Ethical Human Psychology and Psychiatry, *Volume 21*, *Number 2*, *2019*

Electroconvulsive Therapy for Depression: A Review of the Quality of ECT versus Sham ECT Trials and Meta-Analyses

John Read, PhD

University of East London, London, UK

Irving Kirsch, PhD

Harvard Medical School, Boston, MA

Laura McGrath, PhD

University of East London, London, UK

Background: Electroconvulsive therapy (ECT) is still being administered to approximately a million people annually. There have been no ECT versus simulated ECT (SECT) studies since 1985. The five meta-analyses of ECT versus SECT studies all claim that ECT is more effective than SECT for its primary target, severe depression. This review assesses the quality of those meta-analyses and of the 11 studies on which they are based. **Methods:** The meta-analyses were evaluated primarily in terms of whether they considered the quality of the studies they included, but also in terms of whether they addressed efficacy beyond end of treatment. The methodological rigor of the 11 studies included by one or more of the meta-analyses was assessed using a 24-point Quality scale developed for this review. **Results:** The five meta-analyses include between 1 and 7 of the 11 studies. The meta-analyses pay little or no attention to the multiple limitations of the studies they include. The 11 studies have a mean Quality score of 12.3 out of 24. Eight scored 13 or less. Only four studies describe their processes of randomization and testing the blinding. None convincingly demonstrate that they are double-blind. Five selectively report their findings. Only four report any ratings by patients. None assess Quality of Life. The studies are small, involving an average of 37 people. Four of the 11 found ECT significantly superior to SECT at the end of treatment, five found no significant difference and two found mixed results (including one where the psychiatrists reported a difference but patients did not). Only two higher Quality studies report follow-up data, one produced a near-zero effect size (.065) in the direction of ECT, and the other a small effect size (.299) in favor of SECT. **Conclusions:** The quality of most SECT–ECT studies is so poor that the meta-analyses were wrong to conclude anything about efficacy, either during or beyond the treatment period. There is no evidence that ECT is effective for its target demographic—older women, or its target diagnostic group—severely depressed people, or for suicidal people, people who have unsuccessfully tried other treatments first, involuntary patients, or children and adolescents. Given the high risk of permanent memory loss and the small mortality risk, this longstanding failure to determine whether or not ECT works means that its use should be immediately suspended until a series of well designed, randomized, placebo-controlled studies have investigated whether there really are any significant benefits against which the proven significant risks can be weighed.

http://dx.doi.org/10.1891/EHPP-D-19-00014

Keywords: electroconvulsive therapy; placebo; efficacy; meta-analyses; review; methodology

Electroconvulsive therapy (ECT) is still used on approximately a million people annually (Leiknes, Jarosh-von Schweder, & Hoie, 2012; Read, Bentall, Johnstone, Fosse, & Bracken, 2013). A review of 70 studies found, however, "large variation between continent, countries and regions in utilization, rates and clinical practice" (Leiknes et al., 2012, p. 296). For instance, a recent audit found a 12-fold difference in usage between the highest and lowest using regions of England (Read, Harrop, Geekie, & Renton, 2018).

The many recent studies that either compare ECT to other treatments, or compare different types of ECT with each other (Read & Arnold, 2017), typically open with an unqualified statement that ECT is a very effective treatment for depression. Some may consider these types of studies sufficient to justify the use of ECT. We contend, however, that, ECT must be assessed using the same standards applied to psychiatric medications and other medical interventions, with placebo-controlled studies as the primary method for assessment. There have, however, only ever been 11 placebo-controlled studies of the efficacy of ECT. The last study comparing ECT with sham/simulated ECT (SECT), in which the general anaesthetic is administered but the electricity is not, was 35 years ago (Gregory, Shawcross, & Gill, 1985). This review evaluates, for the first time, the impartiality and robustness of the meta-analyses of this small body of literature, and the quality of the studies cited in the meta-analyses. The primary goal is not to assess whether or not ECT is effective. The intent, instead, is to determine whether the available evidence is robust enough to answer that question.

METHOD

A Medline (MESH) search for meta-analyses on the effectiveness of ECT for depression using placebo-controlled trials (ECT vs SECT), was conducted in June 2019, using the following index terms: ["ECT" OR "electroshock therapy" OR "electroconvulsive treatment" OR "electroshock treatment"] AND ["meta-analysis"] AND ["depression" OR "major depressive disorder"].

A 24-point Quality scale was developed to assess the studies cited by the meta-analyses. The scale combined the "risk of bias" domains of the *Cochrane Handbook Risk of Bias Tool* (randomization, blinding, incomplete outcome data and selective reporting; Higgins et al., 2011) with other criteria relating to quality of design and reporting, and some criteria specific to ECT research (see Table 1 for criteria and their definitions). No differential weightings were given to individual items, but the three key issues of randomization, blinding and diagnosis carried extra weight by virtue of having two or three items each. The 11 studies were independently rated, using the definitions in Table 1, by JR and LM, with each rater blind to the other's ratings. "Yes" indicated clear affirmative evidence. "No" meant either no evidence or clear negative evidence. Inconsistencies between raters were resolved by discussion and rereading the articles together. Spearman rank correlations and two-tailed *t* tests were used to assess the relationships between Quality scores and other variables.

TABLE 1. Definitions of the 24 Quality Criteria

RANDOMIZED[a]	Any statement or evidence that the study was randomized, and no evidence that this was not the case
Process described	Any description of the randomization process
BLINDED[a]	Any statement or evidence that the study was blinded, and no evidence that the blind was broken—for raters or patients
Method tested	Any evidence that the blinding of either the raters or patients was tested
No previous ECT	None of the participants had had ECT at any time prior to the study
ALL DEPRESSED	All participants (or a clear subset with separate data) were adjudged, by any method, to be depressed (with or without other features, e.g., psychosis)
Reliable diagnosis	Diagnosis made by two or more independent people, or any standardized depression assessment tool, i.e., not just by one clinician/clinical diagnosis with unspecified diagnoser(s)
Severe	All participants *severely* depressed at outset of study, either any meaningful description of "severe," or ≤ 22 on Hamilton (44 if two raters, most studies), ≤ 29 on Beck scale)
FULL ECT COURSE	At least six ECTs or 6 SECTs; so excluding studies giving ECT to SECT group before six ECT treatments
SUICIDE MEASURE	Any outcome measure of suicide or suicidality (ideation)
VALIDATED DEPRESSION SCALE	e.g., Hamilton, Montgomery, Beck
Means and *SD*s	Means *and SD*s (or SEs or SEMS) reported for the depression scales pre and post treatment (or just the means and SDs of the *change* between pre and post)
NO SELECTIVE REPORTING[a]	Outcomes for all measures and all types of raters (e.g., doctors, patients etc.) reported
INDIVIDUAL PATIENTS' DATA	Any ratings/scores/categorization for individual participants reported
PATIENT RATINGS	Any self-report or patient ratings administered *and* scores reported
QUALITY OF LIFE MEASURE	Any "Quality of Life" ratings administered *and* scores reported (e.g., HONOS)
MORE THAN 1 RATER TYPE	More than one type/group of persons making separate ratings; e.g., psychiatrists, nurses, patients, etc.
DECLINERS DESCRIBED[a]	Any description of people who were approached but declined to participate
WITHDRAWALS DESCRIBED[a]	Any description of people who withdrew (or were withdrawn) from the study after it had started
OTHER TREATMENTS UNSUCCESSFUL	One or more other treatments (antidepressants, CBT etc.) had been tried and did not work prior to ECT
MEDS MATCHED/ CONTROLLED/ STOPPED	Psychiatric meds (e.g., antidepressants) were stopped for the study, or that the two groups (ECT and SECT) were matched or controlled in any way re. psychiatric meds

(Continued)

TABLE 1. Definitions of the 24 Quality Criteria (Continued)

BOTH ECT and SECT SAMPLES ≤ 10	Both sample sizes (ECT and SECT) 10 or larger
AGE and GENDER REPRESENTATIVE	More than 50% female (but not all), and mean age of 50 or more
FOLLOW-UP DATA	Any outcome data gathered beyond end of treatment (more than 1 day after last ECT), without ECT being given to the SECT group

Note. ECT = electroconvulsive therapy; SECT = sham/simulated electroconvulsive therapy.
[a]Relates to one of the four Cochrane 'risk of bias' domains; either directly or, for DECLINERS DESCRIBED and WITHDRAWALS DESCRIBED, relates indirectly to the 'incomplete outcome data' domain.

RESULTS

The search for meta-analyses produced 83 papers (see Figure 1). When the 83 papers were limited to ["SECT" OR "sham ECT"] etc., 14 remained. Three of these were literature reviews (Greenhalgh, Knight, Hind, Beverley, & Walters, 2005; Read & Bentall, 2010; Ross, 2006), one was a meta-analysis in Hungarian (Gábor & László, 2005), one was a meta-analysis of ECT versus SECT for older people only, discussed later (van der Wurff, Stek, Hooogendijk, & Beekman, 2003), and three were about transcranial magnetic stimulation. This left five meta-analyses for review (Janicak et al., 1985; Kho, van Vreewijk, Simpson, & Zwinderman, 2003; Mutz et al., 2019; Pagnin, de Queiroz, Pini, & Cassano, 2004; UK ECT Review Group, 2003). A follow-up search in March 2020 found no further meta-analyses or sham ECT studies.

Independent Quality Ratings

The mean Quality scores of the two raters, for the 11 studies, 10.27 (*SD* 2.45) and 11.91 (*SD* 2.91), were not significantly different ($T(20) = 1.42$, $p = .17$). Their scores for the 11 studies were significantly correlated (rho = .87, $p = .001$). There were 55 inconsistencies out of the 264 ratings, representing an agreement rate of 79.2%. This translates to a *kappa* score (which allows for agreement by chance) of .58, in the "fair to good" range (.40–.75; Fleiss, 1981). The inconsistencies were resolved by discussion. The majority had resulted from raters missing (or misunderstanding) some text; for example, missing methodological information mentioned in a Results section, or missing results in a Discussion section. During this rereading of studies together some instances where both raters had missed some quality evidence were also discovered, and scores increased accordingly.

If ambiguity remained after discussion the raters erred on the side of "Yes." For example, one rater rated Lambourn and Gill "No" for "Means and *SD*s," whereas the other rated it "Yes" because the means were provided and the *SD*s, although not reported, could be calculated from individuals' data. This was finalized as "Yes." Brandon et al. (1984) reported means and *SD*s but only in the form of a graph, with no numbers, leading one rater to rate it as a "No." After discussion, a "Yes" was agreed. One rater had scored Ulett, Smith, and Gleser (1956) as "No" for "Reliable diagnosis" because it was not explicitly stated that

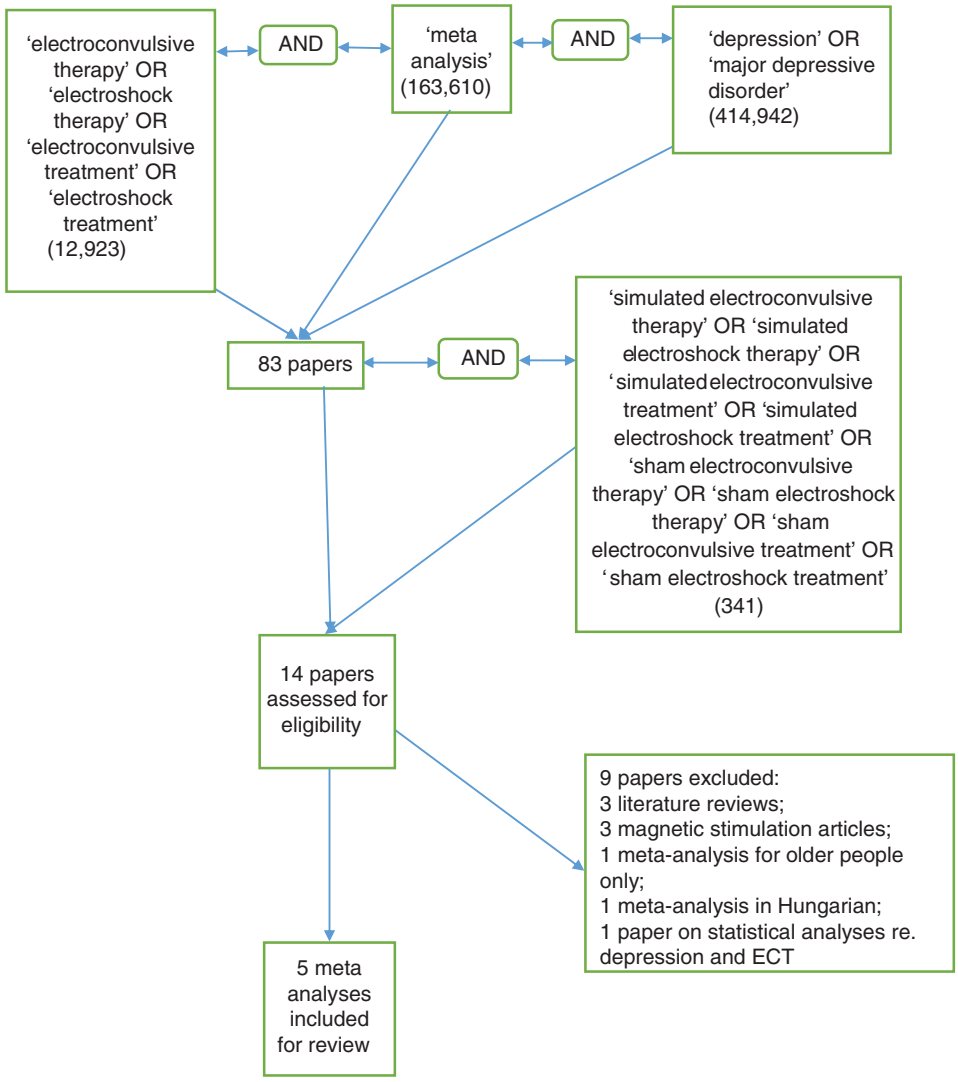

Figure 1. Flowchart of search strategy for meta-analyses.

diagnoses made in the study were independent; but a "Yes" was agreed on as there were two people diagnosing participants.

The mean of the final, agreed, scores was 12.27 (*SD* 3.20), somewhat higher than the original means of the raters.

The 11 SECT versus ECT Studies Included in the Five Meta-Analyses: Findings at the End of Treatment

The 11 ECT versus SECT studies for depression cited by one or more of the five meta-analyses, summarized in Table 2, are the only 11 ever conducted. None since 1985 have

been identified by reviews (Read & Arnold, 2017; Read & Bentall, 2010) or the recent meta-analysis (Mutz et al., 2019). The first five were published between 1956 and 1963; with a second wave, of six, between 1978 and 1985. Three took place in the USA and the other eight in the UK, including all six of the later wave. So there have been no such studies in the UK for 35 years, none in the USA for 57 years, and none anywhere else ever.

Ulett et al. (1956)—Quality Score 10/24. The first SECT versus ECT study, conducted in the USA, compared both ECT and "convulsive photoshock" (using flashing lights) to a sham treatment involving the same "light stage of sleep" as the two treatment groups. There was no significant difference between the ECT and SECT groups on the psychiatrist's ratings, with 33% and 24%, respectively, showing "recovery or marked improvement."

This study, however, does not belong in an evaluation of ECT for depression. The participants were "individuals with the types of mental illness which are thought to respond best to the shock therapies," in 1956. So 24 of the 42 (62%) in the ECT and SECT groups had diagnoses of "schizophrenic reaction" or "involutional psychotic reaction." The study also had no depression outcome measure. Despite this, and numerous other failings (see Table 3) two meta-analyses (Janicak et al., 1985; Pagnin et al., 2004) include this study. Pagnin et al. correctly report the difference between ECT and SECT as nonsignificant. The Janicak meta-analysis, however, wrongly report a significant difference in favor of shock therapy, by inappropriately merging the photoshock and ECT data.

Brill et al. (1959)—9/24. The second study, also in the USA, did not assess outcome until a month after the treatment period so it really belongs as much with the follow-up studies (see below) as with the short-term/end of treatment studies. The study was included in the same two meta-analyses as the Ulett study. It involved 97 men with an average age of 35, so was unrepresentative of the modal ECT recipient—a woman in her 60s (Leiknes et al., 2012; Read et al., 2013, 2018). Only 30 were diagnosed with depression, but fortunately their data were reported separately. A positive outcome was deemed to be "recovery" on two-out-of-three tools: psychiatric evaluation, the Lorr Psychiatric Rating Scale (Lorr, Jenkins, & Holsopple, 1953), and psychological testing. None of the three explicitly assessed depression.

"Nearly half" of the participants had had ECT before, which may have contributed to the fact that "some patients in the nonshock group believed that they were receiving some new variation of ECT" (p. 628). This raises the possibility that some could tell that they may not have had real ECT, because of the absence of headaches and confusion immediately afterwards.

Sixteen of the 21 men in the ECT group (76%) and 4 of the 9 in the SECT group (44%) met the two-out-of-three criterion for recovery. The difference is not statistically significant.

Harris and Robin (1960)—9/24. The first UK study was a trial of the antidepressant phenelzine, but included four women receiving ECT and four receiving SECT (all without phenelzine). The study invalidated any findings on ECT, however, by giving ECT to the SECT group after four ECTs (2 weeks). Despite this, and multiple other flaws (see Table 3), this study was included in two meta-analyses (Janicak et al., 1985; Pagnin et al., 2004). At the 2 week point two of the four ECT recipients and none of the SECT group had shown "great improvement." This difference was not statistically significant.

TABLE 2. Summaries of the 11 SECT vs ECT Studies, Outcomes at End of Treatment

Study and Demographics	*n* ECT vs SECT	ECT Type Frequency % Previous ECT	Outcome Measure(s)	Results	Significant Difference	Two Most Serious Weaknesses
Ulett et al. (1956) 62% fem 17–66 years mean 46 years	21 vs 21	Sine wave Bilateral Three per week 12–15 ECTs 29%	Psychiatrist Ratings of general improvement (1–5) Malamud Scale (psychosis)	33% vs 24% "recovery or marked improvement"	NO	62% not depressed No depression measure
Brill et al. (1959) 100% male 18–68 years mean 35 years	21 vs 9	Sine wave Frontotemporal Three per week 20 ECTs "nearly half"	Three rating scales of general improvement (1 month after treatment ended)"improved" = recovery on two of the three scales	76% vs 44% "improved"	NO	No depression measure All males, mostly middle-aged
Harris and Robin (1960) 100% fem mean 62 years	4 vs 4	? ? Two per week Eight ECTs ?%	General improvement scale (1–4)	2/4 vs 0/4 "great improvement"	NO	Gave ECT to SECT group after 2 weeks (four ECTs)—invalidating end of treatment data No depression measure

(Continued)

TABLE 2. Summaries of the 11 SECT vs ECT Studies, Outcomes at End of Treatment (Continued)

Study and Demographics	*n* ECT vs SECT	ECT Type Frequency % Previous ECT	Outcome Measure(s)	Results	Significant Difference	Two Most Serious Weaknesses
Fahy, Imlah, and Harrington (1963) 55% fem 30–59 years	17 vs 17	? ? Two per week Six ECTs ?	Doctors' rating scale (−1 to 3) Staff rating scale (−1 to 3)	35% vs 12% "recovered /minimal symptoms" %s not reported	NO NO[a]	No proper SECT group—no suggestion to control group that they might be having ECT Severe depression excluded
Wilson, Vernon, Guin, and Sandifer (1963) 100% fem 40–59 years	6 vs 6	? ? Two per week Six ECTs ?%	HAMILTON MMPI-Dep (self-rated)	23 vs 10 points improvement 29 vs 8 points improvement	YES YES	One of the raters not blind No older people, or men
Freeman, Basson, and Crighton (1978) 72% fem mean 52 years (range 20–70)	14 vs 18	Sine wave Bilateral Two per week 6.7 ECTs (mean) 56%	HAMILTON WAKEFIELD (self rating) Visual analogue scale BECK (self rating)	(no means or SDs—just graphs)	YES[b] YES[b] YES[b] NO[b]	Gave ECT to SECT group after 1 week (2 ECTs)—invalidating end of treatment data 4/20 in ECT group (0/20 SECT) withdrew without improvement, not included in calculating means

(Continued)

TABLE 2. Summaries of the 11 SECT vs ECT Studies, Outcomes at End of Treatment (Continued)

Study and Demographics	*n* ECT vs SECT	ECT Type Frequency % Previous ECT	Outcome Measure(s)	Results	Significant Difference	Two Most Serious Weaknesses
Lambourn and Gill (1978) 56% fem	16 vs 16	Brief pulse Unilateral	HAMILTON	26 vs 23 points improvement[c]	NO	66% had had ECT before Clinical diagnosis only
36–69 years Mean 54 (11 ≤ 60, 34%)		Three per week Six ECTs 66%	Doctor's Global assessment (0–3)	6/16 vs 6/16[d]	NO	
Johnstone et al. (1980) 74% fem	31 vs 31	Sine wave Bifrontal	HAMILTON (psychiatrist)	38 vs 28 points improvement	YES[e]	Blindness of raters not assessed No means or *SD*s reported[e]
30–69 years mean 49 years		two per week Eight ECTs 21%	LEEDS (self rating)	not reported data	NO	
			Nurses' Rating	not reported	NO	
West (1981) 59% male	11 vs 11	Sinewave Bilateral	Psychiatrists rating (0–100)	48 vs 7 points improvement	YES	Blindness of raters not assessed
mean 53 years		Two per week Six ECTs	BECK (self-rating)	16 vs 2 points improvement	YES	Not severe depression
		?%	Nurses rating (1–9)	5 vs 1 point improvement	YES	

(Continued)

TABLE 2. Summaries of the 11 SECT vs ECT Studies, Outcomes at End of Treatment (Continued)

Study and Demographics	*n* ECT vs SECT	ECT Type Frequency % Previous ECT	Outcome Measure(s)	Results	Significant Difference	Two Most Serious Weaknesses
Brandon et al. (1984) 64% fem mean 54 years	43 vs 34	Sine wave Bilateral Two per week Eight ECTs (17% given less) 60%	HAMILTON Psychiatrists' ratings (7 point scale of change) "self-ratings"	28 vs 12 points improvement 46 vs 25[g] not reported	YES[e] YES ?	No means or SDs reported[f] 60% had had ECT before
Gregory et al. (1985) gender ? 35/60 60–64 years all under 65	19 uni- lateral 21 bilat- eral 20 SECT	Sine wave Unilateral and Bilateral Two per week approx. Eight ECTs ?%	HAMILTON MADRAS "global assessment of depression"	31 (uni), 28 (bi) vs 14 (SECT) points improvement 24 (uni), 25 (bi) vs 9 (SECT) points improvement Not reported	YES YES	Contradictory reporting of "withdrawers" (36%) Selective reporting of results

Note. ECT = electroconvulsive therapy; fem = female; SECT = sham/simulated electroconvulsive therapy.
[a]graphs show difference is even smaller for staff than for doctors. [b]after 1 week (2 ECTs). [c]Individual scores for each person given so means could be calculated. [d]rated 3 on a 0–3 scale. [e]Significant only for "deluded" and "retarded" subgroups. [f]Means and SDs reported by Buchan et al. (1992). [g]presumably for multiple raters on the 1–7 scale—not stated.

TABLE 3. Quality Ratings of the 11 Studies, on 24 Criteria

	Ulett et al. (1956)	Brill et al. (1959)	Harris and Robin (1960)	Fahy et al. (1963)	Wilson et al. (1963)	Freeman et al. (1978)	Lam-bourn and Gill (1978)	John-stone et al. (1980)	West (1981)	Brandon et al. (1984)	Gregory et al. (1985)	**x/11**
RANDOMIZED	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	11
Process described	✓	✓	✓	✖	✖	✖	✓	✓	✖	✓	✖	6
BLINDED	✓	✖[a]	✓	✖[b]	✖[c]	✓	✓	✓	✓	✓	✓	8
Method tested	✖	✓[d]	✖	✓	✖	✓	✓	✓	✖	✓	✖	6
No previous ECT	✖	✖	✖[e]	✖[e]	✖[e]	✖	✖	✖	✖[e]	✖	✖[e]	0
ALL DEPRESSED	✖	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	10
Severe	✖	✖	✓	✖[f]	✓	✓	✓	✓	✖[g]	✓	✓	7
Reliable diagnosis	✓	✖	✓	✖	✖	✓	✖	✓	✓	✓	✓	7
SUICIDE MEASURE	✖	✖	✖	✖	✖	✖	✓	✖	✖	✖	✖	1
FULL ECT COURSE	✓	✓	✖[h]	✓	✓	✖[i]	✓	✓	✓	✓	✓	9
VALIDATED DEPRESSION SCALE	✖	✖	✖	✖	✓	✓	✓	✓	✓	✓	✓	7
Means and *SD*s	✓	✖	✖	✖	✓	✖	✓	✓	✓	✓[j]	✖[k]	6

(Continued)

TABLE 3. Quality Ratings of the 11 Studies, on 24 Criteria (Continued)

	Ulett et al. (1956)	Brill et al. (1959)	Harris and Robin (1960)	Fahy et al. (1963)	Wilson et al. (1963)	Freeman et al. (1978)	Lam-bourn and Gill (1978)	John-stone et al. (1980)	West (1981)	Brandon et al. (1984)	Gregory et al. (1985)	**x/11**
NO SELECTIVE REPORTING	✓	✓	✓	✖	✓	✓	✖	✖	✓	✖	✖	6
INDIVIDUAL PATIENTS' DATA	✓	✓	✓	✓	✓	✖	✓	✖	✖	✖	✖	6
PATIENT RATINGS	✖	✖	✖	✖	✓	✓	✖	✓	✓	✖[1]	✖	5
QUALITY OF LIFE MEASURE	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	0
MORE THAN 1 RATER TYPE	✖	✖	✖	✓	✓	✓	✓	✓	✓	✓	✖	7
DECLINERS DESCRIBED	✖	✖	✖	✖	✖	✖	✖	✓	✖	✓	✓	3
WITHDRAWALS DESCRIBED	✖	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	10
OTHER TREAT-MENTS UNSUC-CESSFUL	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	0

(Continued)

TABLE 3. Quality Ratings of the 11 Studies, on 24 Criteria (Continued)

	Ulett et al. (1956)	Brill et al. (1959)	Harris and Robin (1960)	Fahy et al. (1963)	Wilson et al. (1963)	Freeman et al. (1978)	Lam-bourn and Gill (1978)	John-stone et al. (1980)	West (1981)	Brandon et al. (1984)	Gregory et al. (1985)	**x/11**
MEDS MATCHED/ CON-TROLLED/ STOPPED	✓	✖	✖	✓	✓	✖	✓	✓	✓	✓	✖	7
BOTH ECT and SECT SAMPLES ≤ 10	✓	✖	✖	✓	✖	✓	✓	✓	✓	✓	✓	8
AGE and GENDER REPRESEN-TATIVE[e]	✖	✖	✖	✖	✖	✓	✓	✖	✖	✓	✖[m]	3
FOLLOW-UP DATA	✖[n]	✓	✖	✖	✖	✖	✓	✓	✖[o,p]	✖[p]	✖[p]	3
SCORE OUT OF 24	10	9	9	9	12	13	17	17	13	16	10	

Note. ECT = electroconvulsive therapy; SECT = sham/simulated electroconvulsive therapy.
[a]"some patients in the nonshock group believed that they were receiving some new variation of ECT." [b]patients and observers knew which treatment had been administered. [c]one of the raters not blind. [d]patients "blindness" tested; raters not. [e]percentage having had ECT prior to the study not reported. [f]"Moderate severity." "Severe depression with high suicidal risk were not included." [g]average baseline Beck scores in "moderate" depression range. [h]gave ECT to SECT group after 4 ECTs. [i]9 gave ECT to SECT group after 2 ECTs. [j]means and SDs represented in figure and published by Buchan et al. (1992). [k]means but no SDs. [l]self-rating scale administered but not reported. [m]gender and age not recorded. [n]not valid because sample mostly not depressed. [o]data gathered but only 5 days after last treatment. [p]most SECT patients had ECT during follow-up period, so no longer a SECT group.

Fahy et al. (1963)—9/24. The second UK study was not a SECT study at all. It compared ECT to sleep induced by general anaesthetic, but: "No attempt was made to suggest to these patients that they were receiving ECT. As far as they knew, the sleep injection was a complete treatment in its own right" (p. 311). Despite this and numerous other flaws (see Table 3) this study was, again, included in the Janicak and Pagnin meta-analyses. Neither mentioned the absence of a SECT group when including the study in their effect-size calculations. The difference, in terms of percentage "recovered or minimal symptoms only" between ECT (35%) and SECT (12%), assessed by a doctor, was not statistically significant. Percentages were not reported for the staff's ratings (thereby meeting the Cochrane "risk of bias" criterion of "selective reporting"), but graphs show that the difference was even smaller than for the doctors. Both meta-analyses use the larger of the two differences in their calculations.

Wilson et al. (1963)—12/24. This small USA project involved 12 ECT patients and 12 SECT patients, with half of each group on an antidepressant. The only meta-analysis to include this study (UK ECT Review Group, 2003) correctly reports only the data for the two groups of six not taking the antidepressant. On both the Hamilton (Hamilton, 1960) and the MMPI-Depression (Schiele, Baker, & Hathaway, 1943) scales the ECT group showed significantly more improvement than the SECT group. The meta-analysis fails to report that one of the two raters before treatment, and one of the three at the end of treatment, knew which patients had received which treatment, so the study was un-blinded. The ratings were not statistically different from each other, and were based on "the same interview" so it is quite possible that the blind raters were influenced by the nonblind rater. Multiple other failings are listed in Table 3, including the exclusion of people aged 60 or older, who are typical ECT patients.

Freeman et al. (1978)—13/24. The first of the second wave of studies (1978–1985) occurred in Scotland. The only meta-analysis to include it was the one by the UK ECT Review Group. Like Harris and Robin (1960), this study invalidated any evaluation of the efficacy of a full course of ECT treatment by giving ECT to the SECT group before the end of the study (after just two ECTs). These two studies evaluate speed of response early in treatment but not efficacy of the whole treatment. After the two ECTs three clinician-rated scales recorded significant differences between the two groups, but there was no difference when the patients rated their own depression. The researchers (Freeman et al., 1978, p. 738) explained:

> The ideal design for such a trial would have been to have compared a full course of S.E.C.T. with a full course of real E.C.T.. . . We felt it ethically unjustified to withhold for a complete course a treatment generally regarded to be effective and to submit patients to perhaps unnecessary general anaesthesia. The method presented here was therefore a compromise.

Four of the 18 ECT patients, but none of the SECT patients, withdrew because they were "nonresponders," but they were not included when calculating means.

This was the only study to report whether participants had been tried on antidepressants prior to the study; 22 (54%) had not.

Lambourn and Gill (1978)—17/24. This study was one of the two highest scorers for Quality. It provided individual Hamilton scores, plus doctors' ratings, for all 32 participants, who had been randomized to the ECT and SECT groups, matching for age and gender. The blindness of the raters was assessed and confirmed. The participants were representative of the age and gender mix of ECT recipients. Most (66%), however, had had ECT before, thereby increasing the probability of un-blinding for those patients.

The study differed from most studies by using unilateral, rather than bilateral, electrode placement. It also differed by studying people diagnosed with "depressive psychosis," although they were severely depressed. The following can be calculated from the individual scores. There was no significant difference in the mean reduction on the Hamilton scale (using the old scoring system in which the ratings of two raters are added together) between the ECT (26.2) and SECT (22.8) groups (t (30) = .50, p = .62). On the doctors' ratings 37.5% of both groups were rated 3 on an undefined 0–3 scale, and 69% of the ECT group vs. 62.5% of the SECT group were rated 2 or 3, a nonsignificant difference (χ^2 = .14, p = .71).

This study was included in all meta-analyses except the recent one (Mutz et al., 2019). Table 4 shows that four different effect sizes were calculated by the four meta-analyses, ranging from .17 (UK ECT Group, 2003) to 0 (Pagnin et al., 2004; Odds Ratio = 1.0). None of them reach the threshold of even a "small" effect size (.2; Hamilton, 1960).

Johnstone et al. (1980)—17/24. The famous Northwick Park study was one of the largest studies, and is the other of the two highest scorers on the Quality scale. Neither the ratings by the nurses nor the self-ratings by the patients produced significant differences between the 31 ECT patients and the 31 SECT patients. There was, however, a significant difference on change in Hamilton scores rated by a psychiatrist. The reporting of the findings is problematic. There were no data or SDs reported for the two outcomes that found no significant difference between ECT and SECT (by nurses and patients), making them harder to include in meta-analyses. There was just one rather basic graph, for the psychiatrist's Hamilton ratings.

Furthermore, despite including three subtypes of depression Johnstone et al. (1980) failed to report separate findings for them. Re-analysis by Buchan et al. (1992) suggests that the difference between ECT and SECT on the Hamilton is only significant for the patients who were deluded as well as depressed (although it is hard to be sure because Buchan et al. (1992) merge the data for the sub groups with data from the Brandon et al. (1984) study).

Only one meta-analysis (UK ECT Group, 2003) includes this relatively rigorously conducted, but poorly reported, study.

West (1981)—13/24. This small study was reported in just two pages, by a sole author. The 11 who received ECT were reported to have improved significantly more than the 11 receiving SECT, on separate ratings by psychiatrists, nurses, and patients. West concluded his findings were "very strong evidence" and that ECT is "an excellent treatment of severe depression." The differences were much larger than in any other studies. Unlike the other studies, there was virtually no change in the SECT group.

The nurses' scale raises concerns about the integrity of the study. The scale was described as a nine point scale from "very much worse" to "very much better," but scores were reported at baseline, before any treatment had taken place. One cannot be "worse" or "better" before a study begins.

TABLE 4. Summaries of the Five Meta-analyses of SECT versus ECT Studies

META-ANALYSES	Janicak et al. (1985)	Kho et al. (2003)	UK ECT Review Group (2003)	Pagnin et al. (2004)	Mutz et al. (2019)
Study	**6 studies**	**2 studies**	**6 studies**	**7 studies**	**1 study**
	n = 205	*n* = 109	*n* = 226[a]	*n* = 245	*n* = 77
Ulett et al. (1956)	χ^2 = 6.36*,[b]			*OR* = .57	
Brill et al. (1959)	χ^2 = 2.37			*OR* = 3.82	
Harris and Robin (1960)	χ^2 = .67			*OR* = 17.0	
Fahy et al. (1963)	χ^2 = 1.09			*OR* = 3.76	
Wilson et al. (1963)			ES = 1.08		
Freeman et al. (1978)			ES = .63		
Lambourn and Gill (1978)	χ^2 = .12	ES = .09	ES = .17	*OR* = 1.00	
Johnstone et al. (1980)			ES = .74*		
West (1981)	χ^2 = 14.85*		ES = 1.25*	*OR* = 86.1*	
Brandon et al. (1984)		ESs = 1.38 – 1.99*		*OR* = 2.16	No data
Gregory et al. (1985)			SES = 1.42*		
Overall finding of meta-analyses	72% v 40% χ^2 = 21.54 $p < .001$	Pooled Effect Size = .95 [95% CI .35–1.54]	Pooled Effect Size = .91 [95% CI · 54–1.27][c]	*OR* = 2.83 [CI 95% 1.30–6.17] χ^2 = 6.87 p = .009	*OR*s Bilateral[e] = 8.91* High-dose unilateral[f] = 7.27* Low dose unilateral[f] = 2.74 Bifrontal[f] = 3.39
	2/6 studies significant[d]	1/2 significant	3/6 significant	1/7 significant	2/4 types significant

Note. Empty cells indicate study excluded by meta-analysis. ECT = electroconvulsive therapy; SECT = sham/simulated electroconvulsive therapy. *OR* = odds ratio between ECT and SECT; ES = standardizsed effect size.

[a]reported as 256 by UK ECT Group, by including withdrawers during four of the studies. [b]wrongly included photoshock data, without which the finding is nonsignificant. [c]"translates to" a mean Hamilton difference of 9.7 (95% CI 5.7–13.5). [d]same year as the meta-analysis so possibly not published in time. [e]extrapolated from one ECT–SECT study (Brandon et al., 1984) and multiple other (not ECT–SECT) studies. [f]no data in the only ECT–SECT study (Brandon et al., 1984) to directly support these *OR*s (see text).
*statistically significant finding.

One patient from each group was withdrawn in week one due to "lack of improvement." If both had been scored as 0 improvement, rather than excluded, this would, in such small groups, have slightly reduced the difference in mean improvement scores between the two groups. For example, the difference between the ECT and SECT groups in the mean amount of change in the psychiatrist's ratings would have fallen from 41.1 (48.4 vs 7.3) to 37.7 (44.4 vs 6.7). An additional ECT patient was withdrawn in week one because s/he "could not complete the Beck Depression Inventory." This person was withdrawn *after* baseline assessments so they must have become unable to respond (to written questions on a 0–3 scale) *after* one or two ECTs. So while it appeared that 11 out of 11 ECT patients improved significantly, the true proportion was 11 out of 13.

Despite the assertion that "These findings confirm the value of electric convulsion therapy in severe depressive illness," the two groups had average baseline Beck scores of only 24 and 27, which are within the "moderate" range of depression (20–28; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961). The baseline scores for the psychiatrists' ratings, on a scale with 100 representing "most severe depressive illness," were only 68 and 71.

Brandon et al. (1984; reviewed next), commenting on the West study, raise concerns about "The sample size, the unusually unequivocal result, problems of selection, and doubts about the extent to which blindness was achieved" (p. 23). West did not tell us how blindness was achieved by either "the psychiatrist in charge" or by the "nurses." We were not told how many nurses were raters, or anything about their role in treatment. We were not told how many patients had enhanced probability of knowing whether they had received ECT in the study because they had had it before. The "blindness" of the raters was not assessed.

Despite all these failings three meta-analyses include this study (using the data that ignored the two withdrawals), and use its aberrantly large pro-ECT findings in their calculations (Janicak et al., 1985; Pagnin et al., 2004; UK ECT Review Group, 2003).

Brandon et al. (1984)—16/24. The largest of the 11 studies (77 patients) took place in Leicester, England. It was a relatively high quality study. The samples were typical of ECT recipients in terms of depression severity, gender, and age. The blinding process was described and tested. Apart from failing to report means and SDs (provided later by Buchan et al., 1992), other failings included the fact that 60% had had ECT before (thereby reducing the probability of genuinely blind ratings by the patients) and that the patients' self-report scores were not reported. No explanation is given for this selective reporting.

On both the Hamilton and a psychiatrist's rating scale the 43 in the ECT group improved significantly more than the 34 in the SECT group. Analysis by Buchan et al. (1992), of the Brandon et al. (1984) and Johnstone et al. (1980) studies combined, however, found that the differences in Hamilton scores were only significant for patients who were "deluded" or "retarded" (slowed thoughts), which was less than half of the participants in the two studies (45%; Buchan et al., 1992, p. 357). None of the three meta-analyses that include the Leicester study (or the one that includes Johnstone's Northwick Park study) acknowledge this. Nor do they wonder why the patients' ratings were not reported.

Gregory et al. (1985)—10/24. The last ever ECT versus SECT study took place 35 years ago. The "Nottingham ECT Study" actually had three groups. The ECT participants were divided into two groups by electrode placement (unilateral or bilateral). It is almost impossible to make sense of the findings. "Of the 69 patients entering the study, 25 received fewer than six study treatments; these were classed as withdrawers" (Gregory et al., 1985,

p. 521). Of these 25 14 were withdrawn because of "failure to improve" and five because they "were better." So 19 of the 69 participants (27%) in a study designed to determine who got better were withdrawn because they did, or did not, get better. (Three of the withdrawers in the ECT group, but none in the SECT group, withdrew consent after the study started.) To further confuse matters Table 1 in Gregory et al. (1985) reports the mean scores of 60 people with "complete data available" although there were only 44 participants remaining after the 25 were withdrawn. A graph portraying changes on the Montgomery–Asberg depression scale (MADRAS; Montgomery & Asberg, 1979) seems to have numbers for each of the three groups closer to those expected when subtracting the withdrawers. Their Table reports "percentage changes" that are more than twice as *large* for the SECT group as for either of the two ECT groups, on both the Hamilton and MADRAS. Finally, a "global assessment of change in depression" was made, but not reported (thereby meeting the Cochrane criterion of "selective reporting").

A Cochrane review on ECT for "the depressed elderly" set out to calculate an effect size for the 35 participants over the age of 60 in this study but found, unsurprisingly, that insufficient data had been provided to make that possible (van der Wurff et al., 2003).

The only meta-analyses that includes this study (UK ECT Group, 2003) fails to acknowledge any of these major problems and unquestioningly included the strong finding in favor of ECT in their calculations of effect sizes.

Follow-Up Findings

Seven of the 11 studies provided follow-up data, but we shall see that only three produced meaningful data for comparing ECT and SECT. An eighth study had stated "We hope to report longer-term effects in a later article" (Fahy et al., 1963, p. 310), but they didn't.

Ulett et al. (1956). Six months after the end of treatment a comparison was made using patients who had been discharged and not received ECT after the end of the study period. Four of the 11 who had had ECT (36%) had relapsed, compared to none of the four in the SECT control group. The majority of patients in this study, however, did not have a depression diagnosis so this finding is irrelevant to the current review.

Brandon et al. (1984) and Gregory et al. (1985). Neither Brandon et al. (1984; 2 and 6 months follow-up) nor Gregory et al. (1985; 1 and 6 months) found significant differences between ECT and SECT at follow-up. Moreover, both studies invalidated any evaluation of long-term benefits by giving ECT to most of the SECT group during follow-up. Brandon et al. (1984), gave ECT to 20 of its 34 SECT participants, and to 17 of the 42 in the real ECT group, during follow-up. Gregory et al. (1985) gave an average of 4.1 ECTs to their SECT group and 1.5 to their ECT group during follow-up.

West (1981). West reported psychiatrists' scores on a 0–100 scale (but not the nurses' or patients' scores), 5 days after the last treatment. The difference in the size of change from baseline was an enormous 53.6 points (52.1 vs −1.5). If such data can be believed they would produce a rather incredible effect size (Cohen's *d*) of 3.22. We have already noted the serious methodological failings of, and ominous questions about, this study.

West then followed up for a further 3 weeks, but like Brandon et al. (1984) and Gregory et al. (1985) gave ECT to most of the SECT group (10 of the 11).

Further suspicion about this study comes from the fact that at the end of the first part of the study the mean psychiatrists' score for the 11 SECT patients was 63.4, but the "base" mean score for the follow-up study, for the 10 remaining SECT patients, was reported to have jumped to 73.4. This is not mathematically possible by excluding just one of 11 people.

Brill et al. (1959). This early USA study did not assess outcomes till a month after treatment ended. As we have seen, 16 of the 21 men in the ECT group (76%) and 4 of the 9 SECT patients (44%) met the researchers' criterion for recovery. The difference is not statistically significant (χ^2 = 2.86, p = .09). The effect size (*d*) is .297 (95% CI −0.44–1.04). This study had extensive methodological flaws, scoring only 9/24 on the Quality scale. As noted earlier, it involved 97 men with an average age of 35, so was totally unrepresentative of the modal ECT recipient—a woman in her 60s. None of the outcome measures explicitly assessed depression. "Nearly half" of the participants had had ECT before.

Lambourn and Gill (1970). Lambourn and Gill also followed up participants for a month. Because they reported detailed data for individual patients it is possible to calculate mean outcomes for the seven ECT patients and eight SECT patients who did not have ECT during the follow-up month. The researchers used a 67% or greater improvement (from baseline) on Hamilton scores as an indicator of improvement. This was achieved, at 1 month follow-up, by four of the seven ECT patients (57%) and five of the eight SECT patients (62%). The mean reductions in Hamilton scores were 30.57 (*SD* = 18.61) for the ECT group and 35.75 (*SD* = 17.65) for the SECT group, producing a difference of 5.18 and an *SD* for the whole sample of 18.10, which produces a "small" effect size (*d*) of .299, in favor of SECT.

(The researchers failed to report their data on number of hospital days during follow-up.)

Johnstone et al. (1980). Johnstone et al. assessed at 1 month and 6 months posttreatment, on three scales. There had been a significantly greater drop in Hamilton scores at the end of treatment for the ECT group, but:

> The advantage of real over simulated ECT was not retained and at the one-month and six-month follow-ups the Hamilton scores of the two groups were almost the same. The Leeds self ratings showed similar trends but these were never significant, and this was also true of the ratings by nurses. (p. 1318)

So none of the three sets of raters found a significant difference between ECT and SECT at one or 6 months after the end of treatment. Johnstone et al. (1980) reported no specific follow-up data, just graphs. Buchan et al. (1992), however, provided Johnstone et al.'s (1980) 6 months mean improvement scores on the Hamilton (but not the nurses' or patients' ratings). The mean reductions were 36.33 for the ECT group and 35.30 for the SECT group. Calculating an effect size for this small difference (1.03 points) is problematic, as we do not know the SDs. The SEs for the data at the end of treatment (3.0 for ECT and 2.7 for SECT) translate into SDs of 16.70 and 15.03 respectively ($SD = SE \times \sqrt{N}$). If we use those as estimates of the SDs after 6 months, the 1.03 difference between the amount

of change in the two groups translates into an effect size of .065. This does not approach the .2 level for a "small" effect size (Cohen, 1988).

Brandon et al. (1984) concluded (p. 23):

> The well designed and carefully-controlled clinical trial. . . (Johnstone et al., 1980) showed that electroconvulsive therapy had only a small effect in depression at the end of the trial period and there was no difference in the condition of patients given real and simulated treatment at one and six months of follow up.

Johnstone et al. (1980), themselves, emphasized this point (p. 1319):

> The most striking finding is that the differences which were present at the end of the course of eight treatments had disappeared one month later and were undetectable also at six months

Conclusion Regarding Long-Term Efficacy. A conservative conclusion from the four studies that provided some relevant data would be that there is no evidence that ECT has any lasting benefits beyond 5 days. Given all the problems with the West study it seems reasonable to exclude it from considerations and conclude that there is no robust evidence of ECT having any benefit at all beyond the last day of treatment.

If we consider only the three studies with data for at least 1 month we are left with one small effect size, .297, in favor of ECT (Brill et al., 1959), one study with a trivial effect size, .065, in favor of ECT (Johnstone et al., 1980) and one with a small effect size, .299, in favor of SECT (Lambourn & Gill, 1978). If we exclude the Brill study because of its multiple methodological flaws (not least its failed blinding process, and its being based on a very atypical sample of middle-aged men) we are left with Lambourn and Gill and Johnstone et al., (1980) two of the three highest Quality studies. Neither of these two studies, one with unilateral electrode placements and one with bilateral, provide any evidence of any long-term benefits of ECT compared to SECT.

The Five Meta-Analyses

The first meta-analysis (Janicak et al., 1985) was published in 1985, possibly too early to consider the last study (Gregory et al., 1985). Three meta-analyses were published nearly 20 years later, in 2003 or 2004. The last was published in 2019. All five concluded that ECT is more effective than placebo.

The five meta-analyses include, between them, the 11 ECT versus SECT studies described above. Table 4 shows the marked variation in the number of studies included in the meta-analyses, from one (Mutz et al., 2019) to seven (Pagnin et al., 2004). No study was included in all five meta-analyses. Most (eight) were included in just one or two meta-analyses.

Janicak et al. (1985). *Inclusion Criteria.* The first meta-analysis, by Janicak and colleagues, includes six studies (Table 4). The "most important" inclusion criterion is the ability "to determine each patient's response to treatment" (p. 298), and "the assessment of each patient's response was determined by the author's designation of the patient as a responder or nonresponder" (p. 298). Five of the six included studies meet the criterion

(see Table 3). One study does not but is included anyway; the one with the strongest outcome in favor of ECT (West, 1981). Although West recorded that it was considered "therapeutically desirable" (without stating by whom or by what criteria) for 10 of the 11 SECT patients to receive ECT in the second part of his study, he neither reported any scores or categorizations for individual patients nor designated participants as "responders." A second criterion is "systematic method for diagnosing the patient as depressed." This is not the case for three of the six (Table 3). A third criterion is that depression be "severe." Only two of the six studies met this criterion (Harris & Robin, 1960; Lambourn & Gill, 1978). One stated "Severe depressions with high suicidal risk were not included" (Fahy et al., 1963, p. 310).

Quality Control. Janicak et al. (1985) make no attempt to evaluate the methodological rigor of the six studies. They are either unaware of, or actively ignore, the 72 specific instances of methodological failings across the six studies (see Table 3). The six included studies had a slightly lower mean Quality score (11.17) than the five excluded studies (13.60), but the difference is not significant (t (9) = 1.30, p = .26).

Short-Term Findings. Efficacy was calculated "by taking the difference in percentage efficacy between real ECT and SECT and averaging across all studies." The reviewers report an "overwhelming statistical superiority of ECT over SECT" (Janicak et al., 1985, p. 301).

The totals they report from their six studies are 72% for ECT and 40% for SECT; hence the assertion that ECT is "32% more effective" (Janicak et al., 1985, p. 298). This is an incorrect calculation of the two percentages from their own Table (Janicak et al., 1985, p. 299). The numbers are, for ECT 73/109, which is 67% not 72%: and, for SECT, 33/96, which is 34% not 40%. These errors do not significantly alter the overall difference between the two conditions, but do indicate carelessness.

More importantly, the reported percentages of two of the six studies are incorrect. In their report of the Ulett et al. (1956) study, Janicak et al. (1985) wrongly include the data of patients subjected to photoshock. Without these patients the correct figures are ECT 7/21 (33%) versus SECT 5/21 (24%), a 9% difference, compared to a 30% difference (65% vs 35%) when the photoshock participants are included. Secondly, Brill et al. (1959) had reported (p. 630; Table 3) that the percentages meeting their criterion of showing improvement on two of their three measures as 76% "shock" versus 44% "nonshock" (16/21 vs 4/9). Janicak, however, report 67% versus 25% (p. 299; Table 1), thereby inflating the difference between real and SECT from 32% to 42%. The percentages using the *correct* numbers for the five studies that *did* report percentages of "responders" (i.e., excluding West, 1981—see above) are: ECT 45/79 (57%) versus SECT 25/67 (37%), a difference of 20%, rather than 32%. This is statistically significant (χ^2 = 5.61; $p < .05$), but not as strongly as Janicak's claim of χ^2 = 21.54 ($p < .0001$).

Four of their six studies (Brill et al., 1959; Fahy et al., 1963; Harris & Robin, 1960; Ulett et al., 1956) have the most methodological flaws of the 11 studies (see Table 3), all four having a Quality score of 10 or less out of 24 (see Table 3).

Follow-up Findings. Janicak et al. (1985) acknowledge that "questions such as those raised by" Johnstone et al. (1980) when they found no difference at follow-up are "left unanswered" (p. 301).

Kho et al. (2003). *Inclusion Criteria.* Eighteen years later Kho et al. (2003) published their meta-analysis in the *Journal of ECT.* It was based on just two studies. They excluded all pre-1978 papers, because of their diagnostic ambiguities (p. 140) and because they wanted to

determine "whether the superior efficacy of ECT is still found using more recently published studies" (p. 140). This assumption, that ECT had already been shown to have "superior efficacy," might be considered a sign of bias on the part of the authors.

Kho et al. (2003) set out to include only studies reporting means and standard deviations generated with depression rating scores such as the Hamilton (1960). They exclude two studies which meet this criterion (Johnstone et al., 1980; West, 1981), without explanation, and rely instead on just two studies (Brandon et al., 1984; Lambourn & Gill, 1978).

Quality Control. Kho et al. (2003) assess the quality of the studies on a 0–5 scale based on randomization, double-blindness and description of withdrawals. Eight of the sixteen various types of studies included in their broader meta-analyses scored 0 out of 5. They fail, however, to report the scores of individual studies.

This is the only meta-analysis where our 24-point Quality scale produces a significantly higher mean score for the included studies (16.50) than the excluded studies (11.33); (t (9) = 2.59, p = .029).

Short-term Findings. The two studies, involving 59 ECT patients and 50 SECT patients, produced four effect sizes. The reviewers calculate a pooled effect size (delta) of .95 (95% interval—-0.35 to +1.54). The reported effect sizes for the three subtypes of depression in the Brandon et al. (1984) study range from 1.38 to 1.99, all far higher than the .77 calculated by Pagnin for the three subtypes combined. Kho et al. (2003) acknowledge that "because the three ESEs from the Brandon study may be correlated, the results from the comparison between ECT and SET may be exaggerated" (p. 145). So three of the four effect sizes may have been "exaggerated" and the fourth (Lambourn & Gill, 1978) was calculated as .09.

Kho et al. (2003) fail to mention any of the problems of the two studies listed in Table 3, including the fact that in the Brandon et al. (1984) study 60% had had ECT before (thereby significantly compromising the blindness of the ratings by the patients) and that the patients' self-report scores scale were not reported.

Follow-up Findings. The issue of efficacy beyond the end of treatment was not mentioned.

UK ECT Review Group (2003). In the same year, 12 reviewers, led by Oxford University psychiatrists, published a meta-analysis funded by (but independent from) the UK Department of Health, and published in the *Lancet*. It is the only one of the four meta-analyses published at the time that was considered to be a "good-quality systematic review of randomized evidence" by a subsequent 170-page UK report for the National Health Service (Greenhalgh et al., 2005).

Inclusion Criteria. The "primary outcome" is "a continuous depressive symptoms scale" but "dichotomous data are merged to produce estimates of odds ratios" and the two are combined using "numerical simulation techniques based on Gibbs sampling" (p. 800).

Six of the 11 studies are included (see Table 4). Freeman et al. (1978) and Harris and Robin (1960) are included despite having invalidated their findings by giving ECT to the SECT group. There is no explanation for excluding four of the other five (although Table 3 shows there are good reasons to do so). Brandon et al. (1984) is excluded "because 43 patients had nondepressive diagnoses" (p. 806). This is incorrect. The 43 had been omitted from the study.

Quality Control. Greenhalgh et al. (2005, p. 15) note that "Little information was provided in the review (UK ECT Review Group, 2003) regarding the characteristics of participants in terms of the nature and severity of their condition, medication history and previous use of ECT." Quality is, however, evaluated, using four criteria: "reporting of allocation concealment, masking, loss to follow up, and length of follow up" (p.799). The UK ECT Group do comment that "The quality of reporting of the trials was poor" (p. 801), but fail to report the performance of individual studies. The reviewers acknowledge the small sample sizes and the absence of data on patients who are "most likely to receive it—e.g., older patients . . ." (p. 806). They are, however, unaware of, or actively ignore, the 47 other specific instances of methodological failings across their six studies (see Table 3).

The quality of the six included studies does not differ significantly from that of the five excluded studies (13.67 vs. 10.60; t (9) = 1.74, p = .12).

Short-term Findings. Unlike the other meta-analyses, which all presuppose that ECT is effective, these reviewers start by acknowledging that views vary, from "it is probably ineffective but certainly causes brain damage . . . through to those who think it is the most effective treatment available in psychiatry and is completely safe" (p. 799).

This is the only meta-analysis to include the Johnstone et al. (1980) study. Only the statistically significant outcome (Hamilton ratings by a single psychiatrist) is included. The nonsignificant findings, from the nurses' and patients' ratings, are ignored, without explanation.

This is also the only meta-analysis to include Freeman et al. (1978). It doesn't mention that ECT was given to SECT patients after a week, or that 20% of ECT patients withdrew unimproved.

The two studies with the largest effect sizes (Gregory et al., 1985; West, 1981) both have multiple methodological shortcomings (see above and Tables 2 and 3).

Ignoring all these problems the reviewers go on to combine the categorical and continuous outcome data to produce a pooled effect size of .91 in favor of ECT. The other four meta-analyses reached a generalized, unqualified conclusion that ECT "is effective." Although the the UK ECT Group (2003) also concluded that "In the short-term (ie at the end of treatment), ECT is an effective treatment for adult patients with depression" (p. 806), they added:

> There is limited randomised evidence on the efficacy of ECT in the specific subgroups of patients who are presently most likely to receive it—eg, older patients or those with treatment-resistant illnesses—or in subgroups of patients in whom ECT is thought to be especially effective. (p. 806)

Multiple emails were sent by JR to the lead author, Professor John Geddes, and other members of the UK ECT Review Group, seeking clarification about all the concerns raised above. Despite polite acknowledgements of the emails none of the questions were answered.

Follow-up Findings. This was the only meta-analysis to investigate longer-term efficacy. Only one study is identified (Johnstone et al., 1980) and "a non-significant two-point difference in final HDRS (Hamilton, 1960) was noted in favour of the simulated group" (p. 801). This is potentially misleading, in favor of SECT. Although the SECT group did end up two points lower, the ECT group had started off more depressed and had actually

changed 1.03 points more than the SECT group (Buchan et al., 1992, p. 358, Table 2), but neither difference is statistically significant.

Pagnin et al. (2004). *Inclusion Criteria.* The fourth meta-analysis was published in the *Journal of ECT*. It includes the largest number of studies, seven, and the largest number of people, 245. Like Janicak et al. (1985), the reviewers include only studies from which they could "determine each patient's response to treatment, using author's own criterion of response or no response." (p. 13), correctly excluding Freeman et al. (1978), Johnstone et al. (1980), and Gregory et al. (1985) on that basis, but, like Janicak et al. (1985) and the UK ECT Review Group, dubiously including West.

Quality Control. Pagnin et al. (2004) make no attempt to rate studies in terms of methodological rigor. The difference between the mean Quality scores of the seven included studies (11.86) was not significantly different from that of the four excluded studies (13.00), (t (9) = .55, p = .60). The reviewers acknowledged problems with "diagnostic heterogeneity," randomization, and maintaining blindness, but without naming any specific studies. They were unaware of, or actively ignored, the 74 other specific instances of methodological failings across the seven studies (see Table 3).

Short-term Findings. Despite only two of the seven studies (Brandon et al., 1984; West, 1981) producing a significant difference, the studies do, when combined, find a significantly greater mean effect size for ECT than for SECT at end of treatment (χ^2 = 6.87, p = .009). Four of the seven included studies had the four lowest Quality scores of the 11 (see Table 3; Brill et al., 1959; Fahy et al., 1963; Harris & Robin, 1960; Ulett et al., 1956) and were excluded by three of the other meta-analyses (Kho et al., 2003; Mutz et al., 2019; UK ECT Group, 2003) (see Table 3). It is also unclear how the effect sizes were calculated. For example Pagnin et al. (2004) report an effect size (D) of 1.341 for the Brill et al. (1959) study (Table 3, p. 15). Yet the 16/21 versus 4/9 improved ratios actually produce an effect size (D) of .297 (95% CI −0.44–1.04; using www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-SMD9.php; see Table 4).

The reviewers acknowledge that any advantage of ECT over SECT is only "specifically among patients with delusions and/or retardation [slowness of thought]" (p. 19).

Follow-up Findings. The absence of any evidence of efficacy beyond the end of treatment is, again, not mentioned.

Mutz et al. (2019). *Inclusion Criteria.* The most recent meta-analysis, from the *Institute of Psychiatry* in London, appeared 15 years later, in the *British Medical Journal* (Mutz et al., 2019). It differs from previous meta-analyses in being a network meta-analysis, making pair-wise comparisons, between four types of ECT and 14 types of brain stimulation, and, when possible, comparing these to sham placebo treatments.

Inclusion criteria required use of the Hamilton or Montgomery scales and a manual-based diagnosis of "major depressive disorder" or "bipolar depression." Outcomes were efficacy and discontinuation/acceptability. Only 2 of the 11 studies were included (Brandon et al., 1984; Gregory et al., 1985). Although not immediately apparent from the article, only one study (Brandon et al., 1984) actually contributed to the analysis regarding efficacy.

A personal communication (Mutz et al., 2019) responding to multiple questions from JR, explained: "The Gregory et al. (1985) study only contributed to the summary odds ratio for all-cause discontinuation as the authors did not report sufficient data in their paper to compute efficacy estimates."

Seven of the other nine studies are not mentioned at all, even in the 13 page "Full Texts Excluded" section of the Supplementary Material (pp. 32–44). The final two studies (Freeman et al., 1978; Johnstone et al., 1980), both published in the *Lancet*, are categorized as "Cannot be obtained" (Supplementary Material, p. 39). The personal communication did not answer the question "Does the Institute of Psychiatry not have access to papers published in the Lancet?" but did state that if they had managed to obtain these two papers (which JR had by now sent to them) neither would have met their inclusion criteria. The personal communication said the same of the seven studies which their paper failed to mention at all, but which they had also subsequently been sent by JR. For example, the Mutz et al. (2019) meta-analysis is the only one not to include the Lambourn and Gill study. The personal communication explained: "This trial was excluded as it did not meet our inclusion criteria of RDC, DSM or ICD diagnosis of major depressive disorder or bipolar depression."

So even after being sent all the studies which their search had missed, or they could not obtain, the *Institute of Psychiatry* reviewers conclude that after 80 years only one ECT–SECT study is robust enough to merit inclusion in meta-analyses.

Quality Control. The meta-analysis by Mutz et al. (2019) is the only one to report any sort of quality ratings for specific studies. Using Cochrane criteria they assess the only study they consider robust, in terms of their inclusion criteria, as having a "high risk" of bias, the worst Cochrane category.

Short-term Findings. Mutz et al. (2019) claim that their "network meta-analysis" produce odds ratios, relative to sham treatment, significantly in favor of ECT for "Bitemporal ECT" (bilateral) and "High-dose Unilateral ECT," but that the odds ratios for "Bifrontal ECT" and "Low to Moderate-Dose Unilateral ECT" are not significant. But the single ECT–SECT study they included only studied bilateral ECT, so conclusions about whether the other three electrode placements were superior to SECT were based on no ECT–SECT data at all. The personal communication explained:

> In the absence of head-to-head clinical trials, network meta-analysis allows us to estimate such treatment effects using data available from other treatment comparisons that share comparison treatments. For example, if we have data on treatment A vs treatment B and data on treatment A vs treatment C, we can estimate the effect of treatment B vs C. Please note that this is a somewhat simplified explanation.

In response to being asked why their review methodology led to an odds ratio for bilateral ECT far higher than the odds ratio calculated by the Pagnin et al. (2004) meta-analysis for the Brandon study, the reviewers replied: "The network meta-analytic ORs are not directly comparable to the individual study OR presented in the Pagnin et al. (2004) meta-analysis." This is very true. The OR calculated by Pagnin et al. (2004), based directly and solely on the ECT–SECT data of the Brandon study was 2.2. The OR calculated by Mutz et al. (2019), based on the Brandon data plus a lot of studies which do *not* compare bilateral ECT and SECT, is an enormous 8.9. Furthermore, their very large 7.3 OR for High-dose Unilateral ECT, is based *entirely* on studies that do not compare ECT and SECT.

We have already noted that the only ECT–SECT efficacy study that met their inclusion criteria was rated, by the reviewers themselves, as "high risk" of bias (Mutz et al., 2019, Supplementary Material, pp. 49, 50). They add:

> Overall risk of bias was deemed high in 19 trials (17%). In a sensitivity analysis excluding these trials, we found that . . . treatment effects of ECT protocols and magnetic seizure therapy versus sham therapy could not be estimated. (Mutz et al., 2019, p 10).

Nevertheless, they ignore their own statement, and proceed to estimate and report the treatment effects, unqualified, in the Abstract:

> 10 out of 18 treatment strategies were associated with higher response compared with sham therapy: bitemporal ECT (summary odds ratio 8.91, 95% confidence interval 2.57 to 30.91), high dose right unilateral ECT (7.27, 1.90 to 27.78). (Mutz et al., 2019, p.1)

Follow-up Findings. The reviewers make no attempt to review the literature regarding longer-term effects of ECT.

DISCUSSION

The Quality of the 11 Studies

Table 3 shows that the 11 studies produced Quality scores, on our 24-point scale, ranging from 9 to 17, with a mean score of 12.27 (*sd* = 3.20). Only three produced scores above 13.

The empirical support for using ECT prior to 1978 had consisted of just five ECT versus SECT studies, on a total of 67 ECT patients and 57 SECT controls, with a mean Quality score of 9.80 out of 24. Four of the five had found no difference between ECT and SECT. The only one finding a significant difference (Wilson et al., 1963) involved just four ECT patients.

The quality of this body of literature as a whole is unimpressive, and is clearly unable to determine whether ECT is more, or less, effective than SECT in reducing depression. Table 3 shows, for example, that 5 of the 11 studies (including three of the second wave) failed to describe their randomization process. Five (including two later studies) reported no attempt to test their blinding process. Of the six that did so, five assessed the blindness of the raters but not that of the patients; mostly by asking raters to guess whether patients had received ECT or SECT and finding no more agreement than that expected by chance (Brandon et al., 1984; Freeman et al., 1978; Johnstone et al., 1980; Lambourn & Gill, 1978), and in one instance by just reporting that it was "easy" for the observers to infer which treatment had been allocated (Fahy et al., 1963). The sixth study (Brill et al., 1959) tested the patients but not the raters, reporting that "some patients in the nonshock group believed that they were receiving some new variation of ECT." So none of the studies tested the blinding process for both the raters and the patients.

The second reason that none of the studies can reasonably claim to be double-blind is that none of them excluded people who had previously had ECT, so some members of the SECT groups would probably know they had not had ECT because they would know that ECT is always followed by headaches and temporary confusion. None of the studies

showed any awareness of this issue. Five of the 11 did not even report how many people had previously had ECT (see Table 3). Table 2 shoes that the other six reported percentages ranging from 21% (Johnstone et al., 1980) to 66% (Lambourn & Gill, 1978), with a weighted mean of 45.1% (the "nearly half" reported by Brill et al. (1959) was interpreted to be 14/30; 47%). So about half the patients in the SECT groups would probably have guessed that they had not had ECT. Therefore, none of the studies could genuinely be described as double-blind.

Two-thirds of ECT recipients are women and the average age is between 60 and 65 (Read & Bentall, 2010; Read et al., 2013, 2018); so the modal ECT person is a woman in her early sixties. Tables 2 and 3 show, however, that only three studies met the criterion of being broadly representative of the demographics of ECT recipients by using samples that were mostly female and had an average age of at least 50. None of the studies showed any interest in age or gender. None analyzed their findings by age or gender. None even reported ethnicity.

ECT is supposed to be given to severely depressed patients. Current guidance from the UK's National Institute for Health and Care Excellence states: "Consider ECT for acute treatment of severe depression that is life-threatening and when a rapid response is required, or when other treatments have failed. Do not use ECT routinely for people with moderate depression . . ." (National Institute of Clinical and Health Excellence [NICE], 2009). Five studies, however, failed to demonstrated that their participants were severely depressed; three did not provide enough information to know, and two clearly had only (Fahy et al., 1963) or mostly (West, 1981) moderately depressed participants. One used participants (62%) without a depression diagnosis at all (Ulett et al., 1956).

Two of the 11 studies invalidated their findings by administering ECT to the SECT group part way through the studies (Freeman et al., 1978; Harris & Robin, 1960). Table 3 reports that only five studies reported means and standard deviations on a dimensional depression scale such as the Hamilton, which is valuable for calculating an effect size and thereby making a meaningful contribution to a meta-analysis.

Only one of the studies reported whether other treatments (e.g., antidepressants or CBT) had been unsuccessfully tried prior to ECT, which would have rendered the studies able to assess whether ECT is effective for people who are today recommended for ECT by NICE guidelines (see above). In the only study that did report, less than half (46%) had been tried on antidepressants prior to the study (Freeman et al., 1978).

Only four studies included ratings by the patients themselves, and none assessed the impact of ECT, positive or negative, on their Quality of Life.

The sample sizes were small, ranging (ECT and SECT groups combined) from eight (Harris & Robin, 1960) and 10 (Wilson et al., 1963) to 77 (Brandon et al., 1984). The mean was 38.3; with 20.4 in the ECT groups, and 17.9 in the SECT groups.

Five studies selectively reported their outcomes, failing to report one or more findings.

The Quality of the Five Meta-Analyses

All five of the meta-analyses claim that ECT is effective for depression but, as we have seen, they are all of a poor standard, not least because none of them pay sufficient attention to the quality of the papers on which they base this claim. The only meta-analysis conducted in the last 15 years, the one from the *Institute of Psychiatry* in London in 2019, is particularly

problematic. Mutz et al. (2019) make strong claims about the efficacy of ECT on the basis of just one ECT–SECT study (Brandon et al., 1984). They not only rated, themselves, that one study as having a "high risk" of bias by Cochrane criteria but stated that exclusion of high risk studies made it impossible to estimate an odds ratio for ECT. Furthermore 67% of the other studies (not ECT–SECT) in their network analysis, used to indirectly calculate odds ratios were, themselves, either "unclear risk" or "high risk" (Mutz et al., 2019, p. 6). As was the case for the other four meta-analyses, major flaws have to be ignored to claim that ECT is more effective than SECT.

Four of the five meta-analyses fail to report the quality of any of the studies they include, most of which are of a very poor standard. The exception is the recent *Institute of Psychiatry* meta-analysis, which, as we have seen, reports that the only study they include had an overall "high risk" of bias. It is worth noting that the study (Brandon et al., 1984) that Mutz et al. (2019) assessed as having a "high risk" of bias is the 3rd most rigorous study of the 11 studies according to our own Quality scale, suggesting that the other eight may be at least as equally problematic.

Given the overall low quality of the 11 studies it would be particularly important that only the best studies are included in meta-analyses. The authors' apparent disinterest in the fact that none of the studies were actually double-blind, in whether the participants were representative of who receives ECT in clinical practice, in whether ECT has any advantage over SECT beyond the end of treatment, and in the pervasive selective reporting, are all indicative of carelessness, bias, or both.

Short-term Efficacy

Contrary to the claims by the authors of all five meta-analyses, the small number of studies, the small samples and the plethora of fundamental methodological flaws of most of the studies, render it impossible to determine whether or not ECT is superior to SECT during the treatment period,. The only three studies scoring 16/24 or higher on the Quality scale produced the following outcomes:

- Brandon et al. (16/24)—significant difference on psychiatrists' ratings, but patients' ratings not reported;
- Johnstone et al. (17/24)—no difference on nurses' ratings, no difference on patients' ratings; significant difference on psychiatrists' ratings (but for only two of three types of depression);
- Lambourn and Gill (17/24)—no difference on Hamilton scores or on psychiatrists' ratings.

This amounts to one of seven sets of ratings being significant and one partially significant.

While most of the 11 studies should never have been included in meta-analyses, it seems desirable to perform a meta-analysis on these three relatively high quality studies (keeping in mind that Mutz et al. (2019) evaluated the Brandon et al. (1984) study as "high risk" of bias). However, this is impossible because all three are guilty of selective reporting. One (Johnstone et al., 1980) failed to provide any data for two of their findings (both were merely reported as nonsignificant) and another (Brandon et al., 1984) failed to report anything at all about one of its two outcome measures (patients' self-ratings). The only

good-quality study to fully report its short-term findings (Lambourn & Gill, 1978) found no difference between ECT and SECT on either of its two measures.

Long-term Efficacy

For the same reasons (but with even fewer studies) it is impossible to know whether or not ECT has any benefits, in terms of depression reduction, beyond the time of the last shock treatment. None of the three studies producing meaningful data found a significant difference. The best two studies found a near-zero effect size toward ECT of .065 (Johnstone, 1999) and a "small" (.299) effect size in favor of SECT (Lambourn & Gill, 1978). So it could be tentatively concluded that there really is no benefit beyond the end of treatment. To do so, however, on the basis of just two or three small studies, would be wrong. The truth is, as is the case for the short term, we don't know.

Severely Depressed / Suicidal / "Treatment Nonresponders"

Even if one were to throw methodological caution to the wind, as the meta-analyses have done, and conclude that taken together there is some evidence that for the participants in the 11 studies there is, in general, an ECT–SECT short-term difference, this could definitely not be said to be true for the people who are supposed to receive ECT today—severely depressed, suicidal patients for whom other treatments have failed (NICE, 2009). Only six of the studies definitely included only or mostly severely depressed people. Two clearly did not. Although suicidal patients would probably have been included by chance in some studies, only two reported whether suicidal patients were actually included. The first actively excluded them (Fahy et al., 1963). In the second, only four of 31 (13%) people starting the trial had previously tried to kill themselves; and three of these four were withdrawn from the study (Harris & Robin, 1960).

We do not know, either, whether ECT is effective for people who have not responded to antidepressants or psychological therapies, the other major criterion for ECT use today, as we do not know how many, if any, such people were studied.

Suicide Prevention

Government and professional guidelines have claimed, for decades, that ECT prevents suicide. Suicidality is said to be a key indicator of suitability for ECT. None of the meta-analyses report any findings that ECT is more effective than SECT at preventing suicide. There are none (Read & Arnold, 2017; Read & Bentall, 2010; Read et al., 2013). Although the Hamilton, MADRAS, and Beck depression scales all include questions about suicidal intent, only one study reported these specific outcomes. Lambourn and Gill (1978) found mean reductions on the suicide item of the Hamilton scale of 3.38 points in the ECT group and 3.32 in the SECT group.

The UK ECT Review Group states: "Although ECT is sometimes thought to be a life-saving treatment, there is no direct evidence that ECT prevents suicide" (p. 806). The 170-page UK government report states: "The evidence did not allow any firm conclusions to be drawn regarding the . . . impact of ECT on all-cause mortality." (Greenhalgh et al., 2005, p. X).

Quality of Life

Quality of life measures can provide a more comprehensive and holistic assessment of our well-being than a depression scale; and one's quality of life can influence one's mood. None of the studies attempted to determine whether ECT improves quality of life, a failing noted by Greenhalgh et al. (2005, p. 15).

Patients' Experience

Only five studies included (and only four reported) any measure completed by the patients themselves. We agree with Kingsley and Patel (2017) that patient-reported outcome measures should be included in clinical trials and meta-analyses of psychiatric conditions. In one of the four studies that did report the patients' assessments of change, the psychiatrists reported a significant difference between ECT and SECT and the patients did not (Johnstone et al., 1980). In another study both the psychiatrists' ratings produced a significant difference but only one of the two self-rated scales did so (Freeman et al., 1978).

Gender

Women are twice as likely to receive ECT as men (Leiknes et al., 2012; Read et al., 2013, 2018). Yet none of the 11 studies or meta-analyses reported whether ECT was more or less effective for this group. Seven of the eight mixed gender studies failed to report data by gender. The two all-female studies produced one positive (Wilson et al., 1963) and one negative finding (Harris & Robin, 1960)—both with tiny samples.

The only study to report data for individuals by gender (Lambourn & Gill, 1978) allows us to calculate that the nine women who received ECT had a mean reduction on the Hamilton of 30.0 points, while the nine in the SECT group had a mean reduction of 18.6, a difference of 11.4 in favor of ECT. The men, however, had mean reductions of 21.4 points with ECT and 27.4 points with SECT, a difference of 6.0 points in favor of SECT. This suggests that ECT may be initially effective for women, but not for men. However, at 1 month follow-up (excluding those who received ECT after the end of treatment) the four women in the SECT group had a mean improvement of 4.0 points greater than the four women in the ECT group, while the four men in the SECT group had a mean improvement of 9.7 points greater than the three men in the ECT group. (This study, like Johnstone et al. (1980), used only one rater for Hamilton scores, and apparently doubled the scores of that person.)

Thus, there is only scant evidence that ECT might be effective in the short-term for one of its major target groups—depressed women; and none that it is effective beyond the end of treatment for them. The 170-page report conducted for the UK's National Health Service concluded "The evidence did not allow any firm conclusions to be drawn regarding the efficacy of ECT in . . . women with psychiatric problems" (Greenhalgh et al., 2005, p. X).

Age

The average age of ECT recipients is usually between 60 and 65 (Leiknes et al., 2012; Read et al., 2013, 2018). One would assume that studies and meta-analyses would therefore pay particular attention to older people. However, with the exception of the smallest study

(Harris & Robin, 1960), the average age of the samples ranged from 35 to 54, and some had no patients at all over 60, or 65 (see Table 1). No analyses by age were conducted by any of the studies.

One study did report individuals' ages and outcomes (Lambourn & Gill, 1978). The six people aged 60 or older who received ECT had a mean fall in Hamilton scores of 16.7, while the 10 aged under 60 had nearly double the improvement (32.0), a large, but non-significant, difference (t (14) = 1.77, p = .09). Improvement in the under 60s was, on average, 10.3 points greater in the ECT group than in the SECT group. In the 60 or over group improvement was an average of 8.7 points greater in the SECT group than the ECT group. Six of the 10 under 60s, but none of the 60 or older group, scored a 3 on the 0–3 doctors' scale, a significant difference (χ^2 = 5.76, p = .016).

One meta-analysis (Kho et al., 2003) found no difference between patients over and under 65 (p. 143; based on 15 ECT samples in studies *without* SECT groups). An additional meta-analysis, a Cochrane review, reported specifically on the effectiveness of ECT for the "depressed elderly" (van der Wurff et al., 2003). It identified only one study comparing ECT and SECT (O'Leary et al., 1994). This was a re-analysis of data for the 35 people aged over 60 in the Gregory et al. (1985) study. Twelve of the 35 had been withdrawn before completion of the study and the reviewers identified additional "major methodological shortcomings" before deciding that "None of the objectives of this review could be adequately tested because of the lack of firm, randomised evidence" (p. 2).

The UK ECT Review Group similarly concluded:

> Despite the reputation of ECT for efficacy in older patients, elderly people tend to be under-represented in trials, which limits the confidence with which results can be used to lend support to clinical practice in this subgroup. (p. 806)

Greenhalgh et al. (2005) concurred, with: "There was no randomised evidence of the efficacy of ECT in people older than 65 years" (p. 45) and "The evidence did not allow any firm conclusions to be drawn regarding the efficacy of ECT in older people." (p. 81)

Thus, there is no evidence that ECT is effective for another of its major target groups —the depressed elderly, either in the short or longer term. Use with this group is especially problematic because it is well established that older people are particularly likely to develop memory loss as a result of ECT (Mosti & Brook, 2019; Sackeim et al., 2007).

Children or Adolescents

No children or adolescents were included in any of the studies. There is no placebo-controlled evidence that ECT is, or is not, effective for these groups, either in the short or longer term.

Involuntary Patients

Many ECT recipients are given it against their will; about 40% in England (Read et al., 2018). None of the studies or meta-analyses addressed the issue of whether the trauma of being forced to undergo ECT after stating that you do not want it reduces the probability of a positive outcome. The UK government's report noted that even what they considered to

be the best of the meta-analyses (UK ECT Review Group, 2003) "did not identify any trials that explored . . . the impact of consumer choice on the outcomes of ECT" (Greenhalgh et al., 2005, p. 15)

Six of the 11 studies made no mention of whether some participants were being coerced to have ECT against their will, or even whether participants gave consent to take part in the study (Brill et al., 1959; Fahy et al., 1963; Freeman et al., 1978; Harris & Robin, 1960; Ulett et al., 1956; Wilson et al., 1963). These studies included most or all patients given ECT in a particular hospital and therefore almost certainly included some patients detained under mental health legislation and/or given ECT against their will. Wilson et al. (1963) refer to the withdrawal of "a voluntary patient signed out by husband" implying that some participants were involuntary. Two studies reported that participants gave consent for the study but made no mention of whether some participants were being coerced to have ECT (Lambourn & Gill, 1978; West, 1981). Three studies explicitly excluded people who were being treated under the Mental Health Act or were been given ECT against their expressed wish (Brandon et al., 1984; Gregory et al., 1985; Johnstone et al., 1980).

Only one of the five studies that found no difference between ECT and SECT, therefore, had excluded people who were having ECT against their will, but the three studies that did make this an exclusion criterion produced positive findings. Thus, it is possible that ECT is even less effective under compulsion than when undertaken voluntarily. This makes intuitive sense, but the evidence is weak. It is all we have to go on, as none of the studies that did include coerced patients analyzed their outcomes separately; and those later studies that (for sound ethical reasons) excluded coerced patients could not answer the question.

What can safely be concluded is that there is no evidence that ECT is effective for coerced patients, either in the short or longer term. This is perhaps the most alarming of all our specific findings. To administer a treatment involving multiple use of general anaesthesia, multiple electric shocks and multiple grand mal convulsions, against someone's will, is unethical. To do so even in the absence of any evidence that there is a good chance of a positive outcome is especially alarming. We have no idea whether this treatment works under compulsion. To do so, therefore, is clearly both unscientific and unethical.

Unilateral versus Bilateral

The purpose of the current review is to determine whether the meta-analyses were correct to claim that ECT is, in general, more effective than SECT, not to compare different types of ECT. We should nevertheless report that only two of the 11 studies used unilateral electrode placements. All the participants in the Lambourn and Gill were administered unilateral ECT, which produced the same outcomes as SECT at the end of treatment and worse outcomes than SECT at follow-up. In the Gregory et al. (1985) study both unilateral and bilateral placement produced significantly better outcomes than SECT at the end of treatment, but no meaningful follow-up occurred. Therefore, the millions of administrations of unilateral ECT over the past 35 years (Leiknes et al., 2012), since the 1985 Gregory et al. (1985) study, have been based on one positive and one negative finding in the short term and one negative finding at follow-up.

Placebo

Hope is a powerful placebo factor in psychiatric treatments, biological or psychological. It effects doctors, nurses, patients, and their loved ones. It can influence not just perceptions

of recovery but actual recovery. In the 1940s psychiatrists were excited about the new treatment. Hope of recovery had returned to some of the most depressing of institutions. Neurologist John Friedberg suggested that in those early days "the influence of ECT was on the minds of the psychiatrists, producing optimism and earlier discharges" (Friedberg, 1976).

Almost all the 11 SECT studies found that having a series of general anaesthetic procedures in the belief that you are having a major medical procedure that the doctors and nurses believe in can temporarily improve mood. Some of the researchers commented on this:

> One possibility is that the effective therapeutic component of ECT is the repeated rapid induction of unconsciousness in the patient. . . . It could very well be that the primary therapeutic agent is the psychological meaning of the treatment to the patient. . . . The influence of the unusual amount of care and attention which all receive could be studied further. (Brill et al., 1959, p. 633).

> Effectiveness . . . is due in large part to the attendant procedures associated with, the administration of an anaesthetic and the mystique associated with an unusual form of treatment. (Lambourn & Gill, 1978, p. 519).

> The results confirm that many depressive illnesses although severe may have a favourable outcome with intensive nursing and medical care even if physical treatments are not given. (Johnstone et al., 1980, p. 1319)

Brandon et al. (1984, p. 23) noted that an early version of convulsive therapy had been abandoned because it was no better than placebo:

> If the undoubted beneficial effects of electroconvulsive therapy were due to an elaborate placebo response the treatment would be comparable with insulin coma therapy, in which Ackner et al. had shown that any effects were not due to the induction of coma with insulin. The absence of a specific antidepressant effect would provide a strong case for abandoning electroconvulsive treatment.

A review focussing just on the placebo response with ECT (Rasmussen, 2009) found "an unexpectedly high rate of response in the sham groups" and concluded "The modern ECT practitioner should be aware that placebo effects are commonly at play" (p. 59). Furthermore:

> It is recognized that through a complex set of circumstances related to the meaning a patient ascribes to encounters with health care providers, which are influenced by cultural factors, individual life experiences, education, and the manner in which doctors communicate, expectations develop in the mind of the patient which by themselves can result in measured improvement in the condition at hand. . . . Finally, one also should not discount the effect of the natural history of depressive episodes. In none of the studies was there an untreated, natural history control group. Patients tend to get better on their own, even without treatment. (p. 58)

Lambourn and Gill reiterated that last, crucially important but often ignored, point:

> The contribution of spontaneous remission during this study remains an unknown factor because of the lack of a totally untreated group. (p. 515)

Does Including Participants Who have had ECT Before a Study Un-blind the Study and Thereby Exaggerate ECT Superiority Over SECT?

Only one study (Lambourn & Gill, 1978) provides data that can test the hypothesis that having previously had ECT un-blinds participants because they know that ECT is always followed by headaches and disorientation and, therefore, know if they have had ECT in a study. Among the 16 people in the SECT group, the 10 who had had one or more previous courses of ECT improved less (20.3 Hamilton points) than the six who had never had it before (27.2). Furthermore, the number of previous courses of ECT was related to degree of improvement on the Hamilton scale (r = .51; p = .044). So greater familiarity with the immediate adverse effects of ECT reduced the probability of benefitting from the placebo effects of SECT because they were more likely to know they had not received ECT. Analysing just the data for the 11 people who had never had ECT before shows that the SECT group had slightly more improvement (27.2 points; *SD* = 17.2) than the ECT group (20.0; *SD* = 17.0). Analysing the data for the 25 who had had ECT previously shows the opposite, with the ECT group improving more (29.1 points; *SD* = 18.3) than the SECT group (20.3; *SD* = 23.2). This suggests that by not excluding people who have previously had ECT all 11 studies exaggerated the difference between ECT and SECT in ECT's favor, and that none were truly blind studies.

Cost-Benefit Analysis

The fact that we don't know whether ECT has any short- or long-term benefits must be weighed against what we do know about its adverse effects, which are summarized briefly.

Brain Damage and Memory Dysfunction. Although ECT has a range of adverse psychological and emotional effects (Johnstone, 1999), the best documented findings are that ECT causes both major types of memory loss: anterograde amnesia (inability to retain new information) and, more commonly, retrograde amnesia (loss of memory for past events).

A 2003 review identified four studies of memory loss at least 6 months post-ECT (n = 597), and found a frequency range of 51% to 79%, and a weighted average of 70% (Rose, Wykes, Leese, Bindmann, & Fleischmann, 2003). Four studies (n = 703) found a range for "persistent or permanent memory loss" of 29% to 55%, with a weighted average of 38% (Rose et al., 2003). In 2007 ECT proponent Professor Harold Sackeim et al. conducted the largest prospective study to date and found that autobiographical memory was significantly worse than pre-ECT levels (p < .0001) 6 months later (Sackeim et al., 2007). Degree of impairment was significantly related to number of treatments. Even with the conservative cut off of two standard deviations worse than pre-ECT scores, 12% had "marked and persistent retrograde amnesia," with higher rates for the two demographic groups who receive ECT disproportionately—women and older people. Impairment was also greater among those who received bilateral ECT rather than unilateral ECT.

The most recent review (Mosti & Brook, 2019, p. 153) concludes that:

> Recent meta-analyses suggest the most prominent deficits are on measures of attentional/executive control (i.e., tests measuring cognitive flexibility, inhibitory control, and processing speed) and auditory verbal learning/recall (i.e., unstructured list learning), a memory task that is also strongly correlated with executive functioning.

ECT proponents often argue that these adverse effects are caused by depression not ECT (Read & Bentall, 2010, p. 343; Read, Cunliffe, Jauhar, & Mcloughlin, 2019), but a 2006 review concluded that "There is no evidence of a correlation between impaired memory/cognition after ECT and impaired mood, much less a causal relationship" (Robertson & Pryor, 2006, p. 230). The Sackeim et al. (2007) study confirmed that conclusion.

A New Zealand Government report stated "ECT may permanently affect memory and sometimes this can be of major personal significance" and noted the "slowness in acceptance by some professional groups that such outcomes are real and significant in people's lives" (Ministry of Health, 2004, p. 16). The American Psychiatric Association (APA) (2001) has admitted "In some patients the recovery from retrograde amnesia will be incomplete, and evidence has shown that ECT can result in persistent or permanent memory loss."

Sadly, the severity and significance of the brain damage and memory loss is rarely studied. It is not hard, however, to find hundreds of personal accounts of debilitating levels of disruption to people's lives. See, for example: https://ectjustice.com/ect-survivor-stories/ and https://www.madinamerica.com/2016/04/comments-by-shock-survivors-and-their-loved-ones/.

A recent USA class action lawsuit was settled on eve of trial at a Federal Court, which had ruled "A reasonable jury could find that the ECT device manufacturer failed to warn plaintiffs' treating physicians of brain damage resulting from ECT ("Breggin, 2018; Schwartzkopff, 2018). The manufacturer, *Somatics*, immediately issued a Regulatory Update to add "permanent brain damage" to the list of risks (Somatics, 2018, p. 4).

"Brain Damaging Therapeutics." The UK ECT Review Group found that bilateral ECT produces greater cognitive impairment than unilateral. Gregory et al. (1985) also discuss the "undoubtedly greater memory impairment produced by bilateral ECT" (p. 523). The 170-page review by Greenhalgh et al. (2005) concluded that any gains of using bilateral rather than unilateral ECT "are achieved only at the expense of an increased risk of cognitive side-effects" (p.1).

If the modest, temporary effects on depression are only to be found if the shock is passed across both temporal lobes, thereby causing maximal memory loss, this would confirm the early theories about how ECT works. Early postmortem examinations had led to the article "Brain damaging therapeutics" where the psychiatrist who introduced ECT to the US wrote, "The greater the damage the more likely the remission. . . . Maybe it will be shown that a mentally ill patient can think more clearly and more constructively with less brain in actual operation" (Freeman, 1941). A colleague had explained: "There have to be organic changes or organic disturbances in the physiology of the brain for the cure to take place. I think the disturbance in memory is probably an integral part of the recovery process" (Myerson, 1942).

A review (involving JR) of the effects of ECT on the brain put it this way:

> We suggest that the temporarily improved scores on depression instruments following ECT reflect the combination of frontal and temporal lobe functional impairments and activation of the HPA axis and the mesocorticolimbic dopamine system. These effects as well as other detailed changes observed in structures such as the hippocampus appear consistent with those typically seen after severe stress-exposure and/or brain trauma. (Fosse & Read, 2013, p. 6)

Mortality Rates. The idea that the mortality rate is "1 per 10,000 patients or 1 per 80,000 treatments" has been promulgated, without supporting evidence, by psychiatric associations (APA, 2001; Royal College of Psychiatry [RCP], 2017) and the USA's Food and Drug Administration (2011). A recent study put it even lower, at "2.1 per 100,000" treatments (Tørring, Sanghani, Petrides, Kellner, & Østergaard, 2017); but this was based on medical records (relying on staff recording that they had caused a death). Numerous studies (see Read & Bentall, 2010; Read et al., 2013) have found mortality rates many times greater than these claims. For example, of 8,148 ECT recipients in Texas, seven died within 48 hours (Shiwach, Reid, & Carmody, 2001). Excluding the two which the researchers argued were "unlikely to have been related to ECT" this is one per 1,630. Eight more died within 2 weeks, of "cardiac event" (the most common ECT-related cause of death). If these are included the rate becomes one per 627. When researchers wanted to interview 183 people, 1 year after ECT, it was reported that two (one in 91.5) had died during the ECT (Freeman & Kendell, 1980). A 1980 study (relying on British psychiatrists' reports of deaths from the ECT they had administered) found that four out of 2,594 ECT patients had died within 72 hours (one per 648.5; Pippard & Ellam, 1981). It could not be determined whether the one death (4 days post-ECT) among 75 French ECT recipients was ECT-related. This study, by anesthetists, found "potentially life-threatening complication" for 12 (16%; Tecoult & Nathan, 2001).

The oft repeated claim that ECT causes no more deaths than general anaesthesia unashamedly ignores the fact that people are subjected to an average of eight such procedures.

LIMITATIONS

The major limitation of any review designed to determine whether ECT works is the low quantity and poor quality of the available studies. The goal of the current review, however, is different; to evaluate the quality of the studies and of the meta-analyses that cite them.

Given the small number of studies, caution should be exercised when interpreting non-significant *t* tests involving the 11 studies, which might have been significant had there been more studies.

CONCLUSIONS

The scarcity and poor quality of most of the findings suggesting that ECT has short-term benefits for some depressed people, the complete lack of evidence of long-term benefits, and the absence of evidence that it prevents suicide, together with the high risk of permanent memory loss and small increased risk of death, broadly confirms the conclusions of previous reviews (Read & Arnold, 2017; Read & Bentall, 2010; Read et al., 2013; Ross, 2006) and books (Andre, 2008; Breggin, 2008). For example (Read & Bentall, 2010):

> Given the strong evidence of persistent and, for some, permanent brain dysfunction, primarily evidenced in the form of retrograde and anterograde amnesia, and the evidence of a slight but significant increased risk of death, the cost-benefit analysis for ECT is so poor that its use cannot

> be scientifically justified (p. 333). . . . The very short-term benefit gained by a small minority cannot justify the significant risks to which all ECT recipients are exposed. (p. 344)

Perhaps, however, given the outcome of this first ever analysis of the quality of the 11 studies that have attempted to determine if ECT is better than placebo, a more accurate conclusion, rather than "a very short-term benefit gained by a small minority" is that we just don't know whether ECT is better than, worse than, or no different from, placebo.

What can the 11 SECT studies tell us about seven specific sub groups? Firstly, we can reasonably conclude that there is no rigorous evidence whatsoever that ECT has any benefit for the three conditions for which it is primarily recommended today: (a) severely depressed people, (b) acutely suicidal people, and (c) people for whom antidepressants and/or psychological therapies do not work. Women and older people are the target demographics for ECT in the 21st century, but there is hardly any specific evidence that ECT is better than SECT for (d) women, in the short-term, and none regarding the long term; plus women are particularly likely to suffer long-term memory loss. There is no evidence whatsoever that ECT is superior to SECT in (e) older people, who are also differentially susceptible to memory loss. There is no evidence that ECT is effective (f) when given under compulsion, as it so often is. There is also no evidence that it is effective for (g) children or adolescents.

Our conclusions regarding depression parallel those of a recent commentary on Cochrane reviews of ECT for "schizophrenia" (Shokraneh, Sinclair, Irving, & Aali, 2019):

> What is common in all versions of these Cochrane reviews is that in spite of seven decades of clinical use of ECT for people with schizophrenia, there still is a lack of strong and adequate evidence regarding its effectiveness and the question "should we stop using electroconvulsive therapy?" is currently unanswered for people with schizophrenia.

The remarkably poor quality of the research in this field, and the uncritical acceptance of that research by psychiatry's meta-analyses, and its professional bodies, all of which endorse ECT as an effective and safe treatment, is a sad indictment of all involved, and a grave disservice to the public.

REFERENCES

American Psychiatric Association. (2001). *The practice of ECT: A task force report* (2nd ed.). Washington, DC: Author.

Andre, L. (2008). *Doctors of deception: What they don't want you to know about shock treatment*. New Brunswick, Canada: Rutgers University Press.

Beck, A., Ward, C., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561–571. https://doi.org/10.1001/archpsyc.1961.01710120031004

Brandon, S., Cowley, P., McDonald, C., Neville, P., Palmer, R., & Wellstood-Eason, S. (1984). ECT results in depressive illness from the Leicestershire trial. *British Medical Journal*, *288*, 22–25. https://doi.org/10.1136/bmj.288.6410.22

Breggin, P. (2008). *Brain-disabling treatments in psychiatry*. New York, NY: Springer Publishing.

Breggin, P. (2018). *Victory in ECT manufacturer lawsuit paves way for more*. Retrieved from https://www.madinamerica.com/2018/10/huge-breakthrough-ect-lawsuit

Brill, N., Crumpton, E., Eiduson, S., Grayson, H., Hellman, L., & Richards, R. (1959). Relative effectiveness of various components of ECT. *Archives of Neurology and Psychiatry*, *81*, 627–635. https://doi.org/10.1001/archneurpsyc.1959.02340170093010

Buchan, H., Johnstone, E., McPherson, K., Palmer, R., Crow, T., & Brandon, S. (1992). Who benefits from electroconvulsive therapy? Combined results of the Leicester and Northwick Park Trials. *British Journal of Psychiatry*, *160*, 355–359. https://doi.org/10.1192/bjp.160.3.355

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Routledge.

Fahy, P., Imlah, N., & Harrington, J. (1963). A controlled comparison of ECT, imipramine and thiopentone sleep in depression. *Journal of Neuropsychiatry*, *4*, 310–314.

Fleiss, J. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York, NY: John Wiley.

Food and Drug Administration. (2011). *FDA executive summary for January 27–28 meeting of the neurological devices panel to discuss the classification of ECT*. Washington, DC: Author.

Fosse, R., & Read, J. (2013). Electroconvulsive treatment: Hypotheses about mechanisms of action. *Frontiers in Psychiatry*, *4*, 94–103. https://doi.org/10.3389/fpsyt.2013.00094

Freeman, C., Basson, J., & Crighton, A. (1978). Double-blind controlled trial of ECT and simulated ECT in depressive illness. *Lancet*, *311*, 738–740. https://doi.org/10.1016/S0140-6736(78)90857-7

Freeman, C., & Kendell, R. (1980). E.C.T, patients' experiences and attitudes. *British Journal of Psychiatry*, *137*, 8–16. https://doi.org/10.1192/bjp.137.1.8

Freeman, W. (1941). Brain-damaging therapeutics. *Diseases of the Nervous System*, *2*, 83.

Friedberg, J. (1976). *Shock treatment is not good for your brain*. San Francisco, CA: Glide.

Gábor, G., & László, T. (2005). The efficacy of ECT treatment in depression: A meta-analysis. *Psychiatria Hungarica*, *20*, 195–200.

Greenhalgh, J., Knight, C., Hind, D., Beverley, C., & Walters, S. (2005). Clinical and cost-effectiveness of electroconvulsive therapy for depressive illness, schizophrenia, catatonia and mania: Systematic reviews and economic modelling studies. *Health Technology Assessment*, *9*, 1–170. https://doi.org/10.3310/hta9090

Gregory, S., Shawcross, C., & Gill, D. (1985). The Nottingham ECT study. *British Journal of Psychiatry*, *146*, 520–524. https://doi.org/10.1192/bjp.146.5.520

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry*, *23*, 56–62. https://doi.org/10.1136/jnnp.23.1.56

Harris, J., & Robin, A. (1960). A controlled trial of phenelzine in depressive reactions. *Journal of Mental Science*, *106*, 1432–1437. https://doi.org/10.1192/bjp.106.445.1432

Higgins, J., Atman, D., Gøtzsche, P., Moher, D., Oxman, A., Savović, J., . . . Sterne, J. (2011). The cochrane collaboration's tool for assessing risk of bias in randomised trials. *British Medical Journal*, *343*, d5928. https://doi.org/10.1136/bmj.d5928

Janicak, P., Davis, J., Gibbons, R., Eriksen, S., Chang, S., & Gallagher, P. (1985). Efficacy of ECT: A meta-analysis. *American Journal of Psychiatry*, *142*, 297–302. https://doi.org/10.1176/ajp.142.3.297

Johnstone, E., Deakin, J., Lawler, P., Frith, C., Stevens, M., McPherson, K., & Crow, T. (1980). The Northwick park ECT trial. *Lancet*, *316*, 1317–1320. https://doi.org/10.1016/S0140-6736(80)92393-4

Johnstone, L. (1999). Adverse psychological effects of ECT. *Journal of Mental Health*, *8*, 69–85. https://doi.org/10.1080/09638239917652

Kho, K., Vreewijk, van., M, Simpson., S, & Zwinderman, A. (2003). A meta-analysis of electroconvulsive therapy in depression. *Journal of ECT*, *19*, 139–147. https://doi.org/10.1097/00124509-200309000-00005

Kingsley, C., & Patel, S. (2017). Patient-reported outcome measures and patient-reported experience measures. *BJA Education*, *17*, 137–144. https://doi.org/10.1093/bjaed/mkw060

Lambourn, J., & Gill, D. (1978). A controlled comparison of simulated and real ECT. *British Journal of Psychiatry*, *133*, 514–519. https://doi.org/10.1192/bjp.133.6.514

Leiknes, K., Jarosh-von Schweder, L., & Hoie, B. (2012). Contemporary use and practice of electroconvulsive therapy worldwide. *Brain and Behaviour*, *2*, 283–344. https://doi.org/10.1002/brb3.37

Lorr, M., Jenkins, R., & Holsopple, J. (1953). Multidimensional scale for rating psychiatric patients. *Veterans Administration Technical Bulletin*, *10*, 507.

Ministry of Health. (2004). *Use of ECT in New Zealand: A review of efficacy, safety and regulatory controls*. Wellington, New Zealand: Author.

Montgomery, S., & Asberg, M. (1979). A new depression scale designed to be sensitive to change. *British Journal of Psychiatry*, *134*, 382–389. https://doi.org/10.1192/bjp.134.4.382

Mosti, C., & Brook, M. (2019). The cognitive effects of electroconvulsive therapy: A critical review. *Psychiatric Annals*, *49*, 152–156. https://doi.org/10.3928/00485713-20190313-02

Mutz, J., Vipulananthan, V., Carter, B., Hurlemann, R., Fu, C., & Young, A. (2019). Comparative efficacy and acceptability of non-surgical brain stimulation for the acute treatment of major depressive episodes in adults: Systematic review and network meta-analysis. *British Medical Journal*, *364*, l1079. https://doi.org/10.1136/bmj.l1079

Myerson, A. (1942). Fatalities following ECT. *Transactions of the American Neurological Association*, *68*, 39.

National Institute of Clinical and Health Excellence. (2009). *Depression in adults: Recognition and management. Clinical guideline [CG90]*. London, England: Author. Retrieved from https://www.nice.org.uk/guidance/cg90/chapter/1-Guidance

O'Leary D, Gill D, Gregory S, Shawcross C. The effectiveness of real versus simulated electroconvulsive therapy in depressed elderly patients. International Journal of Geriatric Psychiatry 1994; 9(7):567-71. https://doi.org/10.1002/gps.930090711

Pagnin, D., Queiroz, de., V., Pini, S, & Cassano, G. (2004). Efficacy of ECT in depression: A meta-analytic review. *Journal of ECT*, *20*, 13–20. https://doi.org/10.1097/00124509-200403000-00004

Pippard, J., & Ellam, L. (1981). *ECT in Great Britain*. London, England: Gaskell.

Rasmussen, K. (2009). Sham electroconvulsive therapy studies in depressive illness: A review of the literature and consideration of the placebo phenomenon in electroconvulsive therapy practice. *Journal of ECT*, *25*, 54–59. https://doi.org/10.1097/YCT.0b013e3181719b23

Read, J., & Arnold, C. (2017). Is electroconvulsive therapy for depression more effective than placebo? A systematic review of studies since 2009. *Ethical Human Psychology and Psychiatry*, *19*, 5–23. https://doi.org/10.1891/1559-4343.19.1.5

Read, J., & Bentall, R. (2010). The effectiveness of electroconvulsive therapy: A literature review. *Epidemiology and Psychiatric Sciences*, *19*, 333–347. https://doi.org/10.1017/S1121189X00000671

Read, J., Bentall, R., Johnstone, L., Fosse, R., & Bracken, P. (2013). Electroconvulsive Therapy. In J. Read & J. Dillon (Eds.), *Models of madness: Psychological, social and biological approaches to psychosis* (2nd ed., pp. 90–104). Hove, UK: Routledge. https://doi.org/10.4324/9780203527160

Read, J., Cunliffe, S., Jauhar, S., & Mcloughlin, D. (2019). Should we stop using electroconvulsive therapy? Maudsley debate. *British Medical Journal*, *364*, k5233. https://doi.org/10.1136/bmj.k5233

Read, J., Harrop, C., Geekie, J., & Renton, J. (2018). An audit of ECT in England 2011–2015: Usage, demographics, and adherence to guidelines and legislation. *Psychology and Psychotherapy: Theory, Research and Practice*, *91*, 263–277. https://doi.org/10.1111/papt.12160

Robertson, H., & Pryor, R. (2006). Memory and cognitive effects of ECT: Informing and assessing patients. *Advances in Psychiatric Treatment*, *12*, 228–238. https://doi.org/10.1192/apt.12.3.228

Rose, D., Wykes, T., Leese, J., Bindmann, J., & Fleischmann, P. (2003). Patients' perspectives of electroconvulsive therapy: Systematic review. *British Medical Journal*, *326*, 1363–1366. https://doi.org/10.1136/bmj.326.7403.1363

Ross, C. (2006). The sham ECT literature: Implications for consent to ECT. *Ethical Human Psychology and Psychiatry*, *8*, 17–28. https://doi.org/10.1891/ehpp.8.1.17

Royal College of Psychiatry. (2017). *Mental health information*. Retrieved from http://www.rcpsych.ac.uk/mentalhealthinformation/therapies/electroconvulsivetherapy,ect.aspx

Sackeim, H., Prudic, J., Fuller, R., Keilp, J., Lavori, P., & Olfson, M. (2007). The cognitive effects of ECT in community settings. *Neuropsychopharmacology*, *32*, 244–254. https://doi.org/10.1038/sj.npp.1301180

Schiele, B., Baker, A., & Hathaway, S. (1943). The minnesota multiphasic personality inventory. *Lancet*, *63*, 292–297.

Schwartzkopff, D. (2018). *DK Law Group announces ECT settlement*. Retrieved from http://ectjustice.com/dk-law-group-announces-ect-settlement

Shiwach, R., Reid, W., & Carmody, T. (2001). An analysis of reported deaths following electroconvulsive therapy in Texas, 1993–1998. *Psychiatric Services*, *52*, 1095–1097. https://doi.org/10.1176/appi.ps.52.8.1095

Shokraneh, F., Sinclair, D., Irving, C. B., & Aali, G. (2019). Electroconvulsive therapy for people with Schizophrenia: Updating a cochrane systematic review for two decades. *BMJ* 2019; 364:k5233. https://www.bmj.com/content/364/bmj.k5233/rr-8

Somatics. (2018). *Regulatory update to Thymatron system IV instruction manual*. Retrieved from http://www.thymatron.com/downloads/System_IV_Regulatory_Update.pdf

Tecoult, E., & Nathan, N. (2001). Morbidity in ECT. *European Journal of Anaesthesiology*, *18*, 511–518. https://doi.org/10.1097/00003643-200108000-00005

Tørring, N., Sanghani, S., Petrides, G., Kellner, C., & Østergaard, S. D. (2017). The mortality rate of electroconvulsive therapy: A systematic review and pooled analysis. *Acta Psychiatrica Scandinavica*, *135*, 388–397. https://doi.org/10.1111/acps.12721

UK ECT Review Group. (2003). Efficacy and safety of ECT in depressive disorders. *Lancet*, *361*, 799–808. https://doi.org/10.1016/S0140-6736(03)12705-5

Ulett, G., Smith, K., & Gleser, G. (1956). Evaluation of convulsive and subconvulsive shock therapies utilizing a control group. *American Journal of Psychiatry*, *112*, 795–802. https://doi.org/10.1176/ajp.112.10.795

van der, Wurff., F, Stek., M, Hooogendijk., W, & Beekman, A. (2003). Electroconvulsive therapy for the depressed elderly. *Cochrane Database of Systematic Reviews*, *2003*(2), CD003593.

West, E. (1981). Electric convulsion therapy in depression: A double-blind controlled trial. *British Medical Journal*, *282*, 355–357. https://doi.org/10.1136/bmj.282.6261.355

Wilson, I., Vernon, J., Guin, T., & Sandifer, M. (1963). A controlled study of treatments of depression. *Journal of Neuropsychiatry*, *4*, 331–337.

Disclosure. The authors have no relevant financial interest or affiliations with any commercial interests related to the subjects discussed within this article.

Funding. The author(s) received no specific grant or financial support for the research, authorship, and/or publication of this article.

Correspondence regarding this article should be directed to John Read, PhD, University of East London, School of Psychology, London E15 4LZ, UK. E-mail: john@uel.ac.uk