TECHNICAL COMMENTARY

NeuRA Discover. Conquer. Cure.

SCHIZOPHRENIA LIBRARY

Estrogen

Introduction

Women tend to have a later age of onset of schizophrenia than men when onset is typically seen in early adulthood. There is also a peak in the onset of schizophrenia in women after menopause. Due to these observations, the hormone estrogen (or 'oestrogen') has been proposed to confer a protective effect for women, as estrogen levels in women decline with age particularly after menopause. This indicates adjunctive estrogen could be an effective treatment for the symptoms of schizophrenia.

Method

We have included only systematic reviews (systematic literature search, detailed methodology with inclusion/exclusion criteria) published in full text, in English, from the year 2000 that report results separately for people with a diagnosis of schizophrenia, schizoaffective disorder, schizophreniform disorder or first episode schizophrenia. Reviews were identified by searching the databases MEDLINE, EMBASE, CINAHL, Current Contents, PsycINFO and the Cochrane library. Hand searching reference lists of identified reviews was also conducted. When multiple copies of reviews were found, only the most recent version was included. Reviews with pooled data are prioritised for inclusion.

Review reporting assessment was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist, which describes a preferred way to present a meta-analysis[1]. Reviews rated as having less than 50% of items checked have been excluded from the library. The PRISMA flow diagram is a suggested way of providing information about studies included and excluded with reasons for exclusion. Where no flow diagram has been presented by individual reviews, but identified studies have been described in the text, reviews have been checked for this item. Note that early reviews may have been guided by less stringent reporting checklists than the PRISMA, and that some reviews may have been limited by journal guidelines.

Evidence was graded using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group approach where high quality evidence such as that gained from randomised controlled trials (RCTs) may be downgraded to moderate or low if review and study quality is limited, if there is inconsistency in results, indirect comparisons, imprecise or sparse data and high probability of reporting bias. It may also be downgraded if risks associated with the intervention or other matter under review are high. Conversely, low quality evidence such as that gained from observational studies may be upgraded if effect sizes are large or if there is a dose dependent response. We have also taken into account sample size and whether results are consistent, precise and direct with low associated risks (see end of table for an explanation of these terms)[2]. The resulting table represents an objective summary of the available evidence, although the conclusions are solely the opinion of staff of NeuRA (Neuroscience Research Australia).

Results

We found six systematic reviews that met inclusion criteria[3-8].

- Moderate to high quality evidence indicates small to medium-sized improvements in positive and negative symptoms in females with schizophrenia taking adjunctive estrogen. There were medium-sized improvements in symptoms in both males and females (including those postmenopausal) with adjunctive raloxifene compared to placebo, with no differences in depression symptoms or cognition.

Margarete Ainsworth Building, Barker Street, Randwick NSW 2031. Phone: 02 9399 1000. Email: info@neura.edu.au

Begemann MJH, Dekker CF, van Lunenburg M, Sommer IE

Estrogen augmentation in schizophrenia: A quantitative review of current evidence

Schizophrenia Research 141; 2012: 179-184

View review abstract online

Comparison	**Adjunctive estrogen therapy (2 to 4 weeks, varying doses) vs. placebo in female patients.**
Summary of evidence	**Moderate to high quality evidence (small to medium-sized samples, consistent, precise, direct) indicates small to medium-sized improvements in positive and negative symptoms with adjunctive estrogen (particularly oestradiol).**
Mental state **Positive and Negative Syndrome Scale (PANSS) or Brief Psychiatric Rating Scale (BPRS)**	
A significant, small to medium-sized improvement in symptoms with adjunctive estrogen; Total symptoms: 4 RCTs, N = 214, g = 0.66, 95%CI 0.21 to 1.11, p = 0.004, I^2 = 58% Positive symptoms: 4 RCTs, N = 214, g = 0.54, 95%CI 0.27 to 0.82, p = 0.002, I^2 = 0% Negative symptoms: 4 RCTs, N = 214, g = 0.34, 95%CI 0.01 to 0.67, p = 0.04, I^2 = 25% *Oestradiol only:* Total symptoms: 3 RCTs, N = 170, g = 0.79, 95%CI 0.28 to 1.31, p = 0.003, I^2 = 56% Positive symptoms: 3 RCTs, N = 170, g = 0.57, 95%CI 0.26 to 0.87, p < 0.001, I^2 = 0% Negative symptoms: 3 RCTs, N = 170, g = 0.45, 95%CI 0.14 to 0.75, p = 0.004, I^2 = 0%	
Consistency in results‡	Consistent, particularly for positive and negative symptoms.
Precision in results§	Precise
Directness of results‖	Direct

Cakici N, van Beveren N, Judge-Hundal G, Koola M, Sommer I

An update on the efficacy of anti-inflammatory agents for patients with schizophrenia: A meta-analysis

Psychological Medicine 2019; 49: 2307-19

View review abstract online	
Comparison	**Adjunctive estrogen vs. placebo.**
Summary of evidence	**Moderate to high quality evidence (large sample, inconsistent, precise, direct) suggests a medium-sized benefit of adjunctive estrogen for improvement in symptoms.**
Symptoms	
Medium-sized effect of reduced symptom severity with estrogen; 11 RCTs, N = 723, g = 0.57, 95%CI 0.25 to 0.90, p = 0.001, I^2 = 74%	
Consistency in results	Inconsistent
Precision in results	Precise
Directness of results	Direct

Cho M, Lee TY, Kwak YB, Yoon YB, Kim M, Kwon JS

Adjunctive use of anti-inflammatory drugs for schizophrenia: A meta-analytic investigation of randomized controlled trials

Australian and New Zealand Journal of Psychiatry 2019; 53: 742-59

View review abstract online

Comparison	**Adjunctive estrogen medications vs. placebo.**
Summary of evidence	**Moderate to high quality evidence (large samples, inconsistent, precise, direct) suggests medium-sized benefits of adjunctive estrogen medications for improving symptoms.**
Symptoms	
Medium-sized effects of reduced symptom severity with estrogen medications; Estrogen: 7 RCTs, N = 327, g = 0.47, 95%CI 0.13 to 0.81, p < 0.05, I^2 = 66% The result was significant for both positive and negative symptoms. Raloxifene: 9 RCTs, N = 583, g = 0.40, 95%CI 0.10 to 0.70, p < 0.05, I^2 = 74% The result was significant for positive but not negative symptoms.	
Consistency in results	Inconsistent

Precision in results	Precise
Directness of results	Direct

De Boer J, Prikken M, Lei WU, Begemann M, Sommer I

The effect of raloxifene augmentation in men and women with a schizophrenia spectrum disorder: A systematic review and meta-analysis

NPJ Schizophrenia 2018; 4(1) doi:10.1038/s41537-017-0043-3

View review abstract online

Comparison	**Adjunctive raloxifene (6 to 24 weeks, 60-120mg vs. placebo in males or females with schizophrenia.**
Summary of evidence	**Moderate to high quality evidence (medium to large samples, some inconsistency, precise, direct) indicates small to medium-sized improvements in positive, negative, and overall symptoms with adjunctive raloxifene. There were no differences in depression or cognition.**

Mental state

Positive and Negative Syndrome Scale (PANSS)

Significant, small to medium-sized effects of improved symptoms with adjunctive raloxifene;

Positive: 9 RCTs, N = 561, g = 0.32, 95%CI 0.05 to 0.59, p = 0.02, I^2 = 54%, p = 0.03

Negative: 9 RCTs, N = 561, g = 0.40, 95%CI 0.08 to 0.72, p = 0.02, I^2 = 67%, p = 0.002

General: 7 RCTs, N = 526, g = 0.46, 95%CI 0.01 to 0.82, p = 0.01, I^2 = 74%, p = 0.001

Total: 8 RCTs, N = 482, g = 0.57, 95%CI 0.41 to 0.99, p = 0.009, I^2 = 77%, p < 0.001

No significant differences in depression;

Depression: 2 RCTs, N = 135, g = 0.14, 95%CI -0.20 to 0.47, p = 0.43%, I^2 = NR, p = 0.57

Cognition

No significant differences in cognition;

Attention and working memory: 4 RCTs, N = 352, g = -0.01, 95%CI -0.28 to 0.26, p = 0.92, I^2 = 27%, p = 0.25

Executive functioning: 2 RCTs, N = 221, g = 0.03, 95%CI -0.23 to 0.29, p = 0.83, I^2 = NR, p = 0.52

Memory: 4 RCTs, N = 361, g = 0.12, 95%CI -0.11 to 0.35, p = 0.31, I^2 = 13%, p = 0.33

Psychomotor speed: 3 RCTs, N = 303, *g* = 0.28, 95%CI -0.26 to 0.81, *p* = 0.31, I^2 = 75%, *p* = 0.02 Verbal fluency: 4 RCTs, N = 361, *g* = 0.06, 95%CI -0.24 to 0.35, *p* = 0.71, I^2 = 42%, *p* = 0.16 Global cognitive functioning: 2 RCTs, N = 256, *g* = -0.13, 95%CI -0.37 to 0.11, *p* = 0.30, I^2 = NR, *p* = 0.64	
Consistency in results	Inconsistent for PANSS symptoms and psychomotor speed.
Precision in results	Precise
Directness of results	Direct

Sommer IE, van Westrhenen R, Begemann M, de Witte L, Leucht S, Kahn RS

Efficacy of Anti-inflammatory Agents to Improve Symptoms in Patients With Schizophrenia: An Update

Schizophrenia Bulletin 2013: 40(1): 181-191

View review abstract online

Comparison	**Adjunctive estrogen (0.05 to 2mg daily for 3 to 4 months) vs. placebo in males or females with schizophrenia.**
Summary of evidence	**Moderate quality evidence (medium-sized sample, inconsistent, precise, direct) suggests a medium-sized improvement in symptoms with adjunctive estrogen.**
Symptoms **Measured by PANSS, BPRS**	
A medium size effect of reduced symptom severity in patients receiving estrogen; 6 RCTs, N = 260, *g* = 0.51, 95%CI 0.04 to 0.97, *p* = 0.03, I^2 = 62% Heterogeneity decreased and the effect size increased when the only male study was excluded.	
Consistency in results	Inconsistent
Precision in results	Precise
Directness of results	Direct

Wang Q, Dong X, Wang Y, Li X

Raloxifene as an adjunctive treatment for postmenopausal women with schizophrenia: a meta-analysis of randomized controlled trials

Archives of Women's Mental Health 2018; 21: 31-41

View review abstract online

Comparison	**Adjunctive raloxifene (8 to 24 weeks, 60-120mg vs. placebo in postmenopausal females with schizophrenia.**
Summary of evidence	**Moderate to high quality evidence (large samples, some inconsistencies, precise, direct) indicates small to medium-sized improvements in positive, negative, and overall symptoms with adjunctive raloxifene in postmenopausal women with schizophrenia. There were no differences in adverse effects.**
Mental state **Positive and Negative Syndrome Scale (PANSS)**	
Significant, small to medium-sized effects of improved symptoms with adjunctive raloxifene; Positive: 6 RCTs, N = 440, SMD = -0.22, 95%CI -0.41 to -0.02, $p = 0.03$, $I^2 = 79\%$, $p < 0.0001$ Removing two outliers found a medium-sized effect (-0.64) and reduced heterogeneity (41%). Negative: 6 RCTs, N = 440, SMD = -0.46, 95%CI -0.89 to -0.02, p = 0.04, I2 = 74%, p = 0.0007 Removing one outlier found a medium-sized effect (-0.57) and reduced heterogeneity (0%). General: 6 RCTs, N = 440, SMD = -0.55, 95%CI -0.99 to -0.11, $p = 0.01$, $I^2 = 75\%$, $p = 0.0005$ Removing one outlier found a medium-sized effect (-0.69) and reduced heterogeneity (0%). Total: 6 RCTs, N = 440, SMD = -0.55, 95%CI -1.01 to -0.09, $p = 0.02$, $I^2 = 77\%$, $p = 0.0002$ Removing one outlier found a medium-sized effect (-0.70) and reduced heterogeneity (0%).	
Risks	There were no significant differences in adverse effects or all cause discontinuation.
Consistency in results	Inconsistent, apart from sensitivity analyses.
Precision in results	Precise
Directness of results	Direct

Estrogen

Explanation of acronyms

BPRS = Brief Psychiatric Rating Scale, CI = confidence interval, I^2 = the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance), N = number of participants, *p* = statistical probability of obtaining that result ($p < 0.05$ generally regarded as significant), PANSS = positive and negative syndrome scale, SMD = standardised mean difference, vs. = versus

Explanation of technical terms

* Bias has the potential to affect reviews of both RCT and observational studies. Forms of bias include; reporting bias – selective reporting of results; publication bias - trials that are not formally published tend to show less effect than published trials, further if there are statistically significant differences between groups in a trial, these trial results tend to get published before those of trials without significant differences; language bias – only including English language reports; funding bias - source of funding for the primary research with selective reporting of results within primary studies; outcome variable selection bias; database bias - including reports from some databases and not others; citation bias - preferential citation of authors. Trials can also be subject to bias when evaluators are not blind to treatment condition and selection bias of participants if trial samples are small[9].

† Different effect measures are reported by different reviews.

Prevalence refers to how many existing cases there are at a particular point in time. Incidence refers to how many new cases there are per population in a specified time period. Incidence is usually reported as the number of new cases per 100,000 people per year. Alternatively some studies present the number of new cases that have accumulated over several years against a person-years denominator. This denominator is the sum of individual units of time that the persons in the population are at risk of becoming a case. It takes into account the size of the underlying population sample and its age structure over the duration of observation.

Reliability and validity refers to how accurate the instrument is. Sensitivity is the proportion of actual positives that are correctly identified (100% sensitivity = correct identification of all actual positives) and specificity is the proportion of negatives that are correctly identified (100% specificity = not identifying anyone as positive if they are truly not).

Weighted mean difference scores refer to mean differences between treatment and comparison groups after treatment (or occasionally pre to post treatment) and in a randomised trial there is an assumption that both groups are comparable on this measure prior to treatment. Standardised mean differences are divided by the pooled standard deviation (or the standard deviation of one group when groups are homogenous) which allows results from different scales to be combined and compared. Each study's mean difference is then given a weighting depending on the size of the sample and the variability in the data. Less than 0.4 represents a small effect, around 0.5 a medium effect, and over 0.8 represents a large effect[9].

Odds ratio (OR) or relative risk (RR) refers to the probability of a reduction (< 1) or an increase (> 1) in a particular outcome in a treatment group, or a group exposed to a risk factor, relative to the comparison group. For example, a RR of 0.75 translates to a reduction in risk of an outcome of 25% relative to those not receiving the treatment or not exposed to the risk factor. Conversely, a RR of 1.25 translates to an increased risk of 25% relative to those not receiving treatment or not having been exposed to a risk factor. A RR or OR of 1.00 means there is no difference between groups. A medium effect is considered if RR > 2 or < 0.5 and a large effect if RR > 5 or < 0.2[10]. lnOR stands for logarithmic OR where a lnOR of 0 shows no difference between groups. Hazard ratios measure the effect of an explanatory variable on the hazard or risk of an event.

Margarete Ainsworth Building, Barker Street, Randwick NSW 2031. Phone: 02 9399 1000. Email: info@neura.edu.au

Correlation coefficients (eg, r) indicate the strength of association or relationship between variables. They can provide an indirect indication of prediction, but do not confirm causality due to possible and often unforseen confounding variables. An r of 0.10 represents a weak association, 0.25 a medium association and 0.40 and over represents a strong association. Unstandardised (*b*) regression coefficients indicate the average change in the dependent variable associated with a 1 unit change in the independent variable, statistically controlling for the other independent variables. Standardised regression coefficients represent the change being in units of standard deviations to allow comparison across different scales.

‡ Inconsistency refers to differing estimates of effect across studies (i.e. heterogeneity or variability in results) that is not explained by subgroup analyses and therefore reduces confidence in the effect estimate. I^2 is the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance) - 0% to 40%: heterogeneity might not be important, 30% to 60%: may represent moderate heterogeneity, 50% to 90%: may represent considerable heterogeneity and over this is considerable heterogeneity. I^2 can be calculated from Q (chi-square) for the test of heterogeneity with the following formula[9];

$$I^2 = \left(\frac{Q - df}{Q}\right) \times 100\%$$

§ Imprecision refers to wide confidence intervals indicating a lack of confidence in the effect estimate. Based on GRADE recommendations, a result for continuous data (standardised mean differences, not weighted mean differences) is considered imprecise if the upper or lower confidence limit crosses an effect size of 0.5 in either direction, and for binary and correlation data, an effect size of 0.25. GRADE also recommends downgrading the evidence when sample size is smaller than 300 (for binary data) and 400 (for continuous data), although for some topics, these criteria should be relaxed[11].

‖ Indirectness of comparison occurs when a comparison of intervention A versus B is not available but A was compared with C and B was compared with C, which allows indirect comparisons of the magnitude of effect of A versus B. Indirectness of population, comparator and/or outcome can also occur when the available evidence regarding a particular population, intervention, comparator, or outcome is not available and is therefore inferred from available evidence. These inferred treatment effect sizes are of lower quality than those gained from head-to-head comparisons of A and B.

Margarete Ainsworth Building, Barker Street, Randwick NSW 2031. Phone: 02 9399 1000. Email: info@neura.edu.au

References

1. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMAGroup (2009): Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *British Medical Journal* 151: 264-9.
2. GRADEWorkingGroup (2004): Grading quality of evidence and strength of recommendations. *British Medical Journal* 328: 1490.
3. Begemann MJH, Dekker CF, van Lunenburg M, Sommer IE (2012): Estrogen augmentation in schizophrenia: A quantitative review of current evidence. *Schizophrenia Research* 141: 179-84.
4. Sommer IE, van Westrhenen R, Begemann MJH, de Witte LD, Leucht S, Kahn RS (2014): Efficacy of Anti-inflammatory Agents to Improve Symptoms in Patients With Schizophrenia: An Update. *Schizophrenia Bulletin* 40: 181-91.
5. De Boer J, Prikken M, Lei WU, Begemann M, Sommer I (2018): The effect of raloxifene augmentation in men and women with a schizophrenia spectrum disorder: A systematic review and meta-analysis. *NPJ Schizophrenia* 4 (1): doi:10.1038/s41537-017-0043-3.
6. Wang Q, Dong X, Wang Y, Li X (2018): Raloxifene as an adjunctive treatment for postmenopausal women with schizophrenia: a meta-analysis of randomized controlled trials. *Archives of Women's Mental Health* 21: 31-41.
7. Cakici N, van Beveren N, Judge-Hundal G, Koola M, Sommer I (2019): An update on the efficacy of anti-inflammatory agents for patients with schizophrenia: A meta-analysis. *Psychological Medicine* 49: 2307-19.
8. Cho M, Lee TY, Kwak YB, Yoon YB, Kim M, Kwon JS (2019): Adjunctive use of anti-inflammatory drugs for schizophrenia: A meta-analytic investigation of randomized controlled trials. *Australian and New Zealand Journal of Psychiatry* 53: 742-59.
9. CochraneCollaboration (2008): Cochrane Handbook for Systematic Reviews of Interventions. Accessed 24/06/2011.
10. Rosenthal JA (1996): Qualitative Descriptors of Strength of Association and Effect Size. *Journal of Social Service Research* 21: 37-59.
11. GRADEpro (2008): [Computer program]. Jan Brozek, Andrew Oxman, Holger Schünemann. *Version 32 for Windows*