60. Farlow MR, Shamliyan TA. Benefits and harms of atypical antipsychotics for agitation in adults with dementia. Eur Neuropsychopharmacol. 2017;27(3):217-231. doi:10.1016/j.euroneuro.2017.01.002
61. Wang J, Yu J, Wang H, et al. Pharmacological treatment of neuropsychiatric symptoms in Alzheimer ' s disease : a systematic review and meta-analysis. J Neurol Neurosurg Psychiatry. 2014.

62. Lee P, Gill S, Freedman M, Bronskill S, Hillmer M, Rochon P. Atypical antipsychotic drugs in the treatment of behavioural and psychological symptoms of dementia: systematic review. BMJ. 2004;329(June):1-5.
63. Tan L, Tan L, Wang H, et al. Efficacy and safety of atypical antipsychotic drug treatment for dementia : a systematic review and meta-analysis. Alzheimers Res Ther. 2015;7(20):1-13.
64. Ma H, Huang Y, Cong Z, et al. The Efficacy and Safety of Atypical Antipsychotics for the Treatment of Dementia: A Meta-Analysis of Randomized Placebo-Controlled Trials. J Alzheimer's Dis. 2014;42:915-937.
65. Seitz D, Gill S, Herrmann N, et al. Pharmacological treatments for neuropsychiatric symptoms of dementia in long-term care: a systematic review. Int Psychogeriatrics. 2013;25(2):185-203.
66. Maher A, Maglione M, Bagley S, et al. Efficacy and Comparative Effectiveness of Atypical Antipsychotic Medications for Off-Label Uses in Adults. J Am Med Assoc. 2011;306(12):1359-1370.
67. Cheung G, Stapelberg J. Quetiapine for the treatment of behavioural and psychological symptoms of dementia (BPSD): a meta-analysis of randomised placebo-controlled trials. N Z Med J. 2011;124(1336):39-50.
68. Lonergan E, Luxenberg J, Colford J, Birks J. Haloperidol for agitation in dementia (Review). Cochrane Libr. 2010;(12):1-42.
69. Carson S, McDonagh M, Peterson K. A systematic review of the efficacy and safety of atypical antipsychotics in patients with psychological and behavioral symptoms of dementia. J Am Geriatr Soc. 2006;54(2):354-361.
70. Ballard C, Waite J, Birks J. Atypical antipsychotics for aggression and psychosis in Alzheimer ' s disease (Review). Cochrane Libr. 2006;(1):1-135.

Appendix

Search string

In title/ abstract: chlorpromazine OR levomepromazine OR promazine OR acepromazine OR triflupromazine OR cyamemazine OR chlorproethazine OR dixyrazine OR fluphenazine OR perphenazine OR prochlorperazine OR thiopropazate OR trifluoperazine OR acetophenazine OR thioproperazine OR butaperazine OR perazine OR periciazine OR thioridazine OR mesoridazine OR pipotiazine OR haloperidol OR trifluperidol OR melperone OR moperone OR pipamperone OR bromperidol OR benperidol OR droperidol OR fluanisone OR oxypertine OR molindone OR sertindole OR ziprasidone OR lurasidone OR flupentixol OR clopenthixol OR chlorprothixene OR tiotixene OR zuclopenthixol OR fluspirilene OR pimozide OR penfluridol OR loxapine OR clozapine OR olanzapine OR quetiapine OR asenapine OR clotiapine OR sulpiride OR sultopride OR tiapride OR remoxipride OR amisulpride OR veralipride OR levosulpiride OR prothipendyl OR risperidone OR mosapramine OR zotepine OR aripiprazole OR paliperidone OR iloperidone OR cariprazine OR brexpiprazole OR pimavanserin

AND

Anywhere: trial

AND

Anywhere: dementia

Table 1. Use of run-in periods and the clinical outcomes of antipsychotics in placebo-controlled trials with conventional OR/AND atypical group

	Trials with conventional OR atypical group							
	Efficacy		**Somnolence**		**EPS**		**Mortality**	
	SMD	N	OR	N	OR	N	OR	N
Conventional antipsychotics								
No run-in	-	0	29.1 (1.5; 578.7)	1	12.4[b] (0.6; 246.1)	1	1.0 (0.1; 8.6)	2
With run-in	-0.424[a] (-0.754; -0.095)	7	3.1 (1.3; 7.1)	2	3.2 (1.4; 7.6)	2	0.7 (0.2; 2.3)	9
Atypical antipsychotics								
No run-in	-0.056 (-0.208; 0.097)	3	4.9 (2.4; 9.9)	2	2.2 (0.7; 6.6)	3	1.7 (0.7; 4.1)	3
With run-in	-0.143 (-0.206; -0.081)	17	2.4 (1.9; 3.1)	12	1.7 (1.3; 2.2)	10	1.6 (1.1; 2.4)	17

	Trials with conventional AND atypical group							
	Efficacy		**Somnolence**		**EPS**		**Mortality**	
	SMD	N	OR	N	OR	N	OR	N
Conventional antipsychotics								
No run-in	-0.389 (-0.669; -0.110)	1	1.2 (0.4; 3.1)	1	2.4 (1.2; 4.6)	1	2.1 (0.2; 23.1)	1
With run-in	-0.310 (-0.505; -0.115)	2	7.5 (2.5; 23.0)	2	3.0 (1.8; 5.0)	2	1.5 (0.6; 3.6)	2
Atypical antipsychotics								
No run-in	-0.387 (-0.665; -0.109)	1	1.0 (0.4; 2.8)	1	0.9 (0.4; 1.8)	1	1.0 (0.1; 16.4)	1
With run-in	-0.072 (-0.266; 0.122)	2	5.0 (2.4; 10.6)	2	1.1 (0.6; 2.0)	2	0.5 (0.1; 1.5)	2

EPS= Extrapyramidal symptoms; SMD= Standardized mean difference; OR=Odds ratio;

[a]random effects model was used

[b]trials with an atypical drug, a conventional drug and placebo group.

Table 2. Trials with run-in in meta-analyses of antipsychotics in dementia

Meta-analysis	Included trials, n	Trials with run-in, n (%)
Farlow 2017[55]	16[a]	13 (81)
Tan 2015[57]	10	7 (70)
Ma 2014[58]	16	12 (75)
Wang 2014[56]	6	5 (83)
Seitz 2013[59]	14	12 (86)
Maher 2011[60]	13	13 (85)
Cheung 2011[61]	5	2 (40)
Lonergan 2010[62]	5	4 (80)
Carson 2006[63]	7	6 (86)
Ballard 2006[64]	15	12 (80)
Lee 2004[65]	5	5 (100)

[a]Individual placebo-controlled start (not extension or stop) trials deduced from references

Chapter 6.

Baseline imbalances and clinical outcomes of atypical antipsychotics in dementia: A meta-epidemiological study of randomized trials

Int J Methods Psychiatr Res. 2019 Mar;28(1):e1757

Hulshof, T.A.
Zuidema, S.U.
van Meer, P.J.K
Gispen-de Wied, C.C.
Luijendijk, H.J.

Abstract

Objectives

To assess baseline imbalances in placebo-controlled trials of atypical antipsychotics in dementia, and their association with neuropsychiatric symptoms (NPS), extrapyramidal symptoms (EPS) and mortality.

Method

We searched for trials in multiple sources. Two reviewers extracted baseline characteristics and outcomes per treatment group. We calculated direction, range, pooled mean, and heterogeneity in the baseline differences, and used meta-regression for the relationship with the outcomes.

Results

We identified 23 trials. Baseline type of dementia, cognitive impairment and NPS were poorly reported. The drug group had a higher mean age than the placebo group in 9 trials and lower mean age in 3 trials (p=.073). The difference in percentage men between the drug and placebo group ranged from -9.7% to 4.4%. There were no statistically significant pooled baseline differences, but heterogeneity was present for age. Higher mean age at baseline in the drug versus placebo group was significantly associated with greater reduction in NPS, and higher percentage of non-white persons with lower risk of EPS. Imbalances were not significantly associated with risk of mortality.

Conclusion

Randomized trials of atypical antipsychotics in dementia showed baseline imbalances that were associated with higher efficacy and lower risk of EPS for atypical antipsychotics versus placebo.

Introduction

Randomization is the cornerstone of clinical trials. Randomization is used to ensure that chance instead of patient characteristics determine treatment assignment. In daily medical practice, patient characteristics affect the choice of treatment and will therefore be distributed unevenly across treatment groups. Moreover, these patient characteristics may also affect prognosis independently of the effect of treatment. Hence, treatment groups of a trial need to be comparable to establish an unbiased effect of a treatment on prognosis (clinical outcomes).

The larger the sample size of a randomized trial is, the smaller the differences between treatment groups at the start of the trial will become and the more comparable the groups will be. Yet, imbalanced groups can occur by chance despite adequately designed and conducted randomization procedures, especially in trials with small sample sizes. In addition, flawed or corrupted randomization procedures can give rise to systematic baseline imbalances between groups.[1,2] If the baseline imbalances are distributed in the same way across trials testing the same treatment, they will bias the pooled results of those trials too.[3–8]

In a previous meta-analysis, we observed baseline imbalances in trials testing antipsychotic drugs in dementia.[9] Atypical antipsychotic have been found to reduce neuropsychiatric symptoms (NPS) and to increase the risk of extrapyramidal symptoms (EPS) and mortality.[10,11] In some trials of atypical antipsychotics, the baseline characteristics of the atypical group seemed unfavorable in comparison with placebo:[12–14] patients were older, and more often men or diagnosed with vascular dementia. These factors are predictive of EPS and death in patients with dementia.[15–19] Therefore, not the atypical drugs themselves but the vulnerability of the patients in the drug compared to the placebo groups could have resulted in a higher risk of EPS and deaths. Moreover, if the more vulnerable patients had more severe NPS and dropped out more often, the remaining group of patients would have had less NPS. Consequently, the effect of atypical antipsychotics on symptom reduction might have overestimated.[20]

Atypical antipsychotics were introduced to the market with the claim that these drug were as effective as haloperidol but had less side-effects.[21–23] At the time, haloperidol – a typical antipsychotic - was the first choice of treatment for agitation and psychosis in dementia. To substantiate the claim of relative benefits and harms, atypical antipsychotics and haloperidol have been compared with placebo simultaneously in a number of trials. We observed that in trials with an haloperidol group, the atypical antipsychotics groups seemed to be less vulnerable than the placebo groups.[9,21–23] This imbalance might have led to overestimation of the effect

of atypical antipsychotics on reduction of NPS and underestimation of the risk of EPS compared to placebo. The variation in baseline imbalances between atypical drug groups and placebo groups across trials enables an evaluation of the effect that the imbalances might have had on trial results. The aim of this study was (1) to assess the presence of systematic baseline imbalances in placebo-controlled trials of atypical antipsychotics in dementia, and (2) to evaluate the association of baseline imbalances with reduction of NPS and risk of EPS and mortality.

Methods

Literature search and selection

Two reviewers (TAH, HJL) searched trials in four sources. First, we used the electronic databases Pubmed, Cinahl and Embase and entered the strings [‘*generic name atypical antipsychotic’ AND trial*] and [*dementia*]. We had composed a list of all atypical antipsychotics from the websites of the World Health Organization and the Food and Drug Administration to this end.[24,25] Secondly, we hand-searched the references of published systematic reviews, which were identified with the same electronic databases and the Cochrane library. Titles and abstracts of potentially eligible studies were retrieved in Pubmed. Thirdly, we sought RCTs in trial registration websites with the keywords (‘*new generation*’, ‘*second generation’* or ‘*atypical*’) and *‘antipsychotic’*. Finally, we searched the databases of the Dutch Medicines Evaluation Board and the FDA for unpublished trials. If studies seemed potentially eligible given title and abstract, full articles of published studies were retrieved as well as online protocols of unpublished studies. Two reviewers (TAH, HJL) reviewed these articles for definitive eligibility. Randomized placebo-controlled trials that reported the effect of orally administered atypical antipsychotics on neuropsychiatric symptoms or mortality in participants with Alzheimer’s disease or vascular dementia were included. Studies in patients with Lewy body or Parkinson’s dementia were excluded, because they are much less tolerant for antipsychotics, as were studies with multiple drugs in one intervention arm. There were no restrictions with respect to publication date, language, flexible or fixed dosing of the active treatment and duration of the study. The search was last rerun in August 2017.

Data extraction

Two reviewers (TAH, HJL) independently extracted data from the included studies. First, we extracted general study characteristics: setting, type of dementia, comparison groups, study duration, number of randomized patients in each treatment group, and publication status (published full article, or unpublished) and commercial funding. We assessed the randomization procedures consisting of the random sequence generation and allocation concealment, defined and scored as

having a low, unclear or high risk of bias in accordance with the Cochrane Risk of Bias assessment tool.[27] We also recorded whether information about baseline characteristics of the treatment group for *all* randomized patients was presented in a baseline table in line with CONSORT requirement.[28]

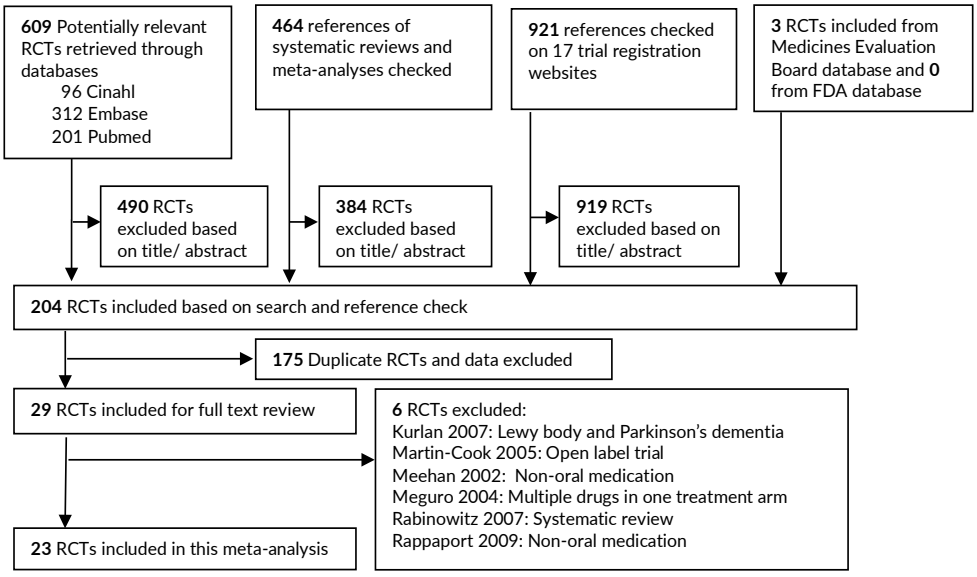


Figure 1. Flow diagram of literature search and study selection

Secondly, we extracted the baseline characteristics of *all* randomized participants in the atypical antipsychotic and placebo groups: mean age and standard deviation (SD), number of men, number of non-white persons, number of vascular/mixed dementia, mean severity of dementia and SD, mean severity of neuropsychiatric symptoms (NPS) and SD. For severity of dementia, we used the mean Mini Mental State Examination (MMSE) score because it was the most frequently reported instrument (see appendix table 3). For severity of NPS, we recorded the reported mean NPI(-NH), BEHAVE-AD, BPRS, BRSD or Neurobehavioral rating scale score. In case of multiple reported generic instruments, we preferred the most commonly used NPI(-NH), or otherwise BEHAVE-AD. Other potentially important prognostic baseline characteristics, such as (cardiovascular) comorbidity, somatic and psychiatric medication use, and extrapyramidal symptoms were reported too infrequently to be of use for our analyses (see appendix table 1). When studies included multiple active medication groups (different dosages or drugs), an average mean or percentage was calculated for these groups. If the SD for mean age, MMSE, and NPS was missing, the SD was imputed with the average SD of the other trials.

Table 1 Characteristics of randomized placebo-controlled trials that tested atypical antipsychotics in dementia

Author, Year	Antipsychotic drug	Setting	Type of dementia
ZIP-128-105, 1993	Ziprasidone	Nursing home	AD-VAS
Satterlee, 1995	Olanzapine	Nursing home	AD
Ris-Bel-14, 1997ŀ	Risperidone	NR	AD
Ris-Int-83, 1997ŀ	Risperidone	NR	AD
De Deyn, 1999	Risperidone Haloperidol	Nursing home	AD-VAS-MIX
Katz, 1999	Risperidone	Nursing home	AD-VAS-MIX
Allain, 2000	Tiapride Haloperidol	Nursing home & Hospital	AD-VAS-MIX
Street, 2000	Olanzapine	Nursing home	AD
Herz, 2002	Risperidone Olanzapine	NR	AD
ILO522, 2002	Iloperidone	NR	AD-VAS-MIX
Brodaty, 2003	Risperidone	Nursing home	AD-VAS-MIX
De Deyn, 2004	Olanzapine	Nursing home	AD
Ballard, 2005	Quetiapine	Nursing home	AD
De Deyn, 2005	Aripiprazole	Outpatients	AD
Deberdt, 2005	Risperidone Olanzapine	Nursing home & Outpatients	AD-VAS-MIX
Kennedy, 2005	Olanzapine	Outpatients	AD (no NPS)
Mintzer, 2006	Risperidone	Nursing home	AD-VAS
Schneider, 2006	Risperidone Olanzapine Quetiapine	Outpatients	AD
Tariot, 2006	Quetiapine Haloperidol	Nursing home	AD
Mintzer, 2007	Aripiprazole	Nursing home	AD
Zhong, 2007	Quetiapine	Nursing home	AD-VAS
Paleacu, 2008	Quetiapine	NR	AD
Streim, 2008	Aripiprazole	Nursing home	AD

AD=Alzheimer Disease; NPS=neuropsychiatric symptoms; NR= Not reported; VAS=Vascular Dementia; Mix=Mixed Dementia; ŀ mortality data were published in Haupt, 2006[26]; °Trial with conference abstracts only were considered as unpublished. ᶲGroups. *Doctors were allowed to stop medication if deemed inefficient or causing too much side-effects.

N randomized	Duration, weeks	Dose, range mg/d	Published°
23	4	2 – 6	No
238	8	1 – 8	No
39	4	1 – 4	No
18	8	0.5 – 1.5	No
344	12	0.5 – 4 0.5 – 4	Yes
625	12	0.5, 1, 2$^{\phi}$	Yes
306	3	100 – 300 2 – 6	Yes
206	6	5, 10, 15$^{\phi}$	Yes
29	10	0.5 – 4 2.5 – 20	No
15	4	0.5 – 6	No
345	12	0.25 – 2	Yes
652	10	1, 2.5, 5, 7.5$^{\phi}$	Yes
62	6	50 – 100	Yes
208	10	2 – 15	Yes
494	10	0.5 – 2 2.5 – 10	Yes
268	26	2.5 – 7.5	Yes
473	8	0.5 – 1.5	Yes
421	2-36	0 – 2* 0 – 17.5 0 – 200	Yes
284	10	25 – 600 0.5 – 12	Yes
487	10	2, 5, 10$^{\phi}$	Yes
333	10	100, 200$^{\phi}$	Yes
40	6	75 – 300	Yes
256	10	0.7 – 15	Yes

Finally, we extracted the clinical outcomes. Efficacy of antipsychotics in dementia is usually measured with a generic instrument for diverse neuropsychiatric symptoms (e.g. NPI, BEHAVE-AD) or with an instrument specific for one type of symptoms such as aggression (CMAI). We preferred the reported total score of a generic instrument to guarantee comparability of outcomes across trials, but if it was lacking we used the reported total score of the specific instrument. If multiple generic instruments were used, we extracted the most commonly reported (NPI(-NH) or otherwise BEHAVE-AD). We extracted the mean change from baseline to end point with its SD for the active drug and placebo groups. If the confidence interval, standard error, or p-value was reported, the SD was calculated with this information (Cochrane handbook). When multiple dosage or multiple drug groups were included in a trial, an average change was calculated. We also recorded the number of patients with EPS and the number of deaths during the trial.

For all extracted information, the published article of a trial was our primary source. Authors provided additional information at our request[21,29–32] and meta-analyses published by industry were our secondary source. Other articles and meta-analyses were our tertiary source of information.[33–40] The reviewers discussed differences in the extracted data until consensus was reached.

Statistical analyses

First, we plotted the difference between group sizes (drug versus placebo) against total trial size for 17 trials with unrestricted randomization (see appendix table 1), and the expected distributions for the 50% and 95% prediction intervals.[41] For trials with more than two active drug groups, we used the first reported active drug group (and a total trial size of placebo group plus first active drug group). For studies that used a randomization ratio other than 1:1 for placebo versus active drug group, we recalculated the size of the active drug group by dividing the true size by the inverse of the ratio, and then re-calculated the hence found difference back to the original total trial size. We then plotted this difference against to the true total number of participants. Trials that reported blocked randomization were excluded from this analysis.

Secondly, we described the range and direction of the baseline imbalances for studies with and without haloperidol groups. The rationale for this distinction is that studies with a haloperidol group seem to suggest higher efficacy, lower risk of EPS, and lower risk of mortality than trials without a haloperidol group (see appendix figures 1-3). We then computed a one-sided sign-test per characteristic to test whether the proportion of studies that reported an imbalance in the most common direction (e.g. higher mean age in antipsychotic versus placebo group) was higher than can be expected by chance (50%). Studies that reported no difference between groups (e.g.

same mean age, which could be due to rounding) and studies with a missing baseline difference are automatically discarded from a sign-test.

Thirdly, we performed meta-analyses to calculate the pooled mean difference (MD) for baseline age, severity of dementia and severity of NPS, and the pooled odds ratio for men, non-white persons, and vascular/mixed dementia with fixed-effects models.[6,42,43] We expected a common effect estimate of zero in these mean baseline variables. Again, we distinguished between studies with and without haloperidol groups. The analyses generate an I^2-statistic for heterogeneity. We calculated 95% confidence intervals around I^2 with the direct command heterogi in Stata.

Fourthly, we performed meta-regression analyses to assess the relationship of the individual baseline imbalances for *all* randomized patients with reduction in NPS, risk of EPS and risk of mortality. The beta-coefficients (betas) were calculated with 95% confidence intervals. We estimated the SMD for NPS reduction and odds ratio (OR) for risk of EPS and mortality. As many trials reported no deaths in one or both treatment groups, we used the Mantel-Haenszel weighted fixed effects model with continuity correction based on the reciprocal of the opposite group arm size to calculate the pooled ORs.[44] A fixed-effects model was applied when heterogeneity (I^2) was found to be below 40%, otherwise a random-effects model (DerSimonian & Laird model with the estimate of heterogeneity being taken from the from the Mantel-Haenszel model).[45] The plot of group size difference against total trial size was made in R,[46] the other analyses were performed with Stata version 14.1.[47]

When we found that a large number of baseline differences were not reported, we decided to pool the outcomes of studies reporting and not reporting a baseline characteristic in a post-hoc analysis. Given the discrepancy in results of trials with and without haloperidol group this analysis was restricted to the latter type of trials.

Results

Our search yielded 1997 potentially relevant RCTs (Figure 1). We obtained the reports of 29 RCTs for full text review and finally identified 23 eligible RCTs with 5853 participants.[13,14,21–23,30,31,48–61] Five trials were relatively small and unpublished (Table 1).[48,50,51,60,61] Twenty trials investigated one atypical antipsychotic drug, three of which included an extra haloperidol group,[21–23] and three trials investigated multiple atypical drugs. The follow-up was <= 12 weeks in 21 trials, and >= 26 weeks in 2 trials. All trials were sponsored completely or partly by industry; one trial did not report the source of funding.[50]

No study described the randomization procedure completely in terms of both the random sequence generation and allocation concealment (see appendix table 1 and table 2). Baseline characteristics were also poorly reported. Only thirteen studies presented a baseline table or baseline information in the text for *all* randomized patients, two studies for a selection of all randomized patients, and eight studies, including four published studies, did not present a baseline table or baseline information in the text (see appendix table 3). Only three trials reported all six patients characteristics. The first author of two trials provided additional data.[21,32] For another trial, we calculated missing baseline information with the provided IPD.[30]

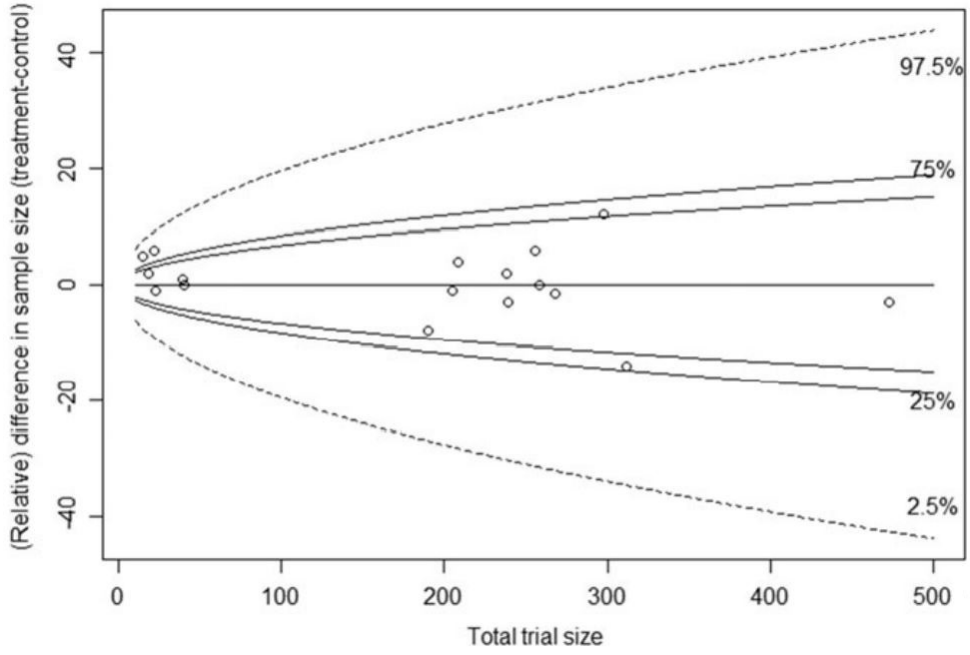


Figure 2. Treatment versus control group size differences and total trial size for 17 unrestricted trials, with expected distributions for 50% and 95% (dotted) prediction intervals.

Figure 2 presents the relation between the difference between group sizes (placebo versus active) and the total sample size for 17 trials with unrestricted randomization, together with the expected distributions for the 50% and 95% prediction intervals. Less than expected trials were outside the 50% and 95% distribution lines, four (24%) and zero (0%) respectively.

Table 2 shows the range of the actual differences between the placebo and treatment groups for each of the baseline characteristics. The percentage men and vascular/

mixed dementia showed imbalances ranging from -9.7% to 4.4% and -9.9% to 2.7% respectively. One trial showed a difference in NPS of 0.271. Baseline imbalances that we investigated were not accounted for in the analyses of all but two trials.[29,31] Table 2 also shows the direction of the baseline imbalances. No statistical differences were found. When there was no haloperidol group, there were numerically more trials with a higher age in the antipsychotics versus placebo group (8 versus 2; p=.055), with a higher percentage of men (7 versus 5; p=.387) and a higher percentage of vascular/ mixed dementia (8 versus 3; p=.113). When combining all trials, the number of trials with a positive versus negative direction was numerically higher for age (9 versus 3; p= .073) and men (10 versus 5; p=.125), and lower for severity of NPS (1 versus 4; p=.188).

Table 3 presents the size and heterogeneity of the pooled baseline differences. The pooled imbalance in the percentage of men in the trials with a haloperidol group stood out numerically (5.4%), but none of the imbalances were statistically significantly different. Four of six baseline characteristics exhibited heterogeneity, when there should have been none. Heterogeneity was 70% for the baseline difference in age in the trials with haloperidol group, and 27% for the difference in severity of NPS in trials without haloperidol group. None of the confidence intervals around I^2 suggested statistically significant heterogeneity.

Table 4 presents the associations between individual baseline imbalances and the clinical outcomes. Only for age, sex and race were there more than 10 trials, the minimum for a reliable meta-regression analysis. A higher mean age, a higher percentage of men and of persons of non-white race in the atypical antipsychotic drug than the placebo group, which was more often the case than not (see Table 2), was associated with greater efficacy and lower risk of EPS. In particular, one percentage more males in the treatment versus placebo group was statistically significantly associated with a higher reduction in NPS (beta -0.027; 95% CI -0.047 to -0.006), and one percentage more non-white persons with a lower risk of EPS (beta -0.4; 95% CI -0.8 to -0.1). An association with mortality risk could be not confirmed for any of the baseline imbalances.

As half of the baseline imbalances we wanted to abstract were not reported, we pooled the clinical outcomes for trials with and without missing baseline information for each baseline characteristic. Efficacy was consistently higher and risk of EPS consistently lower in studies without baseline information than for studies with this information (Table 5). Risk of mortality was, however, lower in studies with missing age, sex and type of dementia, but higher in studies with missing race, MMSE, and severity of NPS.

Table 2. Range and direction of baseline differences between atypical antipsychotic and placebo groups

	Trials with reported characteristic or IPD for all randomized, k/n	Trials without haloperidol group			Trials with haloperidol group			All trials
Patient characteristic		Difference, range[a]	Trials, n-/n0/n+[d]	Sign test, p[e]	Difference, range[a]	Trials, n-/ n0/ n+ [d]	Sign test, p[e]	Sign test, p[e]
Age in years, mean	17/ 23	-1.9 to 2.0	2/ 4/ 8	.055	-2.0 to 1.7	1/ 1/ 1	.750	.073
Male gender, %	15/ 21[f]	-9.7 to 4.4	5/ 0/ 7	.387	2.3 to 7.3	0/ 0/ 3	.125	.151
Non-white race, %	13/ 21[f]	-4.9 to 3.1	3/ 0/ 8	.113	-0.9 to -0.0	2/ 0/ 0	.250	.291
Vascular or mixed dementia, %	5/ 13[f]	-3.3 to 2.7	1/ 0/ 3	.313	-9.9 (1 trial)	1/ 0/ 0	.500	.500
MMSE, mean[b]	8/ 23	-1.2 to 0.7	2/ 1/ 4	.344	-0.2 (1 trial)	1/ 0/ 0	.500	.500
NPS, standardized mean[c]	5/ 22[f]	-0.146 to 0.271	2/ 0/ 1	.500	-0.126 to -0.048	2/ 0/ 0	.250	.188

IPD: individual patient data; MMSE: mini-mental state examination; NPS: neuropsychiatric symptoms

[a]the baseline difference for each trial was calculated as the mean or percentage in the atypical antipsychotic group minus the mean or percentage of the placebo group

[b]MMSE can be scored between 0 and 30; higher is better

[c]NPS were measured with different instruments in the studies and therefore the mean was standardized with the SD; higher is worse.

[d]n- stands for the number of trials with a negative baseline imbalance (f.i. lower age in antipsychotic versus placebo group), and n0 for the number of trials with no

baseline imbalance (f.i. similar mean age in antipsychotic and placebo group), n+ for the number of trials with a positive baseline imbalance (f.i. higher age in antipsychotic versus placebo group),

[e]one-sided sign-test per characteristic to test whether the proportion of studies that reported an imbalance in the direction (in the most common direction) could be attributed to chance

[f]less than 23; 2 trials were performed in men only, 2 trials in white persons only, 10 trials in patients with Alzheimer disease only, and 1 trial in patients without NPS.

Table 3. Pooled baseline difference and heterogeneity in atypical antipsychotic versus placebo groups

	Trials without haloperidol group		Trials with haloperidol group		All trials	
Patient characteristic	**Pooled difference (95%CI)**	**I^2, % (95% CI)**	**Pooled difference (95%CI)**	**I^2, % (95% CI)**	**Pooled difference (95%CI)**	**I^2, % (95% CI)**
Age in years, mean	0.1 (-0.4; 0.6)	0 (0-55)	-0.2 (-1.3; 1.0)	70 (0-91)	0.1 (-0.4; 0.5)	12 (0-49)
Male gender, %	0.3 (-2.8; 3.4)	0 (0-58)	5.4 (-1.9; 12.7)	0 (0-90)	1.1 (-1.8; 3.9)	0 (0-54)
Non-white race, %	-0.1 (-2.5; 2.3)	0 (0-60)	-0.5 (-2.3; 2.0)	0 (nt)	-0.1 (-2.3; 2.0)	0 (0-57)
Vascular or mixed dementia, %	0.3 (-3.7; 4.2)	0 (0-85)	-9.9 (-21.8; 2.0)	nt (nt)	-1.0 (-4.8; 2.7)	12 (0-82)
MMSE, mean[a]	0.1 (-0.4; 0.5)	9 (0-71)	-0.2 (-1.6; 1.2)	nt (nt)	0.0 (-0.4; 0.5)	5 (0-68)
NPS, standardized mean[b]	-.120 (-.252; .011)	27 (0-92)	.013 (-.178; .205)	0 (nt)	-.077 (-.186; .031)	11 (0-81)

NPS: neuropsychiatric symptoms; nt: not testable (too few studies);

[a]MMSE can be scored between 0 and 30; higher is better.

[b]NPS were measured with different instruments in the studies and therefore the mean was standardized with the SD; higher worse.

Table 4. Relationship of individual baseline imbalances with efficacy and risk of EPS and mortality

Imbalance between atypical antipsychotic and placebo group	**Efficacy Change in SMD[a] (95% CI)**	**EPS Change in OR[a] (95% CI)**	**Mortality Change in OR[a] (95% CI)**
Age in years, mean	-0.046 (-0.123; 0.030)	-0.0 (-0.7; 0.7)	0.4 (-0.9; 1.6)
Male gender, %	-0.027 (-0.048; -0.006)	-0.1 (-0.3; 0.1)	-0.2 (-0.5; 0.1)
Non-white race, %	-0.013 (-0.049; 0.023)	-0.4 (-0.8; -0.1)	-0.1 (-0.6; 0.4)
Vascular or mixed dementia, %*	0.015 (-0.011; 0.040)	-0.1 (-0.3; 0.4)	0.2 (-0.4; 0.7)
MMSE, mean[b]	-0.116 (-0.299; 0.067)	0.9 (-2.7; 4.5)	-1.0 (-4.0; 2.1)
NPS, standardized mean[b]	0.109 (-2.093; 2.311)	-7.9 (-30.0; 14.2)	2.7 (-9.4; 14.7)

EPS: extrapyramidal symptoms; NPS: neuropsychiatric symptoms

[a]per unit increase in the baseline difference

[b]results based on less than 10 trials

Table 5. Pooled efficacy, risk of EPS and risk of mortality for trials with reported and missing baseline information[a]

Baseline characteristic	Efficacy SMD (95%CI)		EPS OR (95%CI)		Mortality OR (95%CI)	
Age in years, mean	Reported (12)	-0.102 (-0.173; -0.031)	Reported (11)	1.7 (1.3; 2.2)	Reported (14)	1.7 (1.1; 2.6)
	Missing (2)	-0.243 (-0.390; -0.095)	Missing (1)	1.6 (0.9; 2.7)	Missing (6)	1.5 (0.6; 3.4)
Male gender, %	Reported/ NA (11)	-0.100 (-0.172; -0.028)	Reported/ NA (10)	1.8 (1.3; 2.3)	Reported/ NA (14)	1.7 (1.1; 2.6)
	Missing (3)	-0.243 (-0.387; -0.100)	Missing (2)	1.3 (0.8; 2.2)	Missing (6)	1.5 (0.6; 3.3)
Non-white race, %	Reported/ NA (10)	-0.107 (-0.176; -0.038)	Reported/ NA (9)	1.8 (1.3; 2.3)	Reported/ NA (12)	1.6 (1.1; 2.5)
	Missing (4)	-0.260 (-0.430; -0.090)	Missing (3)	1.4 (0.8; 2.2)	Missing (8)	1.8 (0.8; 4.4)
Vascular or mixed dementia, %[b]	Reported/ NA (13)	-0.107 (-0.174; -0.040)	Reported/ NA (11)	1.7 (1.3; 2.2)	Reported/ NA (19)	1.7 (1.1; 2.5)
	Missing (1)	-0.383 (-0.611; -0.155)	Missing (1)	1.6 (0.9; 2.7)	Missing (1)	1.5 (0.4; 5.4)
MMSE, mean	Reported (6)	-0.079 (-0.174; 0.016)	Reported (5)	2.0 (1.4; 2.8)	Reported (7)	1.6 (0.9; 2.9)
	Missing (8)	-0.171 (-0.257; -0.084)	Missing (7)	1.3 (0.9; 1.9)	Missing (13)	1.7 (1.0; 2.8)
NPS, standardized mean	Reported/ NA (3)	-0.042 (-0.181; 0.097)	Reported/ NA (3)	2.4 (1.3; 4.5)	Reported/ NA (4)	1.2 (0.5; 2.6)
	Missing (11)	-0.152 (-0.225; -0.080)	Missing (9)	1.5 (1.2; 2.0)	Missing (16)	1.8 (1.2; 2.8)

CI: confidence interval; EPS: extrapyramidal symptoms; MMSE: mini-mental state examination; OR: odds ratio; SMD: standardized mean difference

NA stands for not applicable, because some trials were performed in men only, in white persons only, in patients with Alzheimer disease only, or in patients without NPS at baseline.

[a]only trials without an extra haloperidol group;

[b]analyses performed with random effects models

Discussion

We reviewed the randomization procedures and baseline imbalances of 23 randomized placebo-controlled trials of atypical antipsychotics in 5853 patients with dementia. All trials reported the randomization procedures incompletely, and only three trials reported the six baseline characteristics of interest for *all* randomized patients. Numerically more trials reported a higher mean age and a higher percentage of men and of non-white persons in the atypical antipsychotics group than in the placebo group. These imbalances were associated with greater efficacy and lower risk of EPS, but not with risk of mortality. Trials with missing baseline information seemed to have a more favorable pooled efficacy and lower risk of EPS than trials that reported this information.

Randomization procedures

The goal of random sequence generation and concealment of allocation is that investigators, physicians, and patients cannot foresee allocation and then change the decision or time to enroll, or change the allocation itself. If executed correctly, randomization will distribute measured and unmeasured prognostic patient characteristics randomly between groups, hence reducing bias, so that the difference in outcome can be interpreted as an effect of treatment . Baseline tables show whether randomization has led to comparable study groups at the start of individual trials. Random fluctuations will still occur, but in general, the larger the sample size of an individual trial and the larger the number of trials in a review, the smaller the baseline imbalances can expected to be.

This is one of few studies that used objective measures to address risk of bias due to baseline imbalances in trials. Assessments of randomization are usually limited to the procedures, and these assessments can vary widely.[62] E.g. using the Cochrane assessment tool, we found that 22 trials had an unclear random sequence generation and 22 trials an unclear concealment of allocation. In contrast, a Cochrane review reported that only four trials had unclear concealment of allocation.[39] Yet another review found that 100% of trials scored 'high quality' on the Jadad and Van Tulder scale, and 90% on the Brown scale.[11]

We compared the true with expected group size difference and found that the distribution of differences was substantially smaller than could be expected by chance: 76% instead of 50% of the differences fell inside the 50% prediction interval.

Baseline imbalances

CONSORT requires trial articles to present baseline tables for *all* randomized patients. We found four published trials that did not present a baseline table at all. Only a limited number of trials reported the six baseline characteristics we studied. Other characteristics that are likely to predict efficacy or adverse events, such as comorbidity and medication use, were also missing in many articles. Baseline information might not have been missing at random either. In our study, we found that trials with missing information had a more favorable pooled efficacy and risk of EPS than trials that provided the baseline information for each of the six characteristics. Selective reporting is a common problem in the medical scientific literature,[27] and missing information on prognostic baseline characteristics might be another example.

In the articles with baseline information, most of the imbalances seemed small but some were large and obviously clinically relevant. For example, in one study 30% of the participants receiving risperidone had vascular/ mixed dementia versus 41% of the placebo group.[21] The baseline imbalances that we investigated were not accounted for in the analyses of all but two trials.[29,31]

Our next step was to pool the baseline differences and assess heterogeneity, a method recently developed to quantify baseline differences.[2,63] None of the pooled baseline differences we studied were statistically significant from zero. Some baseline differences showed considerable heterogeneity: the difference in mean age in trials with a haloperidol group (70%) and that in severity of NPS in trials without haloperidol group (27%). Heterogeneity for three characteristics was slight (between 5% and 11%). Perhaps, this amount of heterogeneity in baseline imbalances could be considered substantial as well, given that minimal heterogeneity is expected with an appropriate randomization design and conduct.

To quantify baseline imbalances, we also studied whether a positive or negative direction was more common. We found that numerically more trials reporting a higher mean age, higher percentage of men and lower severity of NPS, the primary focus of treatment in the trials, in the atypical antipsychotics group than in the placebo group. Others have suggested that imbalances in age and the primary outcome at baseline could be a good start when studying baseline imbalances.[2] We would like to add baseline imbalance in sex, and also differentiate between trials with and without a treatment arm with the old (patent free) competitor drug.

Clinical outcomes

After assessing the presence of baseline imbalances, we investigated whether they

might have affected the clinical outcomes of the trials. We found that higher mean age, higher percentage of men, and higher percentage of non-white persons at baseline in the antipsychotic than the placebo group was associated with higher efficacy. For the baseline imbalance in sex this was a statistically significant effect. Higher mean age, higher percentage of men and higher percentage of non-white persons at baseline was also associated with a lower risk of EPS. For the baseline imbalance in race this was a statistically significant effect. The effect of the baseline differences on risk of mortality was not so consistent but this was not a targeted outcome either. To our knowledge, there are no other studies that used this approach. In addition, we found a consistent pattern of studies with missing baseline information having more favorable efficacy results and a lower risk of EPS on average. Naturally, the same studies with missing information having been pooled for each of the six baseline characteristics might partly underlie this finding. Again, the pattern was not consistent for the risk of mortality.

Strengths and limitations

This is one of few studies that quantified baseline imbalances in trials. In addition, we performed an extensive literature search to identify unpublished studies. We hypothesized that baseline imbalances were related to outcomes, and hence the imbalance might depend on the publication status of a study. We used FDA and EMA databases amongst other literature sources.[64] The result was that we found six unpublished trials in addition to those included in previous meta-analyses.[10,11] As these were small studies and some did not report all outcomes, efficacy, risk of EPS and risk of mortality for atypical antipsychotics versus placebo were not substantially different from those published before.[10,11]

A limitation of our study is that our analyses depended on the amount of baseline information provided in the articles. Information on type of dementia, MMSE and severity of NPS was often lacking. Power of our study might have been insufficient to detect relevant baseline imbalances and associations of these baseline imbalances with clinical outcomes.

Conclusion

Despite randomization, placebo-controlled trials of atypical antipsychotics in dementia show heterogeneous baseline imbalances. Baseline imbalances that were not taken into account might have mistakenly led to an overestimated efficacy and underestimated risk of EPS. Our findings underscore the need for adequate randomization procedures, and reporting of baseline characteristics for all randomized patients per treatment group. In addition, baseline imbalances need to be assessed objectively as part of systematic reviews.

References

1. Hróbjartsson A, Boutron I, Turner L, Altman D, Moher D. Assessing risk of bias in randomised clinical trials included in Cochrane Reviews: the why is easy, the how is a challenge [editorial]. Cochrane Database Syst Rev. 2013;4.
2. Clark L, Fairhurst C, Hewitt C, et al. A methodological review of recent meta-analyses has found significant heterogeneity in age between randomized groups. J Clin Epidemiol. 2014;67(9):1016-1024.
3. Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessement of trial quality in systematic reviews? Emperical study. Health Technol Assess (Rockv). 2003;7(1).
4. Schulz K, Chalmers I, Hayes R, Altman D. Empirical Evidence of Bias. Dimensions of Methodological Quality Associated with Estimates of Treatment Effects in Controlled Trials. JAMA. 1995;273:408-412.
5. Riley R, Kauser I, Bland M, et al. Meta-analysis of randomised trials with a continuous outcome according to baseline imbalance and availability of individual participant data. Stat Med. 2013;32(16):2747-2766.
6. Trowman R, Dumville J, Torgerson D, Cranny G. The impact of trial baseline imbalances should be considered in systematic reviews: a methodological case study. J Clin Epidemiol. 2007;60(12):1229-1233.
7. Corbett M, Higgins J, Woolacott N. Assessing baseline imbalance in randomised trials: implications for the Cochrane risk of bias tool. Res Synth Methods. 2014;5(1):79-85.
8. Luijendijk H, Hulshof T. Baseline differences in the SAVOR trial. Diabetes Obes Metab. 2015;17(12):1202.
9. Hulshof T, Zuidema S, Ostelo R, Luijendijk H. The Mortality Risk of Conventional Antipsychotics in Elderly Patients: A Systematic Review and Meta-analysis of Randomized Placebo-Controlled Trials. J Am Med Dir Assoc. 2015;16(10):817-824.
10. Schneider L, Dagerman K, Insel P. Risk of Death With Atypical Antipsychotic Drug Treatment for Dementia. J Am Med Assoc. 2005;294(15):1934-1943.
11. Ma H, Huang Y, Cong Z, et al. The Efficacy and Safety of Atypical Antipsychotics for the Treatment of Dementia: A Meta-Analysis of Randomized Placebo-Controlled Trials. J Alzheimer's Dis. 2014;42:915-937.
12. Ballard C, Margallo-Lana M, Juszczak E, et al. Quetiapine and rivastigmine and cognitive decline in Alzheimer's disease: randomised double blind placebo controlled trial. BMJ. 2005;330(7496):874.
13. Street J, Clark S, Gannon K, Cummings J, Bymaster F. Olanzapine Treatment of Psychotic and Behavioral Symptoms in Patients With Alzheimer Disease in Nursing Care Facilities. Arch Gen Psychiatry. 2000;57:968-976.
14. Zhong K, Tariot P, Mintzer J, Mink-

witz M, Devine N. Quetiapine to Treat Agitation in Dementia : A Randomized , Double-Blind , Placebo-Controlled Study. Curr Alzheimer Res. 2007;4(1):81-93.
15. Gambassi G, Landi F, Lapane KL, Sgadari A, Mor V, Bernabei R. Predictors of mortality in patients with Alzheimer's disease living in nursing homes. J Neurol Neurosurg psychiatry. 1999;67:59-65.
16. Fitzpatrick A, Kuller L, Lopez O, Kawas C, Jagust W. Survival following dementia onset: Alzheimer's disease and vascular dementia. J Neurol Sci. 2005;229-230:43-49.
17. Mitchell S, Kiely D, Hamel M, Park P, Morris J, Fries B. Residents With Advanced Dementia. JAMA. 2004;291(22).
18. Rountree S, Chan W, Pavlik V, Darby E, Doody R. Factors that influence survival in a probable Alzheimer disease cohort. Alzheimers Res Ther. 2012;4(3):119. doi:10.1186/alzrt119
19. Garcia-Ptacek S, Farahmand B, Kåreholt I, Religa D, Cuadrado M, Eriksdotter M. Mortality Risk after Dementia Diagnosis by Dementia Type and Underlying Factors: A Cohort of 15,209 Patients based on the Swedish Dementia Registry. J Alzheimer's Dis. 2014;41:467-477. doi:10.3233/JAD-131856
20. Hernán M, Hernández-Díaz S, Robins J. A Structural Approach to Selection Bias. Epidemiology. 2004;15(5):615-625. doi:10.1097/01.ede.0000135174.63482.43
21. De Deyn P, Rabheru K, Rasmussen A. A randomized trial of risperidone, placebo, and haloperidol for behavioral symptoms of dementia. Neurology. 1999;(53):945-956.
22. Allain H, Dautzenberg P, Maurer K, Schuck S, Bonhomme D, Gerard D. Double blind study of tiapride versus haloperidol and placebo in agitation and aggressiveness in elderly patients with cognitive impairment. Psychopharmacology (Berl). 2000;(148):361-366.
23. Tariot P, Schneider L, Katz I, et al. Quetiapine Treatment of Psychosis Associated With Dementia : A Double-Blind, Randomized, Placebo-Controlled Clinical Trial. Am J Geriatr Psychiatry. 2006;14(9):767-776.
24. World Health Organization. WHO Collaborating Centre for Drug Statistics Methodology. http://www.whocc.no/atc_ddd_index/?code=N05A. Published 2013.
25. US Food and Drug Administration. Atypical Antipsychotic Drugs Information. http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm094303.htm. Published 2013.
26. Haupt M, Cruz-Jentoft A, Jeste D. Mortality in elderly dementia patients treated with risperidone. J Clin Psychopharmacol. 2006;26(6):566-570.
27. Higgins J, Altman D, Sterne J. Chapter 8: Assessing risk of bias in included studies. In: Cochrane Handbook. The Cochrane Collaboration; 2014. https://handbook-5-1.cochrane.org/chapter_8/8_assessing_risk_of_bias_in_included_studies.htm.

28. Schulz K, Altman D, Moher D. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol. 2010;63(8):1-6. doi:10.1016/j.ijsu.2010.09.006
29. Kurlan R, Cummings J, Raman R, Thal L. Quetiapine for agitation or psychosis in patients with dementia and parkinsonism. Neurology. 2007;68(17):1356-1363.
30. Paleacu D, Barak Y, Mirecky I, Mazeh D. Quetiapine treatment for behavioural and psychological symptoms of dementia in Alzheimer's disease patients: a 6-week, double-blind, placebo-controlled study. Int J Geriatr Psychiatry. 2008;23(September 2007):393-400.
31. Schneider L, Tariot P, Dagerman K, et al. Effectiveness of Atypical Antipsychotic Drugs in Patients with Alzheimer's Disease. N Engl J Med. 2006;355(15):1525-1538.
32. De Deyn P, Jeste D, Swanink R, Kostic D, Breder C. Aripiprazole for the Treatment of Psychosis in Patients With Alzheimer's Disease. J Clin Psychopharmacol. 2005;25(5):463-467. http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00004714-200510000-00010. Accessed April 17, 2014.
33. Schneider L, Dagerman K, Insel P. Efficacy and adverse effects of atypical antipsychotics for dementia: meta-analysis of randomized, placebo-controlled trials. Am J Geriatr Psychiatry. 2006;14(March):191-210.
34. Seitz D, Gill S, Herrmann N, et al. Pharmacological treatments for neuropsychiatric symptoms of dementia in long-term care: a systematic review. Int Psychogeriatrics. 2013;25(2):185-203.
35. Cheung G, Stapelberg J. Quetiapine for the treatment of behavioural and psychological symptoms of dementia (BPSD): a meta-analysis of randomised placebo-controlled trials. N Z Med J. 2011;124(1336):39-50.
36. Maher A, Maglione M, Bagley S, et al. Efficacy and Comparative Effectiveness of Atypical Antipsychotic Medications for Off-Label Uses in Adults. J Am Med Assoc. 2011;306(12):1359-1370.
37. Lee P, Gill S, Freedman M, Bronskill S, Hillmer M, Rochon P. Atypical antipsychotic drugs in the treatment of behavioural and psychological symptoms of dementia: systematic review. BMJ. 2004;329(June):1-5.
38. Carson S, McDonagh M, Peterson K. A systematic review of the efficacy and safety of atypical antipsychotics in patients with psychological and behavioral symptoms of dementia. J Am Geriatr Soc. 2006;54(2):354-361.
39. Ballard C, Waite J, Birks J. Atypical antipsychotics for aggression and psychosis in Alzheimer ' s disease (Review). Cochrane Libr. 2006;(1):1-135.
40. Lonergan E, Luxenberg J, Colford J, Birks J. Haloperidol for agitation in dementia (Review). Cochrane Libr. 2010;(12):1-42.
41. Schulz K, Chalmers I, Grimes D, Altman D. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecolo-

gy journals. JAMA. 1994;272(2):125-128.
42. Clark L, Fairhurst C, Cook E, Torgerson D. Important outcome predictors showed greater baseline heterogeneity than age in two systematic reviews. J Clin Epidemiol. 2015;68(2):175-181. doi:10.1016/j.jclinepi.2014.09.023
43. Ebell M. Methodological concerns about studies on oseltamivir for flu. BMJ. 2013;347:f7148.
44. Sweeting M, Sutton A, Lambert P. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. Stat Med. 2004;23(9):1351-1375.
45. Deeks J, Higgins J, Altman D. Analysing data and undertaking meta-analyses. In: Cochrane Handbook for Systematic Reviews of Interventions. Vol Version 5. Version 5. The Cochrane Collaboration; 2011.
46. R Core Team. R: A language and environment for statistical computing. 2013.
47. StataCorp. Stata Statistical Software: Release 13. 2013.
48. RIS-INT-83. Janssen Pharmaceutical. 1997.
49. Katz I, Jeste D, Mintzer J, Clyde C, Napolitano J, Brecher M. Comparison of Risperidone and Placebo for Psychosis and Behavioral Disturbances Associated With Dementia: A Randomized, Double-Blind Trial. J Clin Psychiatry. 1999;60(February):107-115.
50. Herz L, Volicer L, Frankenburg F, Colon S, Kittur S. A 6-Week, Double-Blind Comparison of Olanzapine, Risperidone, and Placebo for Behavioral Disturbance in Alzheimer's Disease. J Clin Psychiatry. 2002;63(11):1065.
51. ILO522. A prospective, randomized, double-blind, placebo-controlled, flexible-dose, parallel-group, single-center study to evaluate the safety, tolerability and efficacy of iloperidone (0.5 to 6.0 mg/d given b.i.d.) compared with placebo in treating psychotic and. 2002.
52. Brodaty H, Ames D, Snowdon J, et al. A Randomized Placebo-Controlled Trial of Risperidone for the Treatment of Aggression, Agitation, and Psychosis of Dementia. J Clin Psychiatry. 2003;(February).
53. De Deyn P, Carrasco M, Deberdt W, et al. Olanzapine versus placebo in the treatment of psychosis with or without associated behavioral disturbances in patients with Alzheimer's disease. Int J Geriatr Psychiatry. 2004;19(2):115-126.
54. Deberdt W, Dysken M, Rappaport S, Feldman P, Young C, Hay D. Comparison of Olanzapine and Risperidone in the Treatment of Psychosis and Associated Behavioral Disturbances in Patients With Dementia. Am J Geriatr Psychiatry. 2005;13(August):722-730.
55. Kennedy J, Deberdt W, Siegal A, et al. Olanzapine does not enhance cognition in non-agitated and non-psychotic patients with mild to moderate Alzheimer's dementia. Int J Geriatr Psychiatry. 2005;20(11):1020-1027.
56. Mintzer J, Greenspan A, Caers I, et al. Risperidone in the Treatment of Psychosis of Alzheimer Disease: Results From a Prospective Clini-

cal Trial. Am J Geriatr Psychiatry. 2006;(March):280-291.

57. Mintzer J, Tune L, Breder C, Swanink M, Marcus R. Aripiprazole for the Treatment of Psychoses in Institutionalized Patients With Alzheimer Dementia : A Multicenter , Randomized , of Three Fixed Doses. Am J Geriatr Psychiatry. 2007;(November):918-931.
58. Streim J, Porsteinsson A, Breder C, Swanink R, Marcus R. Placebo-Controlled Study of Aripiprazole for the Treatment of Psychosis in Nursing Home Patients with Alzheimer Disease. Am J Geriatr Psychiatry. 2008;(July):537-550.
59. Satterlee W. Olanzapine versus Placebo in the Treatment of Patients with Psychosis Associated with Dementia (F1D-MC-HGAO). 1995.
60. ZIP-128-105. Third party blind, evaluation of the clinical usefulness of low dose ziprasidone for the symptomatic treatment of dementia in institutionalized patients. 1993.
61. RIS-BEL-14. Janssen Pharmaceutical. 1997.
62. Savovic J, Weeks L, Sterne J, et al. Evaluation of the Cochrane Collaboration's tool for assessing the risk of bias in randomized trials: focus groups, online survey, proposed recommendations and their implementation. Syst Rev. 2014;3(1):1-12. Systematic Reviews.
63. Clark L, Fairhurst C, Cook E, Torgerson DJ. Important outcome predictors showed greater baseline heterogeneity than age in two systematic reviews. J Clin Epidemiol. 2015;68(2):175-181. doi:10.1016/j.jclinepi.2014.09.023
64. Schroll J, Bero L. Regulatory agencies hold the key to improving Cochrane Reviews of drugs [editorial]. Cochrane Database Syst Rev. 2015;4. doi:10.1002/14651858.ED000098

Appendix

Table 1. Randomisation procedures as described verbatim in the trial article, abstract or other publication

Study	Text in article
ZIP-128-105 1993	Randomized
Satterlee 1995	Patients were randomly assigned to olanzapine (1 to 8 mg/day) or placebo.
Ris-bel-14 1997	NR
Ris-int-83 1997	NR
De Deyn 1999	After initial screening and a 1-week,single-blind, placebo washout period, eligible patients were randomly assigned to treatment with risperidone, haloperidol, or placebo according to a predefined randomization 'code generated by the Janssen Research Foundation, Beerse, Belgium. Balancing ensured that an equal number of subjects were allocated to each treatment group at each center. Patient numbers were to be assigned in consecutive order.
Katz 1999	Patients were randomly assigned according to a randomization code provided by the sponsor (Janssen Research Foundation).
Allain 2000	Patients were randomly allocated to tiapride 100 mg/day (50 mg twice a day), haloperidol 2 mg/day (1 mg twice a day) or placebo.
Street 2000	Patients meeting enrollment criteria were randomly allocated to 1 of 4 fixed-dose treatment groups (olanzapine, 5, 10, or 15 mg/d, or placebo) by the assignment of a unique kit number using a permuted block design at each investigational site (block size of 4).
Herz 2002	Randomization to treatments
ILO-522 2002	Patients were randomized to either the iloperidone or placebo treatment group at a ratio of 2:1. The study had two phases: pre-randomization and treatment. If a patient qualified for admission to the treatment phase, the Investigator contacted the pharmacist, who assigned the next sequential number from the treatment randomization list. This randomization number determined treatment assignment. The study pharmacist was provided with a randomization list containing treatment group (iloperidone or placebo) assignments in a ratio of 2 iloperidone to 1 placebo. Only the study pharmacist had access to these codes. Randomization numbers consisted of 4 digits, starting with '1001'.
Brodaty 2003	Eligible patients were randomly assigned to 1 of 2 treatment groups (risperidone or placebo) according to a randomization code that was balanced to ensure even distribution of patients in each treatment group at each center.
De Deyn 2004	Patients randomly assigned to receive Olz 1.0 or Olz 2.5 were respectively given a single 1.0mg or 2.5mg capsule of olanzapine daily (qhs) throughout the study period.
Ballard 2005	The study statistician randomly assigned patients in equal numbers to active quetiapine plus placebo rivastigmine; placebo quetiapine plus active rivastigmine; or placebo rivastigmine plus placebo quetiapine (double dummy). The allocations were computer generated with block randomisation (block sizes of three and six) with Stata software (release 7.0). The randomising clinician faxed a form to the statistician, who communicated allocation to the pharmacy, ensuring concealment.

table continues

Study	Text in article
De Deyn 2005	Eligible patients were randomized to aripiprazole 2 mg/d or placebo, administered once-daily for 10 weeks, in this multicenter, double-blind study.
Deberdt 2005	After a 3- to 14-day placebo/washout period, patients were randomly assigned, in a 2:2:1 ratio, to 10 weeks of double-blind, flexible-dose treatment with olanzapine (2.5 mg–10 mg/day), risperidone (0.5 mg–2 mg/day), or placebo.
Kennedy 2005	Patients were randomized to receive treatment with olanzapine 2.5 mg/d or placebo in a 2:1 ratio.
Mintzer 2006	After completing the run-in, patients were randomized, using a predefined code, to receive either risperidone or placebo during the subsequent eight week treatment phase. Investigators received sealed envelopes for each patient containing coded details of the treatment in this phase.
Schneider 2006	In phase 1 of the study, patients were randomly assigned under double-blind conditions to receive olanzapine, quetiapine, risperidone, or placebo in a 2:2:2:3 ratio. Randomization was performed with the use of permuted blocks of nine per site without stratification and was implemented with the use of an interactive voice-response telephone system.
Tariot 2006	We randomized participants who had either AD with psychosis or another disorder with psychosis in a 3:1 ratio [...].
Mintzer 2007	Patients were randomized to fixed doses of aripiprazole (2 mg/day, 5 mg/day, or 10 mg/day) or placebo for a 10-week period.
Zhong 2007	At baseline, participants who met enrollment criteria were randomly assigned in a 3:3:2 ratio to one of three fixed-dose treatment groups: quetiapine 200 mg/day, 100 mg/day, or placebo. The centralized randomization schedule was generated using a random block size of 8 and was created using random seed and treatment allocation ratios of 3:3:2 and maintained blinded by the sponsor's randomization group.
Paleacu 2008	Patients [...] participated in a 6-week, randomized, double-blind, placebo controlled trial.
Streim 2008	Subjects were randomized to aripiprazole or placebo.

NR: Not Reported.

Table 2. Risk of bias based on randomization procedures

	Random sequence generation	Allocation concealment	Presence of baseline table/ information per group
ZIP-128-105, 1993 (NP)	U	U	H
Satterlee, 1995 (NP/ abstract only)	U	U	L
Ris-Bel-14, 1997 (NP)	U	U	H
Ris-Int-83, 1997 (NP)	U	U	H
De Deyn, 1999	L	U	L
Katz, 1999	U	U	L
Allain, 2000	U	U	L
Street, 2000	U	U	L
Herz, 2002 (NP/ abstract only)	U	U	H
ILO522, 2002 (NP)	U	L	U
Brodaty, 2003	U	U	U
De Deyn, 2004	U	U	H
Ballard, 2005	L	U	L
De Deyn, 2005	U	U	H
Deberdt, 2005	U	U	L
Kennedy, 2005	U	U	L
Mintzer, 2006	U	U	L
Schneider, 2006	U	L	L
Tariot, 2006	U	U	L
Mintzer, 2007	U	U	L
Zhong, 2007	U	U	L
Paleacu, 2008	U	U	H
Streim, 2008	U	U	L

NP= Not Published; L= Low risk of bias; U= Unclear risk of bias; H= High risk of bias

Table 3. Baseline characteristics that were reported per treatment group of the trial

Study, year	Age	Gen-der	Race	Vascular/ mixed dementia	Severity of dementia	NPS
ZIP-128-105,1993	**Yes**	NA	**Yes**	-	-	-
Satterlee, 1995*	**Yes**	**Yes**	**Yes**	NA	MMSE/	Behave-ad/ BPRS/ CGI
Ris-Bel-14, 1997	-	-	-	NA	-	-
Ris-Int-83, 1997	-	-	-	NA	-	-
De Deyn, 1999*	**Yes**	**Yes**	**Yes**	**Yes**	**MMSE/ FAST**	**Behave-ad / CMAI / CGI**
Katz, 1999*	**Yes**	**Yes**	**Yes**	**Yes**	**MMSE / FAST**	Behave-ad
Allain, 2000	**Yes**	**Yes**	NA	-	-	**Moses** irritability/ aggressiveness subscale
Street, 2000*	**Yes**	**Yes**	-	NA	**MMSE**	NPI / BPRS
Herz, 2002	-	NA	-	NA	-	-
ILO522, 2002	Yes	Yes	Yes	-	-	-
Brodaty, 2003*	Yes	Yes	-	Yes	MMSE / FAST	CMAI° / Behave-ad
De Deyn, 2004	-	-	NA	NA	-	NPI
Ballard, 2005*	**Yes**	**Yes**	-	NA	SIB / FAST	CMAI
De Deyn, 2005	**Yes**	**Yes**	**Yes**	NA	MMSE	NPI / BPRS / CGI
Deberdt, 2005*	**Yes**	**Yes**	**Yes**	**Yes**	**MMSE**	**NPI** / BPRS / Cornell / CMAI
Kennedy, 2005*	**Yes**	**Yes**	**Yes**	NA	**ADAS-cog° / MMSE / VASS**	NA (Patients without NPS)
Mintzer, 2006*	**Yes**	**Yes**	**Yes**	**Yes**	**MMSE**	**Behave-ad / CGI**
Schneider, 2006*	**Yes**	**Yes**	**Yes**	NA	**MMSE / ADAS-cog / ADCS-ADL**	**NPI / BPRS**
Tariot, 2006*	**Yes**	**Yes**	**Yes**	NA	MMSE / PSMS	BPRS / CGI / NPI-del+hall/ MOSES
Mintzer, 2007*	**Yes**	**Yes**	**Yes**	NA	MMSE	NPI/CGI / CMAI
Zhong, 2007*	**Yes**	**Yes**	**Yes**	**Yes**	MMSE	CGI/ PANNS / NPI / CMAI
Paleacu, 2008	**Yes**	-	-	NA	**MMSE**	NPI / CGI
Streim, 2008*	**Yes**	**Yes**	**Yes**	NA	MMSE / ADCS-ADL	NPI / CGI / BPRS / Cornell / CMAI

*= had a baseline table; Yes= reported; - = Not reported in baseline table or text; Bold = reported for all randomized patients; NA= Not Applicable; Underlined = used in our analysis

EPS	Cardiovascular risk factors	Medication	Other
-	-	-	-
% EPS	Orthostatic HT/ Weight/ QTc	-	-
-	-	-	-
-	-	-	-
ESRS	-	**Previous neuroleptics**	-
% EPS	-	-	Duration of stay
% EPS	-	-	-
% EPS	**Weight loss**	-	Time from NH admission / Time from first AD symptom
-	-	-	-
-	-	-	-
% EPS	Weight	-	Ónset of dementia / onset of behavioral disturbance
-	-	-	-
% EPS	-	-	-
SAS/ AIMS/ BARS	-	-	-
% EPS	**Hachinski score**	-	Time onset symptoms
SAS/ % EPS	**Weight gain**	-	Barnes
% EPS	**BMI**	-	Age onset dementia / age onset psychosis / years in institution
% EPS	**Weight/ BMI**	**AD, AP, ChEI.**	Education / married / residence
SAS/ AIMS	Weight	ChEI, AD	Time in NH
% EPS	**Weight**	-	Age at onset AD / age at onset psychosis symptoms / prominent symptom at onset
AIMS / SAS / % EPS	**Weight / BMI**	-	Fasting glucose
AIMS / SAS	-	-	-
% EPS	**Weight**	-	Age at onset AD / age at onset psychotic symptoms / predominant psychotic symptoms

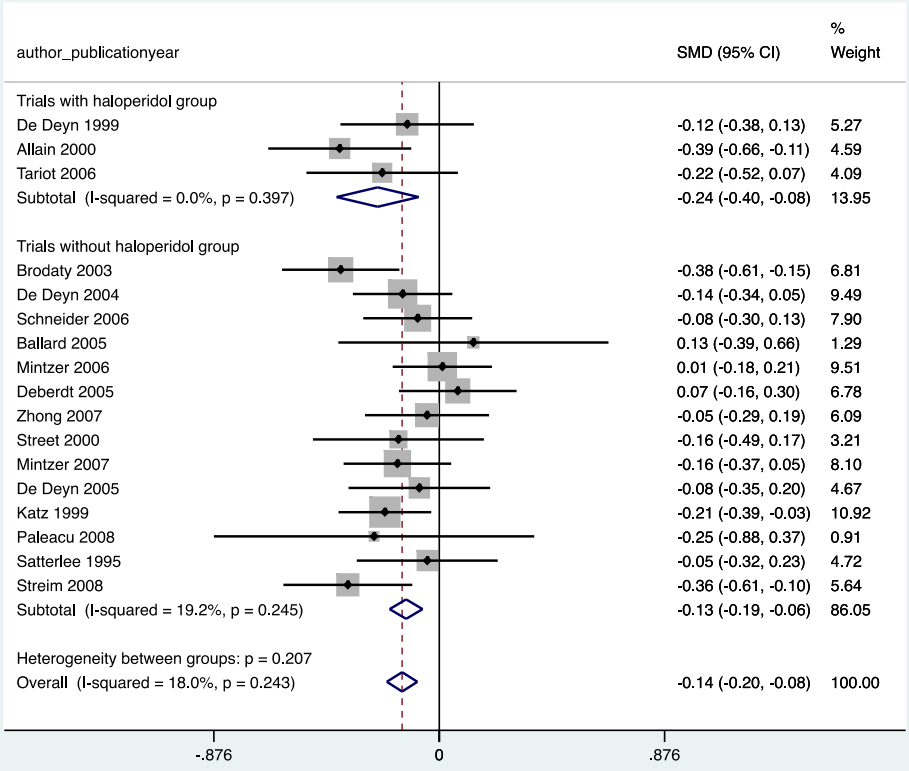


Figure 1. Pooled efficacy (SMD) of atypical antipsychotics versus placebo in dementia

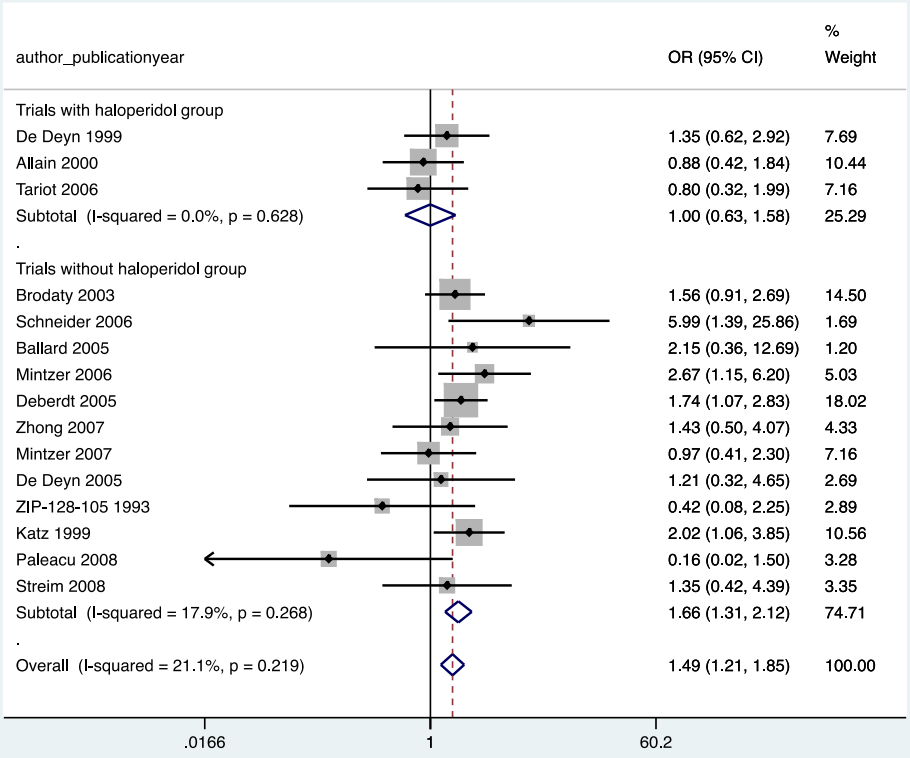


Figure 2. Risk of extrapyramidal symptoms of atypical antipsychotics versus placebo in dementia

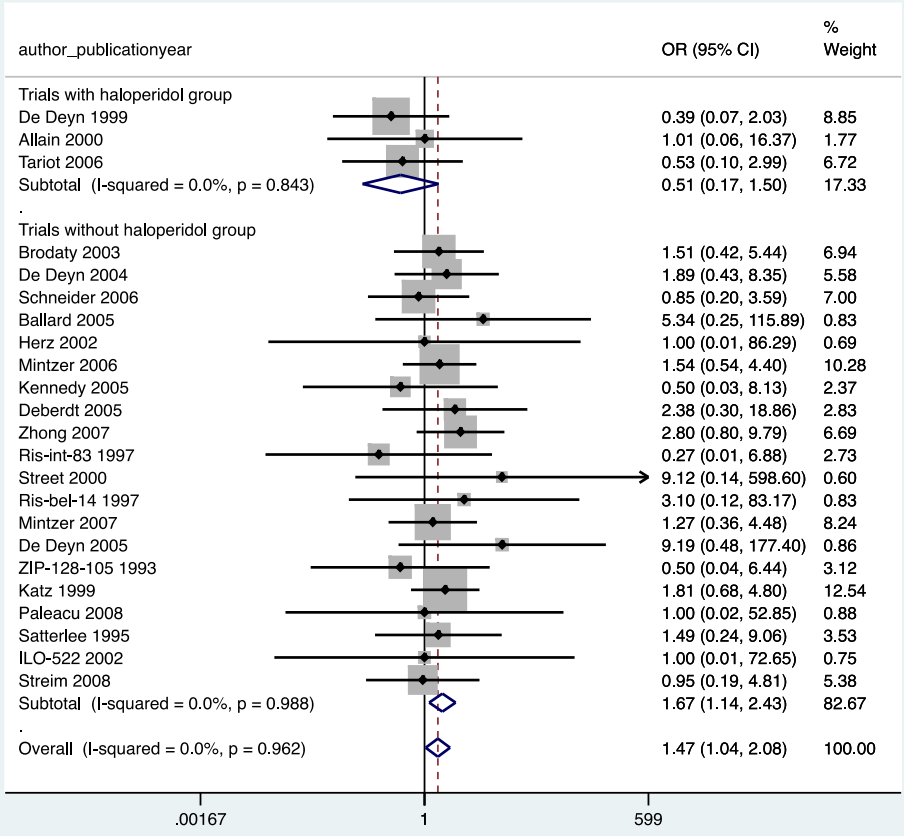


Figure 3. Pooled risk of mortality of atypical antipsychotics versus placebo in dementia

Chapter 7.

Subjective versus objective outcomes of antipsychotics for the treatment of neuropsychiatric symptoms associated with dementia

CNS Drugs. 2019 Sep;33(9):933-942

Vredeveld, E.J.
Hulshof, T.A.
Zuidema, S.U.
Luijendijk, H.J.

Abstract

Background

Knowledge about treatment status can influence effects measured in trials when subjective scales are used. The aim of this study was to compare subjective with objective outcomes of conventional and atypical antipsychotics for neuropsychiatric symptoms (NPS) in dementia.

Methods

We performed a meta-epidemiological study of 28 randomized placebo-controlled trials. For effectiveness, we used change in NPS and response rate as subjective outcomes overall dropout and additional psychotropic use were objective outcomes. For side effects, we used extrapyramidal symptoms (EPS) and somnolence as subjective outcomes, and dropout due to adverse events, medication use for EPS, and participants falling were objective outcomes.

Results

Conventional antipsychotics reduced NPS more than placebo (standardized mean difference (SMD) -0.36; 95% confidence interval (CI) -0.49 to -0.23), as did atypical antipsychotics (SMD -0.14; 95%CI -0.19 to -0.08). Response rates in the drug groups were higher too. Overall dropout did not differ between conventional antipsychotics and placebo (odds ratio (OR) 1.03; 95%CI 0.77 to 1.37) or atypical antipsychotics and placebo (OR 1.01; 95%CI 0.89 to 1.14). Additional psychotropic use did not differ either. The risk of EPS was higher for conventional (OR 2.93; 95%CI 2.04 to 4.22) and atypical antipsychotics (OR 1.52; 95%CI 1.23 to 1.88) versus placebo, as was the risk of somnolence and dropout due to side effects, but medication use for EPS and risk of falls was not.

Conclusions

Effectiveness of antipsychotics for NPS in dementia based on subjective scales was not confirmed with objective outcomes, in contrast to the increased risk of side effects.

Introduction

Doctors often prescribe antipsychotics to treat neuropsychiatric symptoms (NPS) in patients with dementia.[1,2] The prevalence of NPS is 70% to 90% in institutionalized patients with dementia.[3,4] The most common symptoms are aggression, agitation and apathy.[5] NPS have a great impact on the quality of life of patients, informal caregivers and health professionals.[6,7] More than 60% of patients with NPS use psychotropic drugs and anti-psychotics account for almost two-thirds of this use.[8,9]

Knowledge about treatment status can influence the measurement of efficacy and side effects when they are established with subjective rating scales.[10] Such measurement error can bias the trial results (information bias, observer bias). Efficacy might be overestimated, and the risk of side effects under-estimated.[11]

Trials use placebo tablets to blind participants, caregivers and assessors for treatment status to avoid measurement error.[12] However, this way of blinding is not always successful. Treatment status can sometimes be guessed, for instance if the active drug has specific side effects.[13] A systematic review about blinding in randomized controlled trials (RCTs) among psychiatric patients showed that patients in the active treatment group more often correctly guessed the treatment status than the patient receiving placebo.[10] This also applied to the investigators. In particular, a trial comparing alprazolam with placebo in patients with anxiety disorders showed that side effects were associated with correctly guessing treatment status.[14]

Likewise, in a study about the effects of caffeine on cognitive performance, false positive feedback about performance made patients believe they received caffeine pills instead of the placebo they actually received.[15] These patients had faster reaction times than patients who did not get feedback and believed to receive the placebo.[15] It is possible that the effect of antipsychotics on NPS in dementia is systematically overestimated due to partial unblinding by the specific side effects of antipsychotics, such as extra-pyramidal symptoms (EPS).

If treatment efficacy and side effects can be over- or under-estimated due to measurement error, objective outcomes may provide more valid results. Examples of objective outcome measures are the use of rescue medication and medication for side effects. In addition, overall drop-out is an objective measure of effectiveness in terms of the balance between efficacy and acceptability.[16] For instance, meta-analyses of trials comparing paroxetine against placebo for major depression have shown a statistically significant effect on depressive symptoms, but the proportion of patients who discontinued treatment for any reason was not different between the groups.[16]

In trials about antipsychotics for NPS in dementia, more objective outcome measures for efficacy and side effects are available, which may be established with less measurement error. Examples are dropout for any reason, dropout due to adverse events, and use of additional psychotropic medication, rescue medication or medication to treat EPS. The aim of our study was to assess effectiveness and side effects of antipsychotics in randomized placebo-controlled trials for NPS in patients with dementia with subjective and objective outcome measures.

Methods

Search strategy and study selection

We made a list of antipsychotics (conventional and atypical) from the websites of the World Health Organization, the American Food and Drug Administration (FDA) and Wikipedia to enable this search.[17–19] To identify trials, we used three sources. First, two authors (TAH, HJL) searched the electronic databases Pubmed, Cinahl, Embase and Cochrane library with the following key words: '*generic name of atypical/conventional antipsychotic'* and *trial* and *dementia*. We restricted the key words related to drug name to title and abstract. Second, we searched the references of published systematic reviews by hand. We identified these meta-analyses with the abovementioned electronic databases and the Cochrane library. Thirdly, we looked for RCTs in trial registration websites with the same keywords if possible; otherwise we used only the term dementia. Lastly, we searched the European Medicines Agency (EMA) and FDA websites for eligible trials. For a previous project we were also able to search for atypical antipsychotics trials in the databases of the Dutch Medicines Evaluation Board. Titles and abstracts of possibly eligible studies were retrieved from Pubmed. The last search was run in June 2019.

Randomized placebo-controlled trials that investigated the efficacy of orally administered conventional or atypical antipsychotics in patients with NPS and dementia were included. Studies with more than one drug in an intervention arm were excluded. There were no restrictions with respect to dosage, flexible or fixed dosing of the active treatment, trial duration, publication date and language. Two authors determined definitive eligibility (TAH, HJL).

Data extraction

Two authors (EV, TAH) independently extracted data on study characteristics and outcomes. Disagreements were resolved by discussion and consensus with the third author (HJL). We extracted general study characteristics, including setting, type of dementia, type of NPS (agitation, psychosis, or diverse NPS), type of antipsychotic

treatment (conventional or atypical) and the total number of randomized patients in the treatment groups. As subjective measures of effectiveness, we extracted the mean change in symptoms from baseline to the end of the trial (or endpoint if not available). Changes on symptom scales were extracted for the specific indication for which the antipsychotic was tested in the trial. For instance, if trial enrolled patients with psychosis, extracted results were specific for psychosis such as the psychosis subscale of the neuropsychiatric inventory (NPI). The standard deviation (SD) of the difference was extracted, or calculated using the p-value, t-value, or confidence interval (CI).[20] We also extracted the number of patients with a clinically relevant improvement on the subjective symptom scales (as defined by the authors) or the number of patients with any improvement on clinical rating scales. Response rates as measured with both types of scales were combined.[16] As objective measure of effectiveness, we extracted the number of patients that received new additional psychotropic medication including rescue medication during the study. The number of patients that dropped out due to any reason was used as an objective measure of acceptability.[21]

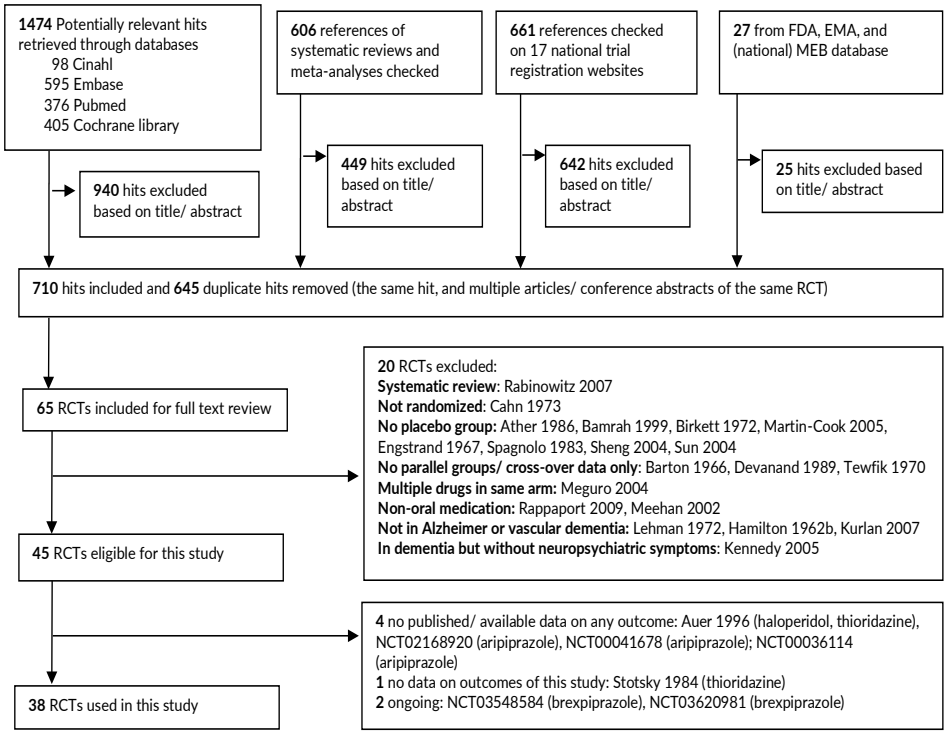


Figure 1. Flow diagram of the literature search and study selection.

EMA: European Medicines Agency; FDA: Food and Drug Administration; MEB: Medicines Evaluation Board.

In our review of side effects, we focused on EPS and somnolence, because these are prevalent and severe side effects of antipsychotics. In most studies, EPS was measured with a specific rating scale, e.g. the Simpson-Angus scale. Somnolence, also called sedation or drowsiness, was measured with a specific rating scale such as the visual analogue sedation scale, or spontaneous reports. We extracted the number of patients with EPS and somnolence measured with these subjective measurement instruments. As objective measure of side effects, we extracted the number of patients that dropped out due to adverse events, and used medication for EPS. Although the level of EPS and the use of medication for EPS are related, the distinction between these outcomes is the degree to which their measurement is sensitive to error. EPS can be rated as more or less severe than they really are, whereas the use of medication of EPS is a verifiable fact. In addition, at the request of a reviewer, we extracted the number of participants that had fallen during the study, because it is an objectively measurable outcome and potential consequence of EPS and somnolence.

When multiple intervention groups with various dosages of a drug were tested in a trial, we calculated an average of the combined groups for all outcomes. The protocol and data extraction form can be obtained with the corresponding author.

Data-analysis

First, we calculated the pooled efficacy of antipsychotics for NPS in dementia in terms of the standardized mean difference (SMD). SMDs were calculated with a 95% confidence interval (CI). Secondly, we calculated an odds ratio (OR) with a 95% CI for all other outcome measures of efficacy and side effects. We used fixed effect models if heterogeneity (presented as I^2) was lower than 40% and p-value > 0.05 for the Chi test; otherwise we used random effect models.

We performed the meta-analyses separately for conventional and atypical antipsychotics, because efficacy and risk of side effects are assumed to differ between these groups. Hence, all comparisons of a conventional drug versus placebo were pooled, and all comparisons of an atypical drug versus placebo. Therefore, the placebo group from a trial that tested both types of antipsychotics was used in both meta-analyses. To assess whether the pooled effects of conventional and atypical antipsychotics differed significantly, we used the standard error (SE) of the difference of the pooled SMD and OR between treatment groups to calculate z with the Z-formula, and then p.

Results

Study characteristics

We found 2768 hits with our search, and 66 studies seemed potentially eligible. Twenty studies were excluded for various reasons, such as only including patients with Lewy body dementia, study medication was not orally administered or the trial did not use a placebo group.[22–41] We identified 45 eligible studies, but five studies did not provide data about any of the outcomes of interest and could not be used in our meta-analyses,[42,43] and NCT02168920, NCT00041678, NCT00036114, and two were ongoing [NCT03548584, NCT03620981]. The other 38 trials with 7726 participants were included in our analyses.[21,44–80] The flowchart in figure 1 shows the results of the search.

Table 1 shows the study characteristics of the included studies. The majority of studies investigated haloperidol (8 trials),[55,71–73,75,77–79] risperidone (10 trials),[21,48–50,53,56,57,60,61,71] quetiapine (5 trials),[21,59,63,64,73] or olanzapine (7 trials),[21,47,51–53,58,60] with placebo. Three studies tested both a conventional antipsychotic and an atypical antipsychotic versus placebo.[71–73] Of the 38 studies, 12 had psychosis as indication,[44,49,73,80,55,57,58,60–62,65,69] 11 agitation,[53,56,79,59,63,67,68,70,72,76,77] 14 diverse NPS,[21,45,71,74,75,78,46–48,50–52,64,66] and one study did not report the type of NPS.[54] Twenty-four studies were performed in patients living in nursing homes,[45,46,61–63,65,67–72,48,73,74,76,49–51,56,58–60] seven in hospitals,[44,55,58,66,72,75,79] seven in outpatients,[21,68,70,77,78,80] and six studies did not report the setting.[47,52–54,57,64]

Subjective measures of efficacy were reported more frequently than objective measures (table 2). In particular, use of additional medication for NPS, use of medication for EPS and number of participants with falls were poorly reported. As a result, some of the meta-analyses yielded large confidence intervals.

Table 1. General characteristics of randomized placebo-controlled trials

Author (year)	Antipsychotic drug	Type of dementia	Type of NPS	Setting	N	Duration, weeks
Conventional antipsychotics (10 trials)						
Hamilton (1962)[44]	Trifluoperazine	CBS	Psychosis	HOS	27	8
Sugerman (1964)[55]	Haloperidol	CBS	Psychosis	HOS	18	6
Rada (1976)[66]	Thiothixene	CBS	Diverse NPS	HOS	42	4
Barnes (1982)[74]	Thioridazine Loxapine	Dementia	Diverse NPS	NH	53	8
Petrie (1982)[75]	Haloperidol Loxapine	Dementia	Diverse NPS	HOS	61	8
Finkel (1995)[76]	Thiotixene	Dementia	Agitation	NH	33	12
Auchus (1997)[77]	Haloperidol	AD	Agitation	OUTP	12	6
Devanand (1998)[78]	Haloperidol	AD	Diverse NPS	OUTP	66	6

table continues

Author (year)	Antipsychotic drug	Type of dementia	Type of NPS	Setting	N	Duration, weeks
Teri (2000)[79]	Haloperidol	AD	Agitation	HOS	70	16
Pollock (2002)[45]	Perphenazine	AD, VAS and MIX	Diverse NPS	NH	54	2,5
Atypical antipsychotics (25 trials)						
Pfizer Ph (1993)[46]	Ziprasidone	AD and VAS	Diverse NPS	NH	23	4
Satterlee (1995)[47]	Olanzapine	AD	Diverse NPS	NR	238	8
Janssen Ph (1997)[48]	Risperidone	AD	Diverse NPS	NH	39	4
Katz (1999)[50]	Risperidone	AD, VAS and MIX	Diverse NPS	NH	625	12
Street (2000)[51]	Olanzapine	AD	Diverse NPS	NH	206	6
Howanitz (2001)[52]	Olanzapine	VAS	Diverse NPS	NR	16	6
Herz (2002)[53]	Risperidone Olanzapine	AD	Agitation	NR	29	6
Novartis Ph (2002)[54]	Iloperidone	AD, VAS and MIX	NR	NR	15	4
Brodaty (2003)[56]	Risperidone	AD, VAS and MIX	Agitation[a]	NH	345	12
Janssen Ph (2003)[49]	Risperidone	AD	Psychosis	NH	18	8
De Deyn (2004)[58]	Olanzapine	AD	Psychosis	NH & HOS	652	10
Ballard (2005)[59]	Quetiapine	AD	Agitation	NH	62	6
De Deyn (2005)[80]	Aripiprazole	AD	Psychosis	OUTP	208	10
Deberdt (2005)[60]	Risperidone Olanzapine	AD, VAS and MIX	Psychosis	NH-OUTP	494	10
Janssen Ph (2005)[57]	Risperidone	AD	Psychosis	NR	33	8
Mintzer (2006)[61]	Risperidone	AD	Psychosis	NH	473	8
Schneider (2006)[21]	Risperidone Olanzapine Quetiapine	AD	Diverse NPS	OUTP	421	36
Mintzer (2007)[62]	Aripiprazole	AD	Psychosis	NH	487	10
Zhong (2007)[63]	Quetiapine	AD and VAS	Agitation	NH	333	10
Paleacu (2008)[64]	Quetiapine	AD	Diverse NPS	NR	40	6
Streim (2008)[65]	Aripiprazole	AD	Psychosis	NH	256	10
Otsuka Ph (2017a)[67]	Brexpiprazole[b]	AD	Agitation	NH	433	12
Otsuka Ph (2017b)[68]	Brexpiprazole	AD	Agitation	NH-OUTP	270	12
Ballard (2018)[69]	Pimavanserin	AD	Psychosis	NH	181	12
ACADIA Ph (2018)[70]	Pimavanserin	AD	Agitation	NH-OUTP	111	12
Conventional and atypical antipsychotic drug (3 trials)						
De Deyn (1999)[71]	Haloperidol Risperidone	AD, VAS and MIX	Diverse NPS	NH	334	12
Allain (2000)[72]	Haloperidol Tiapride	AD, VAS and MIX	Agitation	NH-HOS	306	3
Tariot (2006)[73]	Haloperidol Quetiapine	AD	Psychosis	NH	284	10

AD: Alzheimer's disease; CBS: chronic brain syndrome; HOS: hospital; MIX:mixed dementia (Alzheimer/ Vascular); NH: nursing home; NPS: neuropsychiatric symptoms; OUTP: outpatients; Ph: Pharmaceutical company; NR: not reported; VAS: vascular dementia;

[a]in particular aggression

[b]results reported only for 1mg, 2 mg and placebo groups (total n=413)

Table 2. Effects of antipsychotics on subjective and objective outcomes in patients with neuropsychiatric symptoms in dementia

	Conventional AP vs placebo (13 trials)		**Atypical AP vs placebo (28 trials)**		**$OR_{conventional}$ vs $OR_{atypical}$**
	Trials, n	OR unless indicated otherwise (95% CI)	Trials, n	OR unless indicated otherwise (95% CI)	p-value
Effectiveness: Subjective outcomes					
Change in symptoms, SMD	10	-0.36 (-0.49 – -0.23)	23	-0.14 (-0.19 – -0.08)	<0.001
Response rate	11	1.82 (1.39 – 2.38)	13	1.53 (1.32 – 1.76)	0.267
Effectiveness: Objective outcomes					
Overall dropout	11	1.03 (0.77 – 1.37)	26	1.01 (0.89 – 1.14)	0.897
Use of additional psychotropic medication	3	0.82 (0.55 – 1.22)	9	0.87 (0.73 – 1.03)	0.803
Side effects: Subjective outcomes					
EPS	6	2.93 (2.04 – 4.22)	17	1.52 (1.23 – 1.88)	0.002
Somnolence[a]	6	4.07 (1.80 – 9.20)	18	2.69 (1.99 – 3.62)	0.347
Side effects: Objective outcomes					
Dropout due to adverse events	6	1.78 (1.05 – 3.00)	24	1.51 (1.25 – 1.83)	0.569
Use of medication for EPS	2	1.67 (0.64 – 4.35)	3	0.95 (0.55 – 1.62)	0.312
Falls	1	1.02 (0.55 – 1.91)	11	0.99 (0.80 – 1.22)	0.920

AP: antipsychotics; CI: 95% confidence interval; EPS: extrapyramidal symptoms; OR: odds ratio; SMD: standardized mean difference; n: number; vs: versus;

[a]random effects model

Effectiveness

Conventional antipsychotics had a statistically significant but small effect on the reduction of NPS in dementia: SMD -0.36 (95%CI; -0.49 to -0.23). The response rate was also significantly higher in conventional antipsychotic than placebo groups (OR 1.82; 95%CI 1.39 to 2.38). Yet, there was no statistically significant effect of conventional antipsychotics on NPS in dementia compared to placebo when measured with objective outcome measures. The risk of the use of additional psychotropic medication was numerically lower but not statistically significant in the conventional antipsychotic versus placebo groups (OR 0.82; 95%CI 0.55 to 1.22). Overall dropout did not differ between conventional antipsychotics and placebo either (OR 1.03; 95%CI 0.77 to 1.37).

For atypical antipsychotics, there was a statistically significant but clinically negligible decrease of NPS in dementia compared to placebo: SMD -0.14 (95%CI; -0.19 to -0.08). The response rate was however significantly higher in atypical antipsychotics than placebo groups (OR 1.53; 95%CI 1.32 to 1.76). Again, there was also no effect of atypical antipsychotics on NPS in dementia compared to placebo when measured with objective outcome measures. The risk of the use of additional psychotropic medication was numerically lower but not statistically significant compared to placebo (OR 0.87; 95%CI 0.73 to 1.03). Overall dropout did not differ between atypical antipsychotics and placebo groups (OR 1.01; 95%CI 0.89 to 1.14).

As reported above, both conventional and atypical antipsychotics had an effect on NPS in dementia when measured subjectively. We tested whether these effects differed statistically. Conventional antipsychotics reduced NPS more than atypical antipsychotics compared to placebo (SMD -0.36 versus -0.14; p<0.001), but the response rates did not differ statistically (OR 1.82 versus 1.53; p=0.267). There was no statistical difference between conventional and atypical antipsychotics when measured with objective outcome measure overall dropout (OR 1.03 versus 1.01; p=0.897), or use of additional psychotropic medication (OR 0.82 versus 0.87; p=0.803).

Side effects

When measured with subjective scales for EPS, conventional antipsychotics were associated with significantly more EPS than placebo (OR 2.93; 95%CI 2.04 to 4.22). Somnolence occurred significantly more often in the conventional antipsychotic than placebo group too (OR 4.07; 95%CI 1.80 to 9.20). When measured with objective outcome measures, the risk of dropout due to adverse-events was also significantly higher in conventional antipsychotics than placebo groups (OR 1.78; 95%CI 1.05 to 3.00). Medication for EPS was used more often in conventional antipsychotics compared to placebo but the difference was not statistically significantly different (OR 1.67; 95%CI 0.64 to 4.35). The risk of falls was not increased (OR 1.02; 95% CI 0.55-1.91).

In atypical antipsychotic groups, the risk of EPS was significantly higher compared to placebo (OR 1.51; 95%CI 1.25 to 1.82). Somnolence occurred significantly more often in the atypical antipsychotic than placebo groups too (OR 2.69; 95%CI 1.99 to 3.62). When measured with objective outcome measures for side effects, the risk of dropout due to adverse-events was also significantly higher with atypical antipsychotics than with placebo (OR 1.51; 95%CI 1.25 to 1.83). The risk of using medication for EPS did not differ between atypical antipsychotic and placebo groups (OR 0.95; 95%CI 0.55 to 1.62), nor did the risk of falls (OR 0.99; 95% CI 0.80-1.22).

We also tested whether the risk of side effects differed between conventional and atypical antipsychotics versus placebo. The risk of EPS was higher in conventional antipsychotic than atypical antipsychotic groups versus placebo (OR 2.93 versus 1.52; p=0.002), but the risk of somnolence was not (OR 4.07 versus 2.69; p=.347). There was no statistically significant difference between conventional and atypical antipsychotics when measured with the objective outcome measures dropout due to adverse effects (OR 1.78 versus 1.51; p=0.569), the use of medication for EPS (OR 1.67 versus 0.95; p=0.312), or risk of falls (OR 1.02 versus 0.99; p=0.920).

Three placebo-controlled studies tested both the new generation atypical antipsychotics and haloperidol, the standard (conventional) drug at the time, against placebo.[69-71] In a post-hoc sensitivity analysis without these studies, the risk of drop-out due to adverse events was no longer statistically significantly increased for conventional antipsychotics versus placebo. In addition, the response rate and risk of somnolence for these drugs became close to those of atypical drugs versus placebo. All other results did not change substantially, or could not be reliably interpreted due to too few studies.

Discussion

We performed a meta-epidemiological study of 38 trials testing conventional and atypical antipsychotics for NPS in dementia. Antipsychotics were effective when measured with subjective measures, but not with objective measures. Likewise, conventional antipsychotics were more effective than atypical antipsychotics when measured subjectively, but this difference did not hold when measured objectively. For both drug groups, EPS and somnolence occurred more often in antipsychotic than placebo groups when measured with subjective scales and objectively with drop-out due to adverse events. The use of medication for EPS seemed to be higher for conventional antipsychotics but not for atypical antipsychotics, but power was too low to yield definitive estimates. The risk of falls was not increased for either type of antipsychotic.

Subjective versus objective measures

We found that subjective measures of efficacy suggested that conventional antipsychotics had a small effect on NPS in dementia, and atypical antipsychotics a very small (negligible) effect. If these were unbiased estimates of the true effects, we would have expected them to be confirmed with estimates based on objective measures. However, according to the outcomes overall dropout and use of additional psychotropic medication, antipsychotics were not effective for NPS in

dementia. Prior meta-analyses also found that although conventional and atypical antipsychotics decreased subjectively measured symptoms, dropout rates did not differ between treatment and placebo groups.[81–83] Despite the latter finding, the conclusions of these meta-analyses were that antipsychotics were efficacious for NPS in dementia.

There are a number of explanations for the difference in findings based on subjective and objective measures of efficacy of antipsychotics for NPS in dementia. First, biased outcome reporting, i.e. systematic measurement error, can occur when patients or caregivers can guess which treatment they receive despite blinding.[10] Patients and caregivers might also be more willing to complete the trial, if they or staff believe the patient to be in the treatment group.[15] Likewise, staff might be more tended to motivate patients and caregivers when patients are thought to receive active treatment. In case of antipsychotics, typical side effects such as EPS can give away treatment status and lead to these effects. It is also possible that, apart from bias, antipsychotics are efficacious especially in patients with side effects because benefits and harms stem from the same neurotransmitter inhibition. Or, in case of somnolence, the reduction of NPS is the direct effect of the side effect.

We found that EPS and somnolence occurred more often in conventional and atypical antipsychotic groups compared to placebo as assessed with subjective measures. These findings correspond with those of prior meta-analyses.[81–86] In addition, the risk of EPS was higher for conventional than atypical antipsychotics compared to placebo, which also corresponds with prior meta-analytic findings.[87,88] Part of this finding might be explained by higher doses of haloperidol used in older haloperidol trials.

Remarkably though, drop-out due to adverse events did not differ statistically between conventional and atypical antipsychotics in our study. Probably, EPS and somnolence were not the only adverse events leading to drop-out. Other less prevalent side effects or serious adverse events might have played a role. For example, meta-analyses of trials have shown that atypical antipsychotics had an increased risk of death in patients with dementia, but conventional antipsychotics did not.[89,90] Atypical antipsychotics also increased the risk of cerebrovascular accidents in trials among patients with Alzheimer's disease.[84] In addition, although drop-out is an objective measure, knowledge of the treatment can also influence drop-out, but probably much less than the usual subjective measured outcomes.

In addition, use of medication for EPS was not statistically significantly increased for conventional and atypical antipsychotics versus placebo. However, lack of power is

a problem in both comparisons, with only 2 of 14 trials respectively 5 of 28 trials reporting this outcome. In addition, the use of medicines for EPS will not cover all patients that develop EPS, because physicians might rather discontinue treatment or lower the antipsychotic dose. Possible selective reporting, with studies reporting these outcomes having more favorable results, renders a correct interpretation of our finding even more difficult.

Finally, the risk of falls was not increased for either type of antipsychotic even though the risks of extrapyramidal symptoms and somnolence, which can lead to falls, were. Possibly, antipsychotic use might especially increase the rate of falls, but the mean number of falls per participant was not reported in the studies. In addition, none of the trials identified falls as an outcome a-priori, so it is not clear whether falls had been recorded systematically, if at all.

Strengths and limitations

As far as we know, this is the first meta-analysis that investigated the effects of antipsychotics for NPS in dementia with subjective and objective outcome measures. In addition, we performed a broad search covering unpublished data. This resulted in the inclusion of a relatively large number of trials compared to prior reviews.

Unfortunately, most of the older trials that we included, namely the studies testing conventional antipsychotics, did not report all the variables we were interested in. Particularly, the objective outcome measures use of additional psychotropic medication, use of medication for EPS, and falls were often missing. If this was the result of selective reporting, the risk of side effects might have been underestimated in our analyses with these measures. In addition, due to the lack of data, reliability of some of the pooled effects was low.

Conclusion

Effectiveness of antipsychotics for NPS in dementia based on commonly used subjective scales could not be confirmed with objective measures. Subjective measures of side effects suggested that conventional antipsychotics had a higher risk than atypical antipsychotics, but objective measures did not. Therefore, future trials and reviews about psychotropic medication for NPS in dementia need to address potential information bias by regularly including objective measures. Guidelines need to base recommendations on effects established preferably with objective outcome measures.

References

1. Butler R, Radhakrishnan R. Dementia. BMJ Clin Evid. 2012;(July 2011):1-27.
2. Devanand DP. The Course of Psychopathologic Features in Mild to Moderate Alzheimer Disease. Arch Gen Psychiatry. 1997;54(3):257. doi:10.1001/archpsyc.1997.01830150083012
3. Steinberg M, Sheppard J-M, Tschanz JT, et al. The Incidence of Mental and Behavioral Disturbances in Dementia: The Cache County Study. J Neuropsychiatry Clin Neurosci. 2003;15(3):340-345.
4. Finkel SI. Behavioral and Psychological Symptoms of Dementia: A Current Focus for Clinicians, Researchers, and Caregivers. J Clin Psychiatry. 2001;62(suppl 21):3-6.
5. Zuidema S, Koopmans R, Verhey F. Prevalence and predictors of neuropsychiatric symptoms in cognitively impaired nursing home patients. Int J Geriatr psychiatry Neurol. 2007;20(1):41-49. doi:10.1177/0891988706292762
6. de Vugt M, Riedijk S, Aalten P, Tibben A, van Swieten J, Verhey F. Impact of behavioural problems on spousal caregivers: a comparison between Alzheimer's disease and frontotemporal dementia. Dement Geriatr Cogn Disord. 2006;22(1):35-41. doi:10.137
7. Coen RF, Swanwick GRJ, O'Boyle CA, Coakley D. Behaviour disturbance and other predictors of carer burden in Alzheimer's disease. Int J Geriatr Psychiatry. 1997;12(3):331-336. doi:10.1002/(SICI)1099-1166(199703)12:3<331::AID-GPS495>3.0.CO;2-J
8. Janus SIM, van Manen JG, IJzerman MJ, Zuidema SU. Psychotropic drug prescriptions in Western European nursing homes. Int Psychogeriatrics. 2016;28(11):1775-1790. doi:10.1017/S1041610216001150
9. Nijk RM, Zuidema SU, Koopmans RTCM. Prevalence and correlates of psychotropic drug use in Dutch nursing-home patients with dementia. Int Psychogeriatr. 2009;21(3):485-493. doi:10.1017/S1041610209008916
10. Baethge C, Assall OP, Baldessarini RJ. Systematic review of blinding assessment in randomized controlled trials in schizophrenia and affective disorders 2000-2010. Psychother Psychosom. 2013;82(3):152-160. doi:10.1159/000346144
11. Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: Meta-epidemiological study. Bmj. 2008;336(7644):601-605. doi:10.1136/bmj.39465.451748.AD
12. Schulz KF. The Landscape and Lexicon of Blinding in Randomized Trials. Ann Intern Med. 2002;136(3):254. doi:10.7326/0003-4819-136-3-200202050-00022
13. Moncrieff J, Wessely S, Hardy R. Active placebos versus antidepressants for depression. Cochrane Database Syst Rev. 2004;(1):CD003012. doi:10.1002/14651858.CD003012.pub2

14. Başoğlu M, Marks I, Livanou M, Swinson R. Double-blindness procedures, rater blindness, and ratings of outcome. Observations from a controlled trial. Arch Gen Psychiatry. 1997;54(8):744-748.
15. Colagiuri B, Boakes RA. Perceived treatment, feedback, and placebo effects in double-blind RCTs: An experimental analysis. Psychopharmacology (Berl). 2010;208(3):433-441. doi:10.1007/s00213-009-1743-9
16. Barbui C, Furukawa TA, Cipriani A. Effectiveness of paroxetine in the treatment of acute major depression in adults: a systematic re-examination of published and unpublished data from randomized trials. CMAJ. 2008;178(3):296-305. doi:10.1503/cmaj.070693
17. World Health Organization. WHO Collaborating Centre for Drugs Statistics Methodology.
18. Food and drug administration. Postmarket Drug Safety Information for Patients and Providers - Atypical Antipsychotic Drugs Information. Center for drug evaluation and research.
19. Wikipedia. Typical antipsychotics. Wikipedia, Free Encylcl.
20. Higgins J, Altman D. Assessing Risk of Bias in Included Studies. Cochrane H. (Higgins J, Green S, eds.).; 2008.
21. Schneider L, Tariot P, Dagerman K, et al. Effectiveness of Atypical Antipsychotic Drugs in Patients with Alzheimer's Disease. N Engl J Med. 2006;355(15):1525-1538.
22. Rabinowitz J, Katz I, De Deyn P, Greenspan A, Brodaty H. Treating behavioral and psychological symptoms in patients with psychosis of Alzheimer's disease using risperidone. Int psychogeriatrics. 2007;19(2):227-240. doi:10.1017/S1041610206003942
23. Cahn L, Diesfeldt H. The use of neuroleptics in the treatment of dementia in old age. Psychiat Neurol Neurochir. 1973;(76):411-420.
24. Barton R, Hurst L. Unnecessary use of tranquilizers in elderly patients. Br J Psychiatry. 1966;112(491):989-990. doi:10.1192/bjp.112.491.989
25. Devanand D, Devangere P, Sackeim H, Brown R, Mayeux R. A Pilot Study of Haloperidol Treatment of Psychosis and Behavioral Disturbance in Alzheimer's Disease. Arch Neurol. 1989;46(8):854-857. doi:10.1001/archneur.1989.00520440036018
26. Tewfik G, Jain V, Harcup M, Magowan S. Effectiveness of various tranquillisers. Geront Clink. 1970;12:351-359.
27. Meguro K, Meguro M, Tanaka Y, Akanuma K, Yamaguchi K, Itoh M. Risperidone is effective for wandering and disturbed sleep/wake patterns in Alzheimer's disease. J Geriatr Psychiatry Neurol. 2004;17(2):61-67. doi:10.1177/0891988704264535
28. Rappaport S, Marcus R, Manos G, McQuade R, Oren D. A randomized, double-blind, placebo-controlled tolerability study of intramuscular aripiprazole in acutely agitated patients with Alzheimer's, vascular, or mixed dementia. J Am Med Dir Assoc. 2009;10(1):21-27.
29. Meehan K, Wang H, David S, et al. Comparison of Rapidly Acting Intra-

muscular Olanzapine , Lorazepam , and Placebo : A Double-blind , Randomized Study in Acutely Agitated Patients with Dementia. Neuropsychopharmacology. 2002;26:494-504.

30. Lehmann H, Ban T, Saxena M. Nicotinic acid, thioridazine, fluoxymesterone and their combinations in hospitalized geriatric patients. Can Psychiatr Assoc J. 1972;17(4):315-320.
31. Hamilton LD, Bennett JL. Acetophenazine for hyperactive geriatric patients. Geriatrics. 1962;(17):596-601.
32. Kurlan R, Cummings J, Raman R, Thal L. Quetiapine for agitation or psychosis in patients with dementia and parkinsonism. Neurology. 2007;68(17):1356-1363. doi:10.1212/01.wnl.0000260060.60870.89
33. Kennedy J, Deberdt W, Siegal A, et al. Olanzapine does not enhance cognition in non-agitated and non-psychotic patients with mild to moderate Alzheimer's dementia. Int J Geriatr Psychiatry. 2005;20(11):1020-1027. doi:10.1002/gps.1397
34. Ather SA, Shaw SH, Stoker MJ. A comparison of chlormethiazole and thioridazine in agitated confusional states of the elderly. Acta Psychiatr Scand. 1986;73(10):81-91. doi:10.1111/j.1600-0447.1986.tb10541.x
35. Bamrah S. A multicentre, double-blind, randomised, parallel group comparison of seroquel and haloperidol in the treatment of elderly patients presenting with dementia and psychosis.
36. Birkett D, Boltuch B. Chlorpromazine in geriatric psychiatry. J Am Geriatr Soc. 1972;20(8):403-406.
37. Martin-Cook K, Hynan L, Rice-Koch K, Svetlik D, Weiner M. Responsiveness of the quality of life in late-stage dementia scale to psychotropic drug treatment in late-stage dementia. Dement Geriatr Cogn Disord. 2005;19(2-3):82-85. doi:10.1159/000082353
38. Engstrand E. Buronil og promazinbehandling av senildemente: En dobbeltblind undersøkelse. Nord J Psychiatry. 1967;21(3):234-240. doi:10.3109/08039486709094596
39. Spagnolo C, Dall'asta D, Iannuccelli M, Cucinotta D, Passeri M. A controlled double-blind trial comparing etoperidone with thioridazine in the management of severe senile dementia. Drugs exptl clin res. 1983;(12):873-880.
40. Sheng J, Gao Z, Chen M, Zhang M, Liu J. Risperidone vs haloperidol in treatment of behavioral and psychological symptoms of dementia : a randomized, double-blind trial. Chin J New Drugs Clin Rem. 2004;23(6):359-362.
41. Sun X, Gao Z, Feng F. A randomized double-blind trial of haloperidol and risperidone for behavioral and psychological symptoms of dementia. Chin J Psychiatry. 2004;37(3):156-159.
42. Auer SR, Monteiro IM, Reisberg B. Behavioral symptoms in dementia: Community-based research. Int Psychogeriatrics. 1996;8(SUPPL. 3):363-366. doi:10.1017/S1041610297003633
43. Stotsky B. Multicenter Study Comparing Thioridazine with Diazepam and Placebo in Elderly, Nonpsychotic Patients with Emotional and Behavioral

Disorders. Clin Ther. 1984;6(4):546-559.
44. Hamilton LD, Bennett JL. The use of trifluoperazine in geriatric patients with chronic brain syndrome. J Am Geriatr Soc. 1962;10(2):140-147. doi:10.1111/j.1532-5415.1962.tb00266.x
45. Pollock BG, Mulsant BH, Rosen J, et al. Comparison of citalopram, perphanezine, and placebo for the acute treatment of psychosis and behavioral disturbances in hospitalized, demented patients. Am J Psychiatry. 2002;159(March):460-465.
46. Pfizer. ZIP-128-105. 1993. doi:10.2165/11599010-000000000-00000
47. Satterlee W. Olanzapine versus Placebo in the Treatment of Patients with Psychosis Associated with Dementia (F1D-MC-HGAO). 1995.
48. Janssen Pharmaceutical. Ris-Bel-14.; 1997.
49. Janssen Pharmaceutical. Ris-Int-83 Info.; 2003.
50. Katz IR, Jeste D V, Mintzer JE, Clyde C, Napolitano J, Brecher M. Comparison of risperidone and placebo for psychosis and behavioral disturbances associated with dementia: a randomized, double-blind trial. Risperidone Study Group. J Clin Psychiatry. 1999;60(2):107-115.
51. Street J, Clark S, Gannon K, Cummings J, Bymaster F. Olanzapine Treatment of Psychotic and Behavioral Symptoms in Patients With Alzheimer Disease in Nursing Care Facilities. Arch Gen Psychiatry. 2000;57:968-976.
52. Howanitz E, Wisotzek I. Olanzapine versus placebo in the treatment of behavioral disturbances associated with vascular dementia. 2001.
53. Herz L, Volicer L, Frankenburg F, Colon S, Kittur S. Abstract of the 2002 Annual Meeting of the International College of Geriatric Psychoneuropharmacolgy. 2002.
54. Novartis Pharma. A Prospective, Randomized, Double-Blind, Placebo-Controlled, Flexible-Dose, Parallel-Group, Single-Center Study to Evaluate the Safety, Tolerability and Efficacy of Iloperidone Compared with Placebo in Treating Psychotic.; 2002. doi:10.137
55. Sugarman AA, Williams BH, Adlerstein AM. Haloperidol in the Psychiatric Disorders of Old Age. Am J Psychiatry. 1964;120:1190-1192.
56. Brodaty H, Ames D, Snowdon J, et al. A randomized placebo-controlled trial of risperidone for the treatment of aggression, agitation, and psychosis of dementia. J Clin Psychiatry. 2003;64(2):134-143. doi:http://dx.doi.org/10.4088/JCP.v64n0205
57. Janssen Pharmaceutical. Clinical study report synopsis. 2005:9-12.
58. De Deyn P, Carrasco M, Deberdt W, et al. Olanzapine versus placebo in the treatment of psychosis with or without associated behavioral disturbances in patients with Alzheimer's disease. Int J Geriatr Psychiatry. 2004;19(2):115-126.
59. Ballard C, Margallo-Lana M, Juszczak E, et al. Quetiapine and rivastigmine and cognitive decline in Alzheimer's disease: randomised double

blind placebo controlled trial. BMJ. 2005;330(7496):874.

60. Deberdt WG, Dysken MW, Rappaport S a, et al. Comparison of olanzapine and risperidone in the treatment of psychosis and associated behavioral disturbances in patients with dementia. Am J Geriatr Psychiatry. 2005;13(8):722-730. doi:10.1176/appi.ajgp.13.8.722
61. Mintzer J, Greenspan A, Caers I, et al. Risperidone in the Treatment of Psychosis of Alzheimer Disease: Results From a Prospective Clinical Trial. Am J Geriatr Psychiatry. 2006;(March):280-291.
62. Mintzer JE, Tune LE, Breder CD, et al. Aripiprazole for the treatment of psychoses in institutionalized patients with Alzheimer dementia: a multicenter, randomized, double-blind, placebo-controlled assessment of three fixed doses. Am J Geriatr Psychiatry. 2007;15(11):918-931. doi:10.1097/JGP.0b013e3181557b47
63. Zhong K, Tariot P, Mintzer J, Minkwitz M, Devine N. Quetiapine to Treat Agitation in Dementia: A Randomized , Double-Blind , Placebo-Controlled Study. Curr Alzheimer Res. 2007;4(1):81-93.
64. Paleacu D, Barak Y, Mirecky I, Mazeg D. Quetiapine treatment for behavioural and psychological symptoms of dementia in Alzheimer's disease patients: a 6-week, double-blind, placebo-controlled study. Int J Geriatr Psychiatry. 2008;23:393-400. doi:10.1002/gps
65. Streim JE, Porsteinsson AP, Breder CD, et al. A randomized, double-blind, placebo-controlled study of aripiprazole for the treatment of psychosis in nursing home patients with Alzheimer disease. Am J Geriatr Psychiatry. 2008;16(7):537-550. doi:10.1097/JGP.0b013e318165db77
66. Rada R, Kellner R. Thiothixene in the Treatment of Geriatric Patients With Chronic Organic Brain Syndrome. J Am Geriatr Soc. 1976;24(3):105-107.
67. Otsuka Pharmaceutical. 2017a. www.clinicaltrailsregister.eu.
68. Pharmaceutical O. 2017b. www.clinicaltrailsregister.eu.
69. Ballard C, Banister C, Khan Z, et al. Evaluation of the safety, tolerability, and efficacy of pimavanserin versus placebo in patients with Alzheimer's disease psychosis: a phase 2, randomised, placebo-controlled, double-blind study. Lancet Neurol. 2018;17(3):213-222. doi:10.1016/S1474-4422(18)30039-5
70. Acadia Pharmaceutical. No Title. www.clinicaltrailsregister.eu.
71. De Deyn PP, Rabheru K, Rasmussen A, et al. A randomized trial of risperidone, placebo, and haloperidol for behavioral symptoms of dementia. Neurology. 1999;53(5):946-955. doi:10.1212/WNL.53.5.946
72. Allain H, Dautzenberg PH, Maurer K, et al. Double blind study of tiapride versus haloperidol and placebo in agitation and aggressiveness in elderly patients with cognitive impairment. Psychopharmacology (Berl). 2000;148(4):361-366.
73. Tariot P, Schneider L, Katz I, et al. Quetiapine Treatment of Psychosis Associ-

ated With Dementia : A Double-Blind, Randomized, Placebo-Controlled Clinical Trial. Am J Geriatr Psychiatry. 2006;14(9):767-776.
74. Barnes R, Veith R, Okimoto J, Raskind M, Gumbrecht G. Efficacy of Antipsychotic Medications in Behaviorally Disturbed Dementia Patients. Am J Psychiatry. 1982;139(9):1170-1174.
75. Petrie W, Ban T, Berney S, et al. Loxapine in Psychogeriatrics: A Placebo- and Standard-Controlled Clinical Investigation. J Clin Psychopharmacol. 1982;2(2):122-126.
76. Finkel SI, Lyons JS, Anderson RL, et al. A randomized, placebo-controlled trial of thiothixene in agitated, demented nursing home patients. Int J Geriatr Psychiatry. 1995;10(2):129-136. doi:10.1002/gps.930100208
77. Auchus AP, Bissey-Black C. Pilot study of haloperidol, fluoxetine, and placebo for agitation in Alzheimer's disease. J Neuropsychiatry Clin Neurosci. 1997;9(4):591-593. doi:10.1176/jnp.9.4.591
78. Devanand DP, Marder K, Michaels KS, et al. A radomized, placebo-controlled dose-comparison trial of haloperidol for psychosis and disruptive behaviors in Alzheimer's disease. Am J Psychiatry. 1998;155(11):1512-1520. doi:10.1176/ajp.155.11.1512
79. Teri L, Logsdon RG, Peskind E, et al. Treatment of agitation in AD: a randomized, placebo-controlled clinical trial. Neurology. 2000;55(9):1271-1278. doi:10.1212/WNL.55.9.1271
80. De Deyn P, Jeste D V, Swanink R, et al. Aripiprazole for the treatment of psychosis in patients with Alzheimer's disease: a randomized, placebo-controlled study. J Clin Psychopharmacol. 2005;25(5):463-467. doi:10.1097/01.jcp.0000178415.22309.8f
81. Lonergan E, Luxenberg J, Colford JM, Birks J. Haloperidol for agitation in dementia. Cochrane Database Syst Rev. 2010;(2):1-25. doi:CD002852 [pii]\r10.1002/14651858.CD002852
82. Wang J, Yu J-T, Wang H-F, et al. Pharmacological treatment of neuropsychiatric symptoms in Alzheimer's disease: a systematic review and meta-analysis. J Neurol Neurosurg Psychiatry. 2015;86(1):101-109. doi:10.1136/jnnp-2014-308112
83. Ma H, Huang Y, Cong Z, et al. The Efficacy and Safety of Atypical Antipsychotics for the Treatment of Dementia: A Meta-Analysis of Randomized Placebo-Controlled Trials. J Alzheimer's Dis. 2014;42:915-937.
84. Ballard C, Waite J, Birks J. Atypical antipsychotics for aggression and psychosis in Alzheimer ' s disease (Review). Cochrane Libr. 2006;(1):1-135.
85. Tan L, Tan L, Wang H, et al. Efficacy and safety of atypical antipsychotic drug treatment for dementia : a systematic review and meta-analysis. Alzheimers Res Ther. 2015;7(1):1-13. doi:10.1186/s13195-015-0102-9
86. Yunusa I, Alsumali A, Garba AE, Regestein QR, Eguale T. Assessment of Reported Comparative Effectiveness and Safety of Atypical Antipsychotics in the Treatment of Behavioral and Psychological Symptoms of Dementia: A Network Meta-analysis. JAMA

Netw open. 2019;2(3):e190828. doi:10.1001/jamanetworkopen.2019.0828

87. Carson S, McDonagh MS, Peterson K. A systematic review of the efficacy and safety of atypical antipsychotics in patients with psychological and behavioral symptoms of dementia. J Am Geriatr Soc. 2006;54(2):354-361. doi:10.1111/j.1532-5415.2005.00566.x
88. Lee PE, Gill SS, Freedman M, Bronskill SE, Hillmer MP, Rochon PA. Atypical antipsychotic drugs in the treatment of behavioural and psychological symptoms of dementia : systematic review. BMJ. 2004;55(June):1-5. doi:10.1136/bmj.38125.465579.55
89. Schneider L, Dagerman K, Insel P. Risk of Death With Atypical Antipsychotic Drug Treatment for Dementia. J Am Med Assoc. 2005;294(15):1934-1943.
90. Hulshof TA, Zuidema SU, Ostelo RWJG, Luijendijk HJ. The Mortality Risk of Conventional Antipsychotics in Elderly Patients: A Systematic Review and Meta-analysis of Randomized Placebo-Controlled Trials. J Am Med Dir Assoc. 2015;16(10):817-824. doi:10.1016/j.jamda.2015.03.015

Chapter 8.

General discussion

Discussion

A standard part of performing a systematic review of trials is to critically appraise the quality of the included trials. Commonly used tools for assessing risk of bias in trials are the Cochrane risk of bias tool[1] and the Jadad scale.[2] There are multiple tools for other study designs too.[3] However, despite the use of standardized items and answer categories, the correct application depends on the level of methodological knowledge of the reviewers as well as their attitude towards missing information. This can lead to interrater variation on a tool, and poor agreement between tools.[4,5] Also, the assessments have a qualitative nature in that risk of bias is often categorized as either low, high or unclear. Hence, there is a need for more objective risk of bias tools. The aim of this thesis was to explore quantitative methods to assess the risk of bias in trials as part of reviews.

This thesis focused on antipsychotic drug trials for patients with dementia. This topic served as a test case. The quantitative methods for assessing risk of bias in this thesis can be applied in reviews of other topics too. I attempted to quantify the effect of various sources of bias. These included the enrolment of patients and use of outcomes not corresponding with the research question (PICO), patient deselection after run-in, baseline imbalances between treatment groups despite randomization, and the use of subjective outcome measures. In general, it is shown that in studies in which the reported source of bias was either absent or small, the effectiveness of antipsychotics in dementia seems lower and risk of side effects higher.

In this chapter, a short history of risk of bias assessments is provided first. Then, the main findings of this thesis will be discussed. Finally, considerations for future research are stated.

The history of risk of bias assessments

The importance of clinical trials in drug development received increasing recognition in the 1960s due to the thalidomide crisis (see figure 1).[6] Thalidomide was marketed as a safe drug without having been tested extensively in clinical trials. Multiple observations and cases showed it to cause severe birth defects and malformities when used in pregnancy.[6,7] As a result many countries updated their drug regulatory laws.[6,7]

A decade later, Archie Cochrane criticized the lack of reliable evidence for many accepted healthcare interventions.[8] He explained how summarizing the results of randomized controlled trials (RCTs) in reviews could help to determine the effectiveness of health treatments, and so help doctors and patients make better informed decisions.[9–11]

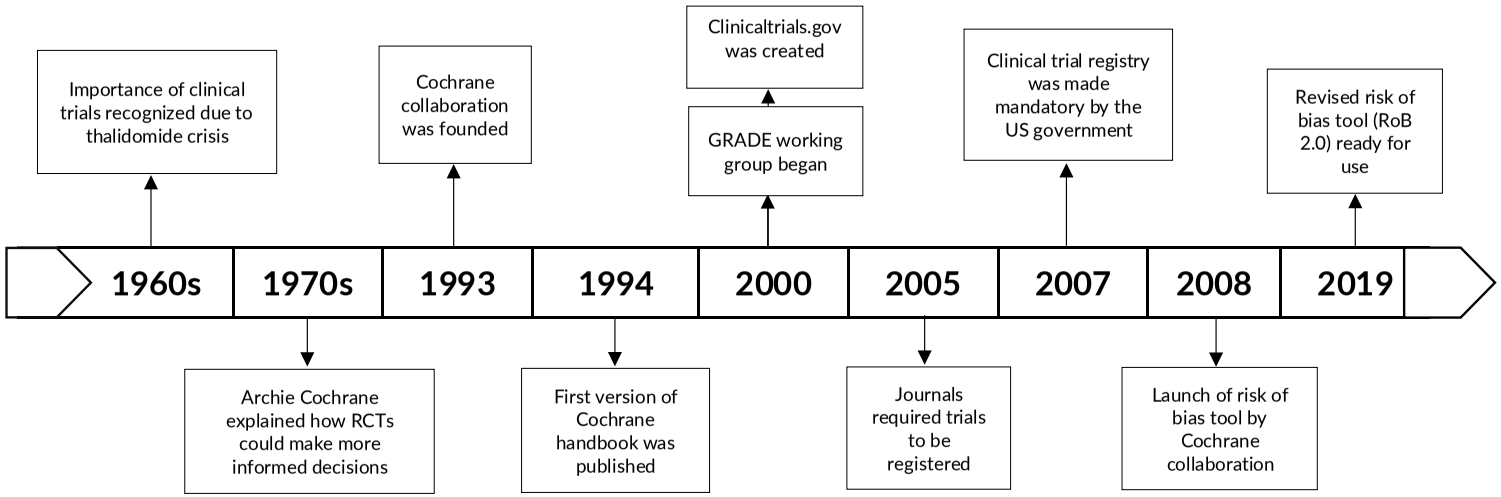


Figure 1. Timeline

In the following years, different researchers outside healthcare also gradually adopted the importance of systematic reviews.[9] Since then, much has improved in both the conduct of RCTs and reviews. For example, sample sizes of trials have increased, and this leads to more accurately estimated results.[12]

However, new challenges arose too. E.g. some trials are not published at all, or only partly. Other trials were terminated early when apparent benefit was found.[13] These trials often failed to adequately report the decision to stop early and showed implausibly large treatment effects.[13] In response, ClinicialTrials.gov was created as the first online registry for clinical trials in 2000.[14] The register presents information about a study's design, provides protocol updates, and shows summary results. Nowadays, many trial registration websites exist, most of which are affiliated with a specific country.[15] In addition, the use of online registration has been stimulated. In 2005, journals required investigators to have registered trials online before submitting the manuscripts with results. In 2007, clinical trial registry was made mandatory by the United States government.[14] The goal of these efforts is that publicly available information about unpublished trials will help reviewers to include them in reviews, and so reduce publication bias.[16]

When performing a review of RCTs, publication bias is not the only challenge to conquer. RCTs may also suffer from other types of bias. Another improvement in review methods was therefore the launch of the Risk of Bias tool by the Cochrane collaboration in 2008.[17] Since then, numerous risk of bias tools have been developed by different reviewers for different study designs. Over 25 risk of bias tools exists to help reviewers assess bias in studies as part of a systematic review.[3] Recently, the Cochrane collaboration has published an updated version of their risk of bias tool.[17] The risk of bias tool is one element of the Cochrane handbook for systematic reviews of interventions. The Cochrane handbook describes the complete process of performing a systematic review in detail and gives guidance to reviewers.[18] Cochrane is recognised internationally as a gold standard for high-quality reviews.[19,20] The Cochrane also adopted GRADE recommendations as part of the review process.[21] The GRADE working group developed recommendations to rate the quality of evidence in systematic reviews, and a large number of organizations, including Cochrane, adopted these principles of GRADE.[21]

Current review methods

Nevertheless, despite the significant improvements over the years, the methods of review can still be improved. One reason is that with new developments in the design and conduct of trials, new pitfalls can potentially arise. This thesis addressed some of these issues.

The work started with a review of placebo-controlled randomized antipsychotic drug trials for patients with dementia. Reviews of observational studies had already reported an increased risk of mortality for conventional antipsychotics in elderly patients. However, these observational studies based on routinely collected health data did not adjust for (delirium due to) terminal illness, an important confounder.[22] Randomisation reduces confounding due to group imbalances in measured and unmeasured factors.

The meta-analysis of placebo-controlled randomized trials did not confirm that conventional antipsychotics in general or haloperidol in particular increase the risk of mortality in elderly patients. This review thus confirms that summarizing trial results can provide useful information. It also prompted the remaining meta-epidemiological studies in this thesis.

Populations and outcomes not matching the research question

Both Cochrane and GRADE recommend to define the research question in terms of patients, intervention, comparison intervention, and outcome (PICO) before the start of the review.[24,25] However, it can be challenging to define the patient population sufficiently narrow without excluding relevant trials. When determining the outcomes of interest, it is advised to focus on patient-important outcomes. Sometimes, trials examined surrogate or substitute outcomes, and outcomes of most importance to patients remained unexplored.[25] If reviewers solely address the former, the review results might be misleading.[25,26] In prior reviews of antipsychotics for dementia, the defined patient population of interest were not always applied well to study selection and data-extraction.[25,26] E.g. trials based on mixed patient populations with various neuropsychiatric symptoms (NPS) were included, even though the main interest was on patients with agitation or psychosis. As a result, the review results were also based on patients with symptoms other than those target symptoms, such as depression. Furthermore, previous reviews of antipsychotics for dementia have pooled results from generic scales measuring other neuropsychiatric symptoms besides agitation and psychosis.

My colleagues and I found that conventional antipsychotics might have a small effect on agitation in agitated patients and on psychosis in psychotic patients. Atypical antipsychotics did not have such an effect. Inclusion of data from mixed populations resulted in an underestimation of the efficacy of conventional antipsychotics on agitation and psychosis, but suggested a small effect of atypical antipsychotics on agitation. Neither drug class seemed to have an effect when it was measured with generic symptoms scales.

Box Summary of findings

I critically appraised the methods and results of antipsychotic drug trials for patients with dementia. **Part 1** addressed the clinical relevance of the reported effects:

- In **chapter 2**, a meta-analysis of placebo-controlled randomized trials – in contrast to reviews of observational studies - did not show that conventional antipsychotics in general, or haloperidol in particular, increase the risk of mortality in elderly patients.
- In **chapter 3**, a meta-epidemiological study reported that conventional antipsychotics might have a small effect in agitated patients (the population of interest) on agitation (the outcome of interest) and in psychotic patients (the population of interest) on psychosis (the outcome of interest). This was not the case for atypical antipsychotics.
- In **chapter 4**, placebo-controlled trials that tested atypical antipsychotics showed statistically significant but clinically negligible effects due to relatively large populations; this may give rise to large sample size fallacy.

Part 2 covered three types of bias, which are commonly acknowledged in observational studies but can occur in trials as well[23], that is selection bias, confounding, and information bias, respectively:

- In **chapter 5,** trials with a run-in period showed a somewhat larger reduction in neuropsychiatric symptoms for antipsychotics versus placebo than trials without a run-in period. In addition, trials with a run-in period showed a lower risk of somnolence, extrapyramidal symptoms (EPS), and mortality than trials without run-in.
- In **chapter 6**, placebo-controlled trials of atypical antipsychotics in dementia show heterogeneous baseline imbalances, despite randomization. Trials with missing baseline information seemed to have a more favorable pooled efficacy and lower risk of EPS than trials that reported this information.
- In **chapter 7**, antipsychotics were found to be effective with subjective symptom scales, but not with more objective outcomes (established with less error). Subjective measures of side effects suggested that conventional antipsychotics had a higher risk than atypical antipsychotics, but objective measures did not.

It seems that inaccurate effect estimates can occur if trials are executed among patients without the specific symptoms described by the PICO, and patients are assessed with nonspecific outcome scales. These findings endorse the Cochrane

and GRADE recommendation to define a PICO that is leading for the conduct of the review. If reviewers prefer not to define patients too specifically, the effect of included trials based among 'wider' patient populations could be investigated in sensitivity analyses.

Large sample size fallacy

The Cochrane collaboration discusses the disadvantages of small and missing sample sizes in its handbook, but those of large sample sizes are lacking.[26] Large sample sizes in reviews can magnify bias associated with error resulting from individual study design.[27] Large sample size can also become problematic when statistically significant results are interpreted as clinically relevant, when they are not.[28,29] This is called the large sample size fallacy. The focus on effect size instead of large samples can prevent this pitfall.

The large sample size fallacy is compounded by a misinterpretation of standardized mean differences (SMDs). Cohen developed the following commonly used cut-offs for interpreting SMDs: smaller than 0.2 (negligible effect), 0.2 to 0.5 (small effect), 0.5 to 0.8 (medium effect), and more than 0.8 (large effect).[30] Cohen emphasized that these guidelines were set with diffidence.[30] He further explained that a medium effect is visible to the naked eye of a careful observer, and a large effect is visible to the naked eye, for example, patients or caretakers.[30] In view of this, small effects should not be presented or interpreted as being clinically relevant, but often are,[31,32] resulting in physicians overvalue the treatment in question.

Over the years, the sample sizes in antipsychotic trials have increased massively. This might be due to, for example, a smaller expected treatment effect or an increase in expected drop-out because recent trials lasted longer. A larger sample size provides more power, given a certain treatment effect, and so a higher chance of a statistically significant result.[33] Larger sample sizes are therefore often viewed as a favorable development, but can become a problem in terms of the large sample size fallacy.

We quantitatively assessed the variation of sample sizes, the size of the reported treatment effects, and the association between study characteristics and sample size in antipsychotic trials for dementia. We found that placebo-controlled trials of atypical antipsychotics showed large sample size fallacy while head-to-head trials were clearly underpowered. Also, less than 45% of the trials reported a sample size calculation.

Large sample sizes should however not be avoided. They are a prerequisite for precise study results especially when treatments effects are expected to be small, e.g. in case of rare adverse effects. It is large sample size fallacy that should be avoided.

Therefore, researchers need to be encouraged to report an effect size with a 95% confidence interval.[34,35] This is commonly done in, for example, GRADE summary of findings tables.[21] Also, in line with the CONSORT statement, all trial reports should include sample size calculation.[36] It would be helpful if journals would mention this in their author instructions as well.

Enriched study populations

Some trials use a run-in phase to decrease drop-out and placebo response so that the estimated effect size and power are increased.[37,38] Run-in periods may lead to exclusion (deselection) of certain patients from randomization. As a result, the remaining randomized group does not represent the population of interest anymore, and this limits the validity of trial results for the defined study population.[38,39] This is a source of selection bias that has not received much attention in the medical literature so far.

The Cochrane collaboration discusses how patient withdrawal, i.e. deselection from a study, after randomization can lead to incomplete outcome data.[40] However, drop-out of eligible patients related to the investigated treatment can also occur before randomization as described above. This is not discussed in the Cochrane Handbook. Risk of bias tools for trials do not require consideration of run-in periods either.

We quantitively studied the use of run-in and its effect on the reported results of antipsychotic trials in patients with dementia. We extracted the presence of a run-in period and the reduction in neuropsychiatric symptoms (NPS), the number of participants with somnolence, extrapyramidal symptoms (EPS), and deaths per treatment group. We found that the reduction in NPS of antipsychotics versus placebo was somewhat higher in trials with a run-in period than in trials without a run-in period. The risk of somnolence, EPS, and mortality was lower in trials with than without run-in. Thus, the use of run-in in trials leads to overestimated efficacy and underestimated risks of side effects of antipsychotics compared with placebo prescribed for NPS in patients with dementia, i.e. the patients of interest as defined a priori.

Run-in needs to be considered as a source of bias in trials and reviews of trials. Not only in antipsychotic trials but also in other fields it is important to take run-in into account. Approximately 5% of published trials use a run-in period.[38] For industry sponsored trials, this figure increases to 11% on average.[38] This percentages varies among disciplines, in our review, the percentage of RCTs that used a run-in period was more than 80%. It is recommended that reviewers address the use of a run-in period in the included trials, and quantitatively assess the effect on the reported results.

Incomparable comparison groups

Comparison groups may differ not just in the treatment that they receive. As a result, a difference in outcome between the groups cannot be attributed to the treatment with certainty. This type of bias is called confounding in the context of observational studies. In trials, randomization is used to create comparable comparison groups. Nevertheless, baseline imbalances can still occur by chance. In addition, flawed or corrupted randomization procedures can give rise to systematic baseline imbalances between groups.[41,42] Therefore, CONSORT requires trials to include baseline tables with information about baseline characteristics for *all* randomized patients per treatment group.[36]

As part of a systematic review, trial methods such as randomization are assessed qualitatively. The Cochrane risk of bias tool includes two items on randomization procedures (random sequence generation and concealment of allocation) and one item on (clinically prognostic) baseline imbalances between the comparison groups.[43] Due to the qualitative nature, the assessment is prone to interrater variability and does not enable a detailed analysis of the effect of baseline imbalances on the estimated treatment effects.[44] The only and uncommonly applied option is to exclude trials with clearly imbalanced groups from the analysis.[45]

Quantitative methods to assess the presence of baseline imbalances and their effect on the reported results of antipsychotic trials for dementia are needed. We extracted baseline characteristics prognostic of the efficacy and side effects of atypical antipsychotics in dementia. We found that all included trials reported the randomization procedures incompletely, and despite randomization, the trials showed heterogeneous baseline imbalances. Baseline imbalances were associated with higher efficacy and lower risk of EPS for atypical antipsychotics versus placebo. Moreover, trials with missing baseline information seemed to show a more favorable pooled efficacy and lower pooled risk of EPS than trials that reported this information. We concluded that baseline imbalances that were not taken into account might have mistakenly led to an overestimated efficacy and underestimated risk of EPS.

Imbalanced comparison groups in trials may be more common than is often assumed, also in large trials.[41,46] These baseline imbalances are often dismissed with the argument that they probably occurred by chance. However, chance imbalances can also influence the trial results. We recommend that baseline imbalances are investigated quantitatively in all reviews of RCTs. The question will be how to select the relevant prognostic patient characteristics without increasing the workload too much. The investigators may want to select the most obvious prognostic patient characteristics or those that are reported most often, such as age and sex.

Outcomes measured with error

Blinding of patients, health care professionals and outcome assessors to treatment status is used to stimulate impartial measurement of outcomes in trials. However, when the active drug has specific side effects, treatment status can sometimes be guessed. This 'knowledge' about treatment status can then influence the measurement of the outcome.[47] This type of bias is called information bias in the context of observational studies and trials. Studies are less susceptible to this type of bias when objective outcomes are used.

The Cochrane advises reviewers to take the subjectiveness of outcome measures into account when assessing the adequacy of blinding methods.[1] However, the use of objective measures is not included as a specific item in the Cochrane risk of bias tool 2.0, or other risk of bias tools.[2,43,48]

In antipsychotic trials for dementia, the persons involved may be 'unblinded' by fairly specific side effects such as stiffness, drooling and drowsiness. Objective outcomes may provide more valid results than subjective outcomes. We assessed the effects of antipsychotics for neuropsychiatric symptoms superimposed in dementia, using not only subjective outcomes such as a change in NPS, EPS, and somnolence, but also objective outcomes such as the use of additional psychotropic medication and the drop-out due to adverse events. We found that antipsychotics were not effective when using objective outcomes, while subjective outcomes suggested otherwise. Subjective outcomes suggested an increased risk of side effects for conventional versus atypical antipsychotics, while objective outcomes did not.

However, subjective outcomes were reported more frequently than objective outcomes, because they were often the primary outcome. Therefore, reviewers are advised to include frequently reported objective outcome measures, such as drop-out, whenever possible. Also, guidelines need to base recommendations preferably on the effects of interventions on objective outcomes.

Further considerations

This thesis quantified methods for assessing risk of bias in trials as part of reviews. A few considerations can be identified and are described in this paragraph.

Overall, a small number of trials was included in the studies (17 to 38 trials). Some (sub)analyses were performed on only a fraction of those trials due to missing information within the trials. Missing information was observed especially in trials that were published before the 2000s. Therefore some (sub)analyses were performed on less than 10 observations, the recommended minimum, resulting

in very imprecise estimates with wide confidence intervals. Conclusions based on such results should be interpreted with caution. For example, in the study about the use of run-in periods, we found different effects of antipsychotics on efficacy and side-effects in trials with and without a run-in phase. However, due to the small number of studies included, confidence intervals were very wide and overlapped, and conclusions could not be drawn with certainty.

I did not study reporting bias explicitly in this thesis. Reporting bias occurs when systematic differences exist between reported and unreported findings.[1] I observed during the studies that many articles about antipsychotic use in patients with dementia had missing information. E.g. when effects on psychosis, agitation or adverse events have not been reported. This information is crucial for readers to critically assess the effect of antipsychotics on clinical outcomes adequately.

Furthermore, a source of bias that was not studied in this this thesis is patient drop-out during a study, which can lead to incomplete outcome data. This can lead to biased results when the withdrawal is related to treatment group and outcome, and the participants are omitted from the analyses.[1] A common way to avoid this source of selection bias is to include information on all participants who underwent randomization in the groups to which they were originally allocated (intention-to-treat analysis).[36] Another quantitative method to investigate this type of bias is to study withdrawal itself as an outcome. When patients drop out during a study, the overall clinical benefit was probably insufficient due to either lack of efficacy, or adverse events.[50]

When substantial information is missing, the risk of bias cannot be studied adequately. In the 2000s, online trial registration sites, CONSORT reporting requirements and the risk of bias assessment were introduced, to ensure adequate trial conduct and complete reporting. Although the reporting of trials seems to have improved over the years, the reporting of some aspects, such as randomization procedures, baseline information for all randomized patients and sample size calculations, still remain behind.[49] The quantitative assessment described in this thesis depend how well trial methods and results are reported. More complete reporting will ensure a better quantitative assessment of the risk of bias.

Recommendations for future research

It is recommended that the effect of bias in trials on the results of a review is investigated quantitatively in every review. This thesis shows examples of how to objectively assess bias in reviews of antipsychotic trials for dementia. These methods are applicable in other reviews as well. Table 1 presents an overview of recommendations for researchers performing a review based on the methods

described in this thesis. Also, prerequisites are stated for reviewers implementing the recommendations given.

Table 1. Recommendations for trialists and reviewers

Section	*Recommendations for reviewers*	*Prerequisites*
PICO	Patients and outcomes should be defined a priori and accurately (using PICO), and trials should be selected accordingly.	Patients included in the studies should be representative of the patients in clinical practice.
Sample sizes	Quantitatively assess the variation of sample sizes, the size of the reported treatment effects, and the association between study characteristics and sample size between studies.	All trial reports should include a sample size calculation, and effect sizes with 95% confidence intervals to avoid large sample size fallacy.
Run-in	The risk of bias assessment could cover the presence of a run-in period. Analyze the effect of a run-in period on the reported results with sensitivity analysis.	Deselection of patients after run-in should be avoided. If run-in and deselection is used nonetheless, the number and characteristics of deselected patients should be described.
Randomization	Baseline imbalances can be assessed objectively as part of systematic reviews with metaregression. It can be helpful to select a limited set of the most obvious prognostic patient characteristics.	Baseline characteristics prognostic of the outcomes for all randomized patients per treatment group should be reported. Randomization procedures such as the random sequence generation and concealment of allocation should be reported. Investigators, physicians, and patients should not be able to foresee allocation, or change the allocation itself.
Objective outcomes	When subjective measures are included as primary outcome, sensitivity analyses could be performed using objective measures to compare study results.	Including objective outcomes as primary outcome instead of subjective outcomes whenever possible.

When a trial or review is finished, publication in a journal is usually the next step. Publication is not only important for researchers, but also for journals; the number and quality of the articles determines who reads and cites the content.[51] Journals can make a contribution to a more transparent research field by making more publications freely accessible. Open access of research publication increases visibility and reuse of academic research results.[52] However, open access does not solve publication bias, the withholding of negative results from publication.[53] A way

to reduce publication bias by journals is by publishing high-quality studies regardless of novelty or unexciting results.[54] If more study results, also negative results, are published and made available, reviews including these studies will be more accurate. Therefore, researchers should also be motivated to publish studies with negative results as well. Thus, collaboration of both researchers and journals is needed to prevent publication bias.

This thesis examined issues related to review methodology and the objective methods that can be used to improve the quality. Performing a trial or review is a time-consuming activity, especially if high quality standards are to be upheld. Upcoming technologies such as machine learning and artificial intelligence (AI) gain popularity because they could substantially reduce reviewer workload.[55] Also, with the development of algorithms for AI and machine learning the process of clinical trials and reviews can be modernized.[56] In clinical trials, these algorithms can for example simulate study control arms.[56] In reviews, the risk of bias assessment can be automated with a machine learning system called 'RobotReviewer', generating equal results as to manually identified assessments.[55] It is conceivable that machine learning can also be used to extract quantitative data about bias from studies, leaving time for reviewers to improve other aspects of the research. Recently, a systematic review was completed in two weeks using automation tools, among which 'RobotReviewer'.[57] However, new innovate systems may also bring new risks of bias.[58] Systems like AI and machine learning, learn to make decisions based on training data. If this training data includes biased human decisions or over- or underrepresented groups, than the training data may be flawed and output data might be biased as well.[59] Researchers should be aware of these biases when using these systems to aid their research.

Conclusion

Since the 70s the improvement of review methodology has come a long way. However, as illustrated in this thesis, some aspects can still be improved as bias was found in antipsychotic drug trials for dementia. The topic in this thesis served as a test case, but general conclusions can be drawn from the results. For reviewers it is recommended that the effect of bias in trials is investigated quantitatively in every review. Further, if possible, the quantitative assessment should be repeated in reviews of other interventions, other patient populations, and more recently published RCTs. For journals it is recommended to published high-quality studies regardless of the direction of results (positive or negative). In addition, in the future, emerging innovative technologies can support the review process in general, and the risk of bias assessment in particular.

References

1. Higgins J, Savovic J, Page M, Elbers R, Sterne J. Chapter 8: Assessing risk of bias in a randomized trial. In: Higgins J, Thomas J, Chandler J, et al., eds. Cochrane Handbook for Systematic Reviews of Interventions. Version 6. Cochrane; 2019. https://training.cochrane.org/handbook/current/chapter-08.
2. Jadad A, Moore A, Carroll D, et al. Assessing the quaility of reports of randomized clinical trials: Is blinding necessary? Control Clin Trials. 1996;17:1-12.
3. Pubrica-academy. What Is The Risk Of Bias Assessment And Different Tools Used To Assess Systematic Review? https://pubrica.com/academy/research/what-is-the-risk-of-bias-assessment-and-different-tools-used-to-assess-systematic-review/. Published 2020. Accessed March 16, 2021.
4. Moseley AM, Rahman P, Wells GA, et al. Agreement between the Cochrane risk of bias tool and Physiotherapy Evidence Database (PEDro) scale: A meta-epidemiological study of randomized controlled trials of physical therapy interventions. PLoS One. 2019;14(9):1-16. doi:10.1371/journal.pone.0222770
5. Hartling L, Bond K, Vandermeer B, Seida J, Dryden DM, Rowe BH. Applying the risk of bias tool in a systematic review of combination long-acting beta-agonists and inhaled corticosteroids for persistent asthma. PLoS One. 2011;6(2):2-7. doi:10.1371/journal.pone.0017242
6. Junod S. FDA and Clinical Drug Trials: A Short History. In: Davies M, Kerimani F, eds. A Quick Guide to Clinical Trials. Washington: Bioplan, Inc; 2008:25-55. https://www.fda.gov/media/110437/download.
7. Rehman W, Arfons LM, Lazarus HM. The rise, fall and subsequent triumph of thalidomide: Lessons learned in drug development. Ther Adv Hematol. 2011;2(5):291-308. doi:10.1177/2040620711413165
8. Shah HM, Chung KC. Archie cochrane and His vision for evidence-based medicine. Plast Reconstr Surg. 2009;124(3):982-988. doi:10.1097/PRS.0b013e3181b03928
9. Poklepovic Pericic T, Tanveer S. Why systematic reviews matter. A brief history, overview and practical guide for authors. 2019. https://www.elsevier.com/connect/authors-update/why-systematic-reviews-matter.
10. Cochrane A. Effectiveness and Efficiency: Random Reflections on Health Services. Nuffield Trust; 1972.
11. Chalmers I. The Cochrane collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. Ann N Y Acad Sci. 1993;Dec(703):156-163.
12. Schünemann H, Vist G, Higgins J, et al. Chapter 15: Interpreting results and drawing conclusions. In: Higgins J, Thomas J, Chandler J, et al., eds. Cochrane Handbook for Systematic

Reviews of Interventions. Version 6. Cochrane; 2019. https://training.cochrane.org/handbook/current/chapter-15.
13. Montori V, Devereaux P, Adhikari N, et al. Randomized trials stopped early for benefit: a systematic review. JAMA. 2005;294(17):2203-2209. doi:10.1001/jama.294.17.2203
14. ClinicalTrials.gov. ClinicalTrials.gov Background. https://www.clinicaltrials.gov/ct2/about-site/background. Published 2018. Accessed March 16, 2021.
15. Wikipedia. List of clinical trial registries. https://en.wikipedia.org/wiki/List_of_clinical_trial_registries. Published 2020. Accessed March 16, 2021.
16. ClinicalTrials.gov. Why Should I Register and Submit Results? https://clinicaltrials.gov/ct2/manage-recs/background. Published 2020. Accessed March 16, 2021.
17. Methods C. Risk of Bias 2 (RoB 2) tool. https://methods.cochrane.org/risk-bias-2. Published 2021. Accessed March 16, 2021.
18. The Cochrane Collaboration. About the handbook. https://training.cochrane.org/handbook#about-the-handbook. Published 2021. Accessed May 13, 2021.
19. Anthony J. Cochrane Library: an improved online platform to guide health decision-making across the world. Cochrane collaboration. https://www.cochrane.org/news/cochrane-library-improved-online-platform-guide-health-decision-making-across-world. Published 2018. Accessed May 13, 2021.
20. Useem J, Brennan A, LaValley M, et al. Systematic differences between cochrane and non-cochrane meta-analyses on the same topic: A matched pair analysis. PLoS One. 2015;10(12):1-17. doi:10.1371/journal.pone.0144980
21. Schunemann H, Higgins J, Vist G, et al. Chapter 14: Completing 'Summary of findings' tables and grading the certainty of the evidence. In: Cochrane Handbook for Systematic Reviews of Interventions. Version 6. Cochrane; 2019. https://training.cochrane.org/handbook/current/chapter-14.
22. Luijendijk H, De Bruin N, Hulshof T, Koolman A. Terminal illness and the increased mortality risk of conventional antipsychotics in observational studies: a systematic review. Pharmacoepidemiol Drug Saf. 2016;25(2):113-122.
23. Luijendijk HJ, Page MJ, Burger H, Koolman X. Assessing risk of bias: a proposal for a unified framework for observational studies and randomized trials. BMC Med Res Methodol. 2020;20(1):237. doi:10.1186/s12874-020-01115-7
24. McKenzie J, Brennan S, Ryan R, Thomson H, Johnston R, Thomas J. Chapter 3: Defining the criteria for including studies and how they will be grouped for the synthesis. In: Higgins J, Thomas J, Chandler J, et al., eds. Cochrane Handbook for Systematic Reviews of Interventions. Version 6. Cochrane; 2019. https://training.cochrane.org/handbook/current/chapter-03.
25. Guyatt GH, Oxman AD, Kunz R, et

al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol. 2011;64(4):395-400. doi:10.1016/j.jclinepi.2010.09.012
26. Higgins J, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions. Version 5. The Cochrane Collaboration; 2011. https://training.cochrane.org/handbook.
27. Kaplan RM, Chambers DA, Glasgow RE. Big data and large sample size: A cautionary note on the potential for bias. Clin Transl Sci. 2014;7(4):342-346. doi:10.1111/cts.12178
28. Lantz B. The large sample size fallacy. Scandanavian J Caring Sci. 2013;27(1):487-492.
29. Lin M, Henry C. Lucas J, Galit Shmueli. Research Commentary—Too Big to Fail: Large Samples and the p-Value Problem. Inf Syst Res. 2013;24(4):906-917.
30. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. New York: Routledge; 1988. doi:https://doi.org/10.4324/9780203771587
31. Ballard C, Waite J, Birks J. Atypical antipsychotics for aggression and psychosis in Alzheimer ' s disease (Review). Cochrane Libr. 2006;(1):1-135.
32. Ma H, Huang Y, Cong Z, et al. The Efficacy and Safety of Atypical Antipsychotics for the Treatment of Dementia: A Meta-Analysis of Randomized Placebo-Controlled Trials. J Alzheimer's Dis. 2014;42:915-937.
33. Schneider L, Dagerman K, Insel P. Risk of Death With Atypical Antipsychotic Drug Treatment for Dementia. J Am Med Assoc. 2005;294(15):1934-1943.
34. Thompson B. Effect Sizes, Confidence Intervals, and Confidence Intervals for Effect Sizes. Psychol Sch. 2007;44(5).
35. Lee DK. Interval and effect size. Korean J Anesthesiol. 2016;69(6):555-562.
36. Schulz K, Altman D, Moher D. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol. 2010;63(8):1-6. doi:10.1016/j.ijsu.2010.09.006
37. Ulmer M, Robinaugh D, Friedberg J, Lipsitz S, Natarajan S. Usefulness of a run-in period to reduce drop-outs in a randomized controlled trial of a behavioral intervention. Contemp Clin Trials. 2008;29(5):705-710.
38. Laursen DRT, Paludan-Müller AS, Hróbjartsson A. Randomized clinical trials with run-in periodsfrequency, characteristics and reporting. Clin Epidemiol. 2019;11:169-184. doi:10.2147/CLEP.S188752
39. Agid O, Siu C, Potkin S, Kapur S, Watsky E, Vanderburg D. Meta-regression analysis of placebo response in antipsychotic trials 1970-2010. Am J Psychiatry. 2013;170(11):1335-1344.
40. Higgins J, Altman D, Sterne J. Chapter 8: Assessing risk of bias in included studies. In: Cochrane Handbook. The Cochrane Collaboration; 2014. https://handbook-5-1.cochrane.org/chapter_8/8_assessing_risk_of_bias_in_included_studies.htm.
41. Clark L, Fairhurst C, Hewitt C, et al. A methodological review of recent meta-analyses has found significant

heterogeneity in age between randomized groups. J Clin Epidemiol. 2014;67(9):1016-1024.
42. Hróbjartsson A, Boutron I, Turner L, Altman D, Moher D. Assessing risk of bias in randomised clinical trials included in Cochrane Reviews: the why is easy, the how is a challenge [editorial]. Cochrane Database Syst Rev. 2013;4.
43. The Cochrane Collaboration. The Cochrane Collaboration's tool for assessing risk of bias. 2005:1-4.
44. Savovic J, Weeks L, Sterne J, et al. Evaluation of the Cochrane Collaboration's tool for assessing the risk of bias in randomized trials: focus groups, online survey, proposed recommendations and their implementation. Syst Rev. 2014;3(1):1-12. Systematic Reviews.
45. Gøtzsche PC, Jørgensen KJ. Screening for breast cancer with mammography. Cochrane Database Syst Rev. 2013;2013(6). doi:10.1002/14651858.CD001877.pub5
46. Luijendijk H, Hulshof T. Baseline differences in the SAVOR trial. Diabetes Obes Metab. 2015;17(12):1202.
47. Baethge C, Assall OP, Baldessarini RJ. Systematic review of blinding assessment in randomized controlled trials in schizophrenia and affective disorders 2000-2010. Psychother Psychosom. 2013;82(3):152-160. doi:10.1159/000346144
48. van Tulder M, Furlan A, Bombardier C, Bouter L. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. Spine (Phila Pa 1976). 2003;28:1290-1299.
49. Latif L, Eduardo J, Amadera D, Pimentel D, Pimentel T, Fregni F. Sample Size Calculation in Physical Medicine and Rehabilitation: A Systematic Review of Reporting, Characteristics, and Results in Randomized Controlled Trials. Arch Phys Med Rehabil. 2011;92(2):306-315.
50. Schneider L, Dagerman K, Insel P. Efficacy and adverse effects of atypical antipsychotics for dementia: meta-analysis of randomized, placebo-controlled trials. Am J Geriatr Psychiatry. 2006;14(March):191-210.
51. McGlashan D, Hadley C. How journal editors can improve journal quality. Inlexio. https://www.inlexio.com/how-journal-editors-can-improve-journal-quality/. Published 2016. Accessed May 10, 2021.
52. Openaccess.nl. What is open access? https://www.openaccess.nl/en/what-is-open-access. Published 2021. Accessed May 10, 2021.
53. Joober R, Schmitz N, Annable L, Boksa P. Publication bias: What are the challenges and can they be overcome? J Psychiatry Neurosci. 2012;37(3):149-152. doi:10.1503/jpn.120065
54. Song F, Hooper L, Loke YK. Publication bias: What is it? How do we measure it? How do we avoid it? Open Access J Clin Trials. 2012;5(1):71-81. doi:10.2147/OAJCT.S34419
55. Marshall IJ, Kuiper J, Wallace BC. RobotReviewer: Evaluation of a system for automatically assessing bias in clinical trials. J Am Med Informatics Assoc. 2016;23(1):193-201. doi:10.1093/jamia/ocv044

56. Shah P, Kendall F, Khozin S, et al. Artificial intelligence and machine learning in clinical development: a translational perspective. npj Digit Med. 2019;2(1). doi:10.1038/s41746-019-0148-3
57. Clark J, Glasziou P, Del Mar C, Bannach-Brown A, Stehlik P, Scott AM. A full systematic review was completed in 2 weeks using automation tools: a case study. J Clin Epidemiol. 2020;121:81-90. doi:10.1016/j.jclinepi.2020.01.008
58. Jaspreet. Understanding and Reducing Bias in Machine Learning. Towards data science. https://towardsdatascience.com/understanding-and-reducing-bias-in-machine-learning-6565e23900ac. Published 2019. Accessed May 10, 2021.
59. Manyika J, Silberg J, Presten B. What Do We Do About the Biases in AI? Harvard Bisiness Review. https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai. Published 2019. Accessed May 14, 2021.

Appendices.

Scientific summary

Wetenschappelijke samenvatting

Dankwoord

Curriculum Vitae

Previous dissertations

Scientific summary

A standard part of a systematic review is to critically appraise the quality of the included studies. Despite significant improvements of review methods over the years, the assessment tools have a qualitative nature and depend on the level of methodological knowledge of the reviewer. This thesis quantitatively appraised the methods of antipsychotic drug trials for patients with dementia, and how the quality affected the reported results. It consists of two parts. The first part addressed the clinical relevance of the reported effects.

In chapter 2, the mortality risk of conventional antipsychotics in elderly patients with dementia or at risk of delirium, in randomized controlled trials was investigated. Health authorities warned against use of conventional antipsychotics in dementia, because numerous observational studies reported an increased risk of mortality for conventional antipsychotics in elderly patients. We performed a meta-analysis of 17 randomized trials in 2387 patients. In contrast to reviews of observational studies, we did not find that conventional antipsychotics in general, or haloperidol in particular, increase the risk of mortality in elderly patients with dementia.

In chapter 3, the pooled efficacy of antipsychotics in patients with dementia and agitation or psychosis was assessed. Prior reviews on the efficacy of antipsychotics in dementia included trials that enrolled mixed patient populations with various neuropsychiatric symptoms (NPS) and used outcome scales that were not specific for the target symptom (agitation or psychosis). We investigated how this selection of studies might have affected the pooled efficacy. The meta-epidemiological study showed that conventional antipsychotics might have a small effect in agitated patients (the population of interest) on agitation (the outcome of interest) and in psychotic patients (the population of interest) on psychosis (the outcome of interest). This was not the case for atypical antipsychotics. The pooled efficacy of atypical antipsychotics was larger when based on trials that included patients without these target symptoms and used generic outcome scales.

Chapter 4 focused on large sample size fallacy. Larger sample sizes provide more power to identify a treatment effect that is really present. A disadvantage of (very) large sample size is that a difference in outcomes between the groups will become statistically significant, no matter how (very) small or clinically meaningless it is. If these results are interpreted as clinically relevant, large sample size fallacy occurs. Therefore, sample size, size of treatment effect and general study characteristics related to sample size were investigated in 51 antipsychotic trials for dementia. Overall, sample size calculations were poorly reported in these trials. In addition, in

33 placebo-controlled trials that tested atypical antipsychotics, we found statistically significant but clinically negligible effects in relatively large populations; this may have given rise to large sample size fallacy. On the other hand, the 18 head-to-head trials were all underpowered.

In part two of this thesis, bias in antipsychotic trials in dementia was studied. Chapter 5, presents a study about the association of run-in periods with reported treatment effects. Run-in periods are used to identify placebo-responders and washout medication that patient already in the period between screening and before randomization. At the end of a run-in period, patients can be deselected for participation in a trial. We performed a meta-epidemiological study and included 35 trials. Trials with a run-in period showed a lower risk of somnolence, extrapyramidal symptoms, and mortality than trials without run-in. Meta-analyses should include sensitivity-analyses of trials with and without run-in periods.

In chapter 6, the presence of baseline imbalances in placebo-controlled trials of atypical antipsychotics in dementia was assessed. Randomization is used to ensure that chance instead of patient characteristics determine treatment assignment. Therefore, we assessed the presence of baseline imbalances that had occurred despite randomization, and their association with neuropsychiatric symptoms, extrapyramidal symptoms, and mortality. We included 23 placebo-controlled trials of atypical antipsychotics in 5853 patients with dementia. All trials reported the randomization procedures incompletely and showed heterogeneous baseline imbalances. Trials with missing information about baseline characteristics seemed to have a more favorable pooled efficacy and lower risk of extrapyramidal symptoms than trials that reported this information. Imbalances were not significantly associated with risk of mortality. Baseline imbalances need to be assessed objectively as part of systematic reviews.

In chapter 7, the effectiveness and risk of side effects of antipsychotics has been reported in terms of subjectively and objectively measured outcomes. Knowledge about treatment status can influence the measurement of subjective outcomes when blinding for the allocated treatment is suboptimal. Therefore, we compared subjective with objective outcomes in trials of conventional and atypical antipsychotics in dementia. We performed a meta-epidemiological study of 38 randomized, placebo-controlled trials. Antipsychotics were found to be effective with subjective symptom scales (change in neuropsychiatric symptoms and response rate), but not with more objective outcomes (overall dropout and additional psychotropic use). Subjective measures of side effects (extrapyramidal symptoms and somnolence) suggested that conventional antipsychotics had a higher risk than atypical antipsychotics, but

objective measures (dropout due to adverse events, medication use for extrapyramidal symptoms, and participants falling) did not. Future trials and reviews need to address potential information bias by including objective measured outcomes.

This thesis shows examples of quantitative assessments of bias in reviews of antipsychotic trials for dementia. The presented methods in this thesis can be applied in other reviews of other interventions, other patient populations, and in more recently published RCTs.

Wetenschappelijke samenvatting

Een standaard onderdeel van een systematische review is de kritische beoordeling van de kwaliteit van de geïncludeerde onderzoeken. Ondanks de aanzienlijke verbeteringen van de reviewmethoden door de jaren heen, hebben de beoordelingsinstrumenten voor de kwaliteit van de trials een kwalitatief karakter. Daardoor is de beoordeling afhankelijk van het niveau van methodologische kennis van de beoordelaar. Dit proefschrift presenteert een aantal studies waarin met kwantitatieve methoden de kwaliteit van antipsychotische geneesmiddelen trials is onderzocht. Het proefschrift bestaat uit twee delen. Het eerste deel gaat in op de klinische relevantie van de gerapporteerde effecten.

In hoofdstuk 2 werd het sterfterisico van conventionele antipsychotica, bij oudere patiënten met dementie of hoog risico op delier, in gerandomiseerde gecontroleerde trials onderzocht. Gezondheidsautoriteiten waarschuwden voor het gebruik van conventionele antipsychotica bij dementie, omdat meerdere observationele studies een verhoogd risico op sterfte voor conventionele antipsychotica bij oudere patiënten hadden gemeld. Wij voerden een meta-analyse uit van 17 gerandomiseerde trials bij 2387 patiënten. In tegenstelling tot reviews van observationele studies, vonden wij niet een verhoogd risico op sterfte bij gebruik van conventionele antipsychotica in het algemeen, of haloperidol in het bijzonder, door oudere patiënten met dementie. In hoofdstuk 3 werd de werkzaamheid van antipsychotica bij patiënten met dementie en agitatie of psychose onderzocht. Eerdere reviews naar de werkzaamheid van antipsychotica bij dementie omvatten onderzoeken waarin gemengde patiëntenpopulaties met verschillende neuropsychiatrische symptomen (NPS) werden opgenomen en waarbij uitkomstschalen werden gebruikt die niet specifiek waren voor het doelsymptoom (agitatie of psychose). We hebben onderzocht hoe deze selectie van onderzoeken de gepoolde werkzaamheid zou kunnen hebben beïnvloed. De meta-epidemiologische studie rapporteerde dat conventionele antipsychotica een klein effect kunnen hebben bij geagiteerde patiënten (de populatie van interesse) op agitatie (de uitkomst van interesse) en bij psychotische patiënten (de populatie van interesse) op psychose (de uitkomst van interesse). Dit was niet het geval voor atypische antipsychotica. De gepoolde werkzaamheid van atypische antipsychotica was groter wanneer deze is gebaseerd op onderzoeken bij patiënten zonder de doelsymptomen en op generieke uitkomstschalen.

Hoofdstuk 4 richtte zich op de effecten van grote steekproeven. Grotere trials hebben meer power om een behandeleffect dat echt aanwezig is te identificeren. Een nadeel van een (zeer) grote steekproef is dat een verschil in uitkomsten tussen de groepen statistisch significant wordt, hoe (zeer) klein of klinisch irrelevant het

ook is. Als deze resultaten als klinisch relevant worden geïnterpreteerd, treedt er 'large sample size fallacy' op. Daarom werd de steekproefomvang, de grootte van het behandeleffect en algemene studiekenmerken gerelateerd aan steekproefomvang onderzocht in 51 antipsychotica trials voor dementie. Over het algemeen werden berekeningen van de steekproefomvang slecht gerapporteerd in de trials. In 33 placebogecontroleerde trials waarin atypische antipsychotica werden getest, vonden we statistisch significante maar klinisch verwaarloosbare effecten bij relatief grote populaties; dit kan aanleiding hebben gegeven tot 'large sample size fallacy'. De 18 head-to-head trials waren allemaal 'underpowered'.

In deel twee van dit proefschrift is bias in antipsychotica trials bij dementie onderzocht. Hoofdstuk 5 presenteert een onderzoek naar de associatie van een 'run-in' periode met gerapporteerde behandeleffecten. Run-in periodes worden gebruikt om patiënten die op placebo reageren te identificeren en medicatie die al gebruikt wordt te stoppen (washout) in de periode tussen screening en randomisatie. Na de run-in periode kan deselectie van patiënten voor deelname aan de studie plaatsvinden. We voerden een meta-epidemiologisch onderzoek uit, waarin 35 onderzoeken werden geïncludeerd. Trials met een run-in periode lieten een iets grotere afname van neuropsychiatrische symptomen zien dan studies zonder run-in periode. Bovendien lieten onderzoeken met een run-in periode een lager risico op sufheid, extrapiramidale symptomen en sterfte zien dan onderzoeken zonder run-in periode. Meta-analyses zouden subanalyses van trials met en zonder run-in periodes moeten includeren.

In hoofdstuk 6 werd de aanwezigheid van baseline verschillen in placebo-gecontroleerde trials met atypische antipsychotica bij dementie beoordeeld. Randomisatie wordt gebruikt om ervoor te zorgen dat toeval in plaats van patiëntkenmerken de behandeltoewijzing bepalen. Wij hebben onderzocht of baseline verschillen aanwezig waren ondanks randomisatie, en hun verband met neuropsychiatrische symptomen, extrapiramidale symptomen en sterfte. We includeerden 23 placebogecontroleerde trials met atypische antipsychotica bij 5853 patiënten met dementie. Alle trials rapporteerden de randomisatieprocedures onvolledig en vertoonden heterogene baseline verschillen. Onderzoeken met ontbrekende informatie over baseline verschillen leken een gunstiger gepoolde werkzaamheid en een lager risico op extrapiramidale symptomen te hebben dan onderzoeken die deze informatie rapporteerden. Verschillen waren niet significant geassocieerd met het risico op sterfte. Baseline verschillen moeten objectief worden onderzocht als onderdeel van systematische reviews.

In hoofdstuk 7 werden de effectiviteit en het risico op bijwerkingen van antipsychotica in termen van subjectief en objectief gemeten uitkomsten onderzocht. Kennis over de behandelstatus van een deelnemer kan de meting van subjectieve uitkomsten beïnvloeden. Daarom vergeleken we subjectieve uitkomsten met objectieve uitkomsten in trials van antipsychotica bij dementie. We voerden een meta-epidemiologische studie uit van 38 gerandomiseerde, placebo-gecontroleerde studies. Antipsychotica bleken effectief te zijn bij subjectieve symptoomschalen (zoals verandering in neuropsychiatrische symptomen en responspercentage), maar niet bij objectievere uitkomsten (zoals algehele uitval tijdens studie en aanvullend psychotroop gebruik). Subjectieve metingen van bijwerkingen (zoals extrapiramidale symptomen en slaperigheid) suggereerden dat conventionele antipsychotica een hoger risico hadden dan atypische antipsychotica, maar objectieve metingen (zoals uitval door bijwerkingen, medicatiegebruik voor extrapiramidale symptomen en valincidenten van deelnemers) niet. In toekomstige trials en reviews moet mogelijke informatiebias worden aangepakt door objectieve uitkomsten op te nemen.

Dit proefschrift toont aan hoe bias kwantitatief kan worden beoordeeld en onderzocht in antipsychotica trials voor dementie. De gepresenteerde methoden zijn ook toepasbaar in andere reviews over andere interventies, andere patiëntenpopulaties en recenter gepubliceerde gerandomiseerde onderzoeken.

Dankwoord

Nooit gedacht dat ik nog eens een dankwoord zou schrijven voor mijn proefschrift. Maar het is gelukt! Voor het mogelijk maken van dit mooie resultaat wil ik dan ook graag de volgende personen hartelijk bedanken:

Professor Zuidema, beste Sytse, bedankt voor de mogelijkheid om dit proefschrift af te ronden via deze, toch wel bijzondere weg. Je ziet altijd mogelijkheden en ondanks je drukke agenda maak je tijd vrij om vragen te beantwoorden of feedback te geven. Ik wil je heel erg bedanken voor alle begeleiding de afgelopen jaren.

Dr. Luijendijk, beste Dika, bedankt dat je me tijdens dit gehele proces hebt begeleid. Vanaf het begin, met mijn scriptie voor mijn masteropleiding, tot aan dit proefschrift. Het was vast niet altijd even makkelijk, met mijn wisselende motivatie en het herhaaldelijk uitleggen van bepaalde concepten, maar ik vond je begeleiding heel waardevol. Bedankt voor je geduldige en gepassioneerde uitleg. Er zijn vast niet veel promovendi die zulke support tijdens hun proefschrift ontvangen, dus daar ben ik heel dankbaar voor.

Alle co-auteurs die hebben bijgedragen aan de artikelen. Veel dank voor jullie input en goede feedback.

Dank aan mijn collega's bij het RIVM. Voor alle interesse tijdens het gehele proces en voor input op mijn vragen.

Mijn lieve vriendengroep, Sarah, Boris, Jesse, Fien, Alexander, Ilena, Teun, Marije, Marloes, Laurens… Holy Moly het boekje is er! Bedankt voor jullie input voor de conclusie. Maar bovenal bedankt voor al die geweldige feestjes, etentjes, gezellige momenten en mooie herinneringen. Hopelijk volgen er nog vele.

Anne, Inge en Anouk, mijn lieve vriendinnen. Elkaar tig jaar geleden leren kennen bij het voetbal, maar inmiddels zijn we al heel wat kreupele knieën verder. Wat is het ontzettend fijn dat ik altijd bij jullie kan aankloppen als er iets is. Bedankt voor alle goede gesprekken (vooral die bij Jan) en dat we na al die jaren nog steeds zulke goede vriendinnen zijn.

Beste Marcel, ook jij verdient een plekje in dit dankwoord. We zijn al goede vrienden sinds groep 1 van de basisschool en dat is inmiddels dus al bijna 30 jaar. Waar blijft de tijd, op naar de volgende 30!

Sarah en Inge, mijn paranimfen, twee toppers aan mijn zijde! Ik weet dat ik altijd op jullie kan bouwen en dat is heel fijn. Bedankt voor alle hulp bij de afronding.

Joris, mijn grote kleine broer. Samen hebben we mooie avonturen beleefd, en wat gaaf dat je nu je eigen avontuur aangaat in Australië. Hopelijk kun je Felice snel weer in je armen sluiten om dit avontuur af te maken. Lieve Joris en Felice, of jullie nou dichtbij zijn of ver weg, afstand maakt niet uit, want ik weet dat ik op jullie kan rekenen.

Anke en Henk, mijn lieve schoonouders, dank voor al jullie steun. Niet alleen tijdens het afronden van mijn proefschrift maar ook daarnaast. Henk, bedankt voor de hulp bij het maken van de cover. Wat is hij mooi geworden.

Jos en Gerdy, lieve pap en mam, waar te beginnen? Een stukje in dit dankwoord is eigenlijk te kort om te benoemen hoe dankbaar ik ben voor alle liefde, steun en mogelijkheden die jullie mij hebben gegeven. Dit zorgt voor de beste basis die er is, en kijk waar dat op uit is gekomen! Ik hoop dat Lars en Isa later net zo trots zijn op mij als ik op jullie ben.

Allerliefste David, dank dat je er altijd voor me bent. Dat je zegt dat je trots op me bent en dat het goed komt, ook als ik daar zelf aan twijfel. Mijn rots in de branding. Wat hebben we het fijn samen met onze kleine man en sinds kort ook ons kleine meisje. Samen met jou nog vele jaren. Ik hou van jou!

Lieve Lars, zonder al jouw knuffels was het nooit gelukt.
Welkom lieve Isa. Wat ben ik blij dat jij er bent.
Ik hou van jullie!

Curriculum Vitae

Tessa was born on December 19th 1988, in Zevenaar. She grew up in Zevenaar with one younger brother and attended secondary education at Liemers College. After graduating from Havo in 2006, she moved to Kearney, Missouri, USA, and completed a senior year in high school. In 2007 she moved back to Zevenaar and started her bachelor's degree in nursing at the Hogeschool van Arnhem en Nijmegen (HAN) at Nijmegen. After several internships, she graduated in 2011. In the same year, she started her premaster Health Sciences at Vrije Universiteit (VU) Amsterdam and completed her master's degree the following year. Her master thesis resulted in her first scientific article: 'The Mortality Risk of Conventional Antipsychotics in Elderly Patients: A Systematic Review and Meta-analysis of Randomized Placebo-Controlled Trials'.

After graduating in Amsterdam, she started working at the University Medical Center Groningen (UMCG), where she worked on several articles included in this thesis. In 2016, she was looking for an opportunity to develop her research skills in a different field and started working at the National Institute for Public Health and the Environment (RIVM). In addition, she continued working on the research projects that she started in Groningen. This resulted in several publications. In 2020 she decided to finish her Ph.D. thesis. She currently still works at RIVM.

Previous dissertations

Research Institute SHARE

This thesis is published within the **Research Institute SHARE** (Science in Healthy Ageing and healthcaRE) of the University Medical Center Groningen / University of Groningen.

Further information regarding the institute and its research can be obtained from our internet site: http://www.share.umcg.nl/

More recent theses can be found in the list below.
(supervisors are between brackets)

2021

Peters RM
Total hip arthroplasty: register based evaluation of current practice
(prof SK Bulstra, dr M Stevens, dr WP Zijlstra)

Simmelink L
Combined arm-leg ergometry in persons with a lower limb amputation
(prof R Dekker, prof JHB Geertzen, prof LHV van der Woude)

Valk TA
When multiple become one; understanding the coordination of degrees of freedom in synergies in goal-directed pointing movements
(dr RM Bongers, prof E Otten, dr LJ Mouton)

Bos FM
Ecological momentary assessment as a clinical tool in psychiatry: promises, pitfalls, and possibilities
(prof M Wichers, prof R Bruggeman, dr JAJ van der Krieke, dr E Snippe)

Akbari F
Chronic pain in the context of the lives of dyads; cognitions, behaviors, and well-being
(prof M Hagedoorn, prof R Sanderman, dr M Dehghani)

Hepping AM
Grip on recovery after paediatric forearm fractures
(prof SK Bulstra, prof JHB Geertzen, dr M Stevens)

Beijers L
Parsing the heterogeneity of major depression
(prof RA Schoevers, dr KJ Wardenaar, dr HM van Loo)

Köhler TC
Providing color to the pharmacy technician; a new profession within the pharmacy team
(prof ADC Jaarsma, dr M Westerman)

Bunk SF
Frontal brain functioning and pain; possible underlying mechanisms of increased pain responses in age- and dementia-related cognitive impairment
(prof SU Zuidema, dr M Kunz)

Bosáková L
Breaking the cycle of poverty; routes to counteract intergenerational transmission of socioeconomic health differences
(prof SA Reijneveld, prof A Madarasová-Gecková)

Vendeloo SN van
Optimizing learning environments and resident well-being in postgraduate medical education
(prof PLP Brand, prof SK Bulstra, dr CCPM Verheyen)

Sampurna MTA
Improving the management of hyperbilirubinemia in a limited-resource area
(prof AF Bos, dr CV Hulzebos, dr PH Dijk, dr R Etika)

Siswanto JE
Retinopathy of prematurity; how to prevent retinopathy of prematurity in preterm infants in Indonesia?
(prof AF Bos, prof A Adisasmita, dr PH Dijk)

Kaper MS
Improving communication in healthcare for patients with low health literacy; building competencies of health professionals and shifting towards health literacy friendly organizations
(prof SA Reijneveld, prof AF de Winter)

Schuurmans J
Population-based expanded carrier screening reporting couple results only: a mixed methods approach
(prof IM van Langen, prof AM Lucassen, prof AV Ranchor, dr M Plantinga)

Vrijsen J
Towards dementia risk reduction among individuals with a parental family history of dementia
(dr N Smidt, prof SEJA de Rooij)

For earlier theses visit our website