Economic evaluation of outcomes from the Acorn program

Analysis commissioned by Hopscotch Foundation

March 2022

Dr Tim Schultz[1] and Dr Paul Aylward[2]

[1] Senior Research Fellow, Flinders Health and Medical Research Foundation, Flinders University

[2] Action Research Partnerships

Suggested citation:

Schultz, T. J. and P. Aylward (2022). Economic evaluation of outcomes from the Acorn program. Adelaide, Flinders University.

Contents

List of Figures

List of Tables

Executive summary

The Acorn program funded by the Hopscotch Foundation has been delivered in South Australia over a 6 year period to over 400 mothers of young children (up to 3 years of age) who are experiencing mental health challenges that impact on the quality of their parenting interactions (Aylward 2022). Multidisciplinary teams of professionals including clinicians, family support practitioners, and dance workers deliver and support activities of Dance Play and Reflective Journaling and provide a weekly Therapeutic Letter sent out to mothers between Acorn sessions. A range of benefits for mothers and children have been identified through on-going action research evaluation conducted in tandem with the program, including improvements in maternal depression scores after attending Acorn program which was employed as an indicator of client wellbeing. Program costs and benefits have not previously been quantified. This study provides an economic evaluation of the Acorn program focussing on depression as one key indicator of improved wellbeing for participating mothers.

Two methods were tested. Method 1 was a cost benefit approach that estimated the reduced number of women with depression following participation in Acorn, and the societal savings attributed to maternal and infant outcomes from reduced maternal depression. Method 2 was a cost effectiveness method that relied on conversion of depression (PHQ-9 scores) into Quality Adjusted Life Years gained. Both methods relied on estimates of the costs to deliver Acorn, and an estimate of what would have happened to Acorn participants if they had not attended the program, which was inferred from results of seven wait-list control studies of post-natal depression.

For Method 1, an estimate of reliable and clinically significant improvement in depression (Jacobson and Truax 1991, McMillan, Gilbody et al. 2010) was made for both Acorn participants and wait-list control studies. For Acorn participants, two estimates provide an upper and lower bound for the % improvement in depression: 26.3% of eligible mothers (Kroenke standard definition), to 30.0% of eligible mothers (RCSC c criteria). The median estimate for clinically significant improvement in wait-list control studies was 18.3%, indicating:

- a lower estimate of Acorn effect (8.0%, calculated as 26.3% – 18.3%)
- an upper estimate of Acorn effect (11.7%, calculated as 30% – 18.3%)

Across 190 eligible Acorn participants (ie those commencing Acorn with at least moderate depression) this equates to a **reduction of 15.2 to 22.2 mothers with clinical depression**.

The cost to Australian society of an untreated case of postnatal depression is estimated to range from AU$41,367 at the lower range, and AU$62,051 at the upper range. Using the lower estimate of effect (8%), the projected benefits from Acorn were calculated as $628,778 (using the lower range of costs to society) to $943,175 (using the upper range of cost to society). For the upper range of Acorn effectiveness (11.7%), the projected benefits from Acorn were calculated as $918,347 (using the lower range of costs to society) to $1,377,532 (using the upper range of cost to society). Based on a program cost of $2,756,098, it is estimated by Method 1 that between 22.8% and 50.1% of the costs of Acorn are realised as a societal benefit from reduced postnatal depression.

For Method 2, the greatest impact on health utility (Quality Adjusted Life Years) was for Acorn mothers with severe depression. However, differences in the mental health status of mothers in wait-list control studies and the Acorn program impacted the validity of this approach, and it was not possible to estimate cost effectiveness.

There are a number of assumptions and limitations resulting from the use of wait-list controls, the conversion of symptom scores into utility scores, and estimated costs of untreated post-natal depression in Australia that impact on the strength of these findings.

1 Introduction – Acorn program

The Acorn program funded by the Hopscotch Foundation has been delivered face-to-face in South Australia for over a six year period[1]. The program establishes 'Acorn groups' to help mothers of young children (up to 3 years of age) who are experiencing mental health challenges that impact on quality of their parenting interactions with their children (Aylward 2022). Multidisciplinary teams of professionals including clinicians, family support practitioners, and dance workers deliver and support activities of Dance Play and Reflective Journaling and provide a weekly Therapeutic Letter sent out to mothers between Acorn sessions. Since inception over 400 individual mothers and children have attended at least one Acorn group, with 353 mothers completing the program (Aylward 2022). The Acorn program aimed 'To holistically nurture and enhance parental wellbeing and the quality of the parent/child relationship for mothers experiencing identified mental health illnesses and their young children aged 0-36 months'. In addition to enhancing the quality of the parent-child interaction and building parental confidence, competence and social connectivity, an objective of the program was to enhance the mother's wellbeing, coping, resilience and self-efficacy. As many Acorn mothers were experiencing depression on commencing the program, improvements in depression were selected as one indicator of enhanced wellbeing.

A recent review of Acorn concluded that the "program has clearly nurtured the quality of the parent-child interaction for both mothers with these mental health issues and their young children with enduring positive outcomes for both which has enhanced their relationship" (p. 9) (Aylward 2020). There were significant improvements in maternal depression scores measured before and after the Acorn program, for example, the number of mothers reporting major depression decreased from 97 (28.2%) before the program, to 60 (17.4%) after the program (Aylward 2022). However, the participatory action research model adopted to evaluate this community based program and enable ongoing improvement to the intervention did not include a control group and an economic evaluation was not included in the Acorn review. Therefore, this study provides an economic evaluation of the Acorn program focussing on depression as one key indicator of improved wellbeing for participating mothers and considering wait-list controls of comparable studies focussing on post-natal depression.

2 Background

2.1 Perinatal mental health

Mental health problems are common in pregnant women and new mothers in the first year post-partum (the perinatal period). Perinatal mental illness occurs in 10-20% of women, and is associated with increased morbidity and maternal death in high income countries (Sambrook Smith, Lawrence et al. 2019). Perinatal mental illness may also compromise the healthy emotional, cognitive and physical development of the child, with potential serious long term consequences (Stein, Pearson et al. 2014, Rees, Channon et al. 2019). While postnatal depression (PND) has traditionally been the focus of much research, it is highly correlated with antenatal depression (Underwood, Waldie et al. 2016). In addition to depression, other mental health conditions that may occur during the post-natal period include anxiety, psychosis (such as bipolar disorder and schizophrenia), borderline personality disorder, and post-traumatic stress disorder. The majority of people suffering from depression have at least one comorbid mental disorder and the prevalence of mental health comorbidity has been shown to increase with severity of depression (Steffen, Nübel et al. 2020).

[1] An online version was developed more recently but this report addressed the face-to-face model of the program.

A recent systematic review estimated the mean prevalence of PND in the community was 13.1% (Underwood, Waldie et al. 2016). Risk factors for PND include previous depression prior to or during pregnancy, anxiety during pregnancy, experiencing stressful life events during pregnancy or the early puerperium, low levels of social support or partner support, low socioeconomic status, and obstetric complications (Fitelson, Kim et al. 2010). The adverse effects of PND are well established and include detrimental social, emotional cognitive and behavioural outcomes of the infant with potential longer term consequences on child development (Hall 2012). Impaired mother-child bonding, increased odds of developmental delay at 18 months of age, behavioural/emotional problems in children and reduced adolescent IQ have all been associated with PND (Poobalan, Aucott et al. 2007, Hall 2012). The impacts of PND on mothers and other family members is also significant, particularly in light of societal expectations of motherhood as a uniquely joyful experience, including disconnection from the mothering process, risks of substance abuse, self harm and suicide (Fitelson, Kim et al. 2010, Hall 2012).

Postnatal depression is underdiagnosed, with the mean diagnosis rate estimated at 30.8% (95% CI, 30.2%-31.4%) (Cox, Sowa et al. 2016). Treatment only occurs in about 16% of cases of PND (95% CI, 14.8%-16.9%), and only 6.3% of women with PND receive an adequate trial of treatment (Cox, Sowa et al. 2016). Therefore, there is a clear need for effective treatment programs for PND.

Treatment programs for PND include pharmacological, psychotherapeutic (such as interpersonal therapy, cognitive-behavioral therapy, counselling, and psychodynamic psychotherapy, as well as psychosocial interventions) and non-pharmacologic interventions (Fitelson, Kim et al. 2010). Pharmacologic interventions alone have a mean remission rate of 49.8% (95% CI, 49.0-50.6%) (Cox, Sowa et al. 2016); however, maternal concerns about infant exposure to medication through breast milk limits its utility in practice (O'Hara, Stuart et al. 2000, Fitelson, Kim et al. 2010). Psychotherapy interventions alone have similar effects to pharmacologic interventions and are effective treatment options for women suffering from postpartum depression (Dennis and Hodnett 2007, Huang, Zhao et al. 2018), with a mean remission rate of 51.2% (95% CI, 49.1%-53.3%) (Cox, Sowa et al. 2016). Effects of psychological interventions are stronger immediately after treatment but are also effective at reducing maternal symptomology at 6 months' follow-up (Stephens, Ford et al. 2016).

2.2 Economic costs

The estimated two-generational annual economic cost of not treating one mother with peripartum depression is US$22,647 (Diaz and Chase 2010). The cost of not treating the mother is $7,211; while the cost attributable to a child born to a depressed mother reaches $15,323. While these costs are attributable to mothers depressed at the time of birth, they provide an indication of the potential cost savings from improved management of PND.

A UK-based study estimated the total long term costs to society of three perinatal mental health conditions (depression, anxiety and psychosis) as just under £10,000 per birth for every live birth in the UK (Bauer, Parsonage et al. 2014). Nearly three quarters (72%) of these costs related to adverse impacts on the child, rather than the mother. The average costs per case of PND was around £74,000, of which £23,000 relates to the mother and £51,000 relates to impacts on the child (Appendix 3). The costs per case of perinatal anxiety (on its own, not co-morbid with depression) was similar for mothers (£21,000) but less for the child (£14,000) (Bauer, Parsonage et al. 2014). It was estimated that an additional £400 spent per birth on enhancing perinatal mental health in the National Health Service (NHS) could save the NHS around £2,100 from a total societal cost of around £10,000 per birth.

Psycho-social interventions for PND are cost-effective at funding levels used by the UK NICE (Morrell, Warner et al. 2009, Henderson, Dixon et al. 2019). However, apart from the PoNDER study (Morrell,

Warner et al. 2009, Henderson, Dixon et al. 2019), there are few published economic analyses of psychosocial interventions for PND (Morrell, Sutcliffe et al. 2016).

2.3 Study aim

The aim of this study was to conduct a health economic evaluation of Acorn and its impact on PND. Two methods were tested. Method 1 was a cost benefit approach that estimated the reduced number of women with depression following participation in Acorn, and the societal savings attributed to maternal and infant outcomes from reduced maternal depression. Method 2 was a cost effectiveness method using a conversion of PHQ-9 scores into QALYs gained. Both methods used estimates of the direct and indirect program costs to deliver Acorn. Due to the lack of control data, both methods also relied on an estimate of what would have happened to Acorn participants if they had not attended the program, in other words if there was a control group not exposed to Acorn against which the Acorn results could be compared.

3 Methods

3.1 Measurement of depression

Data collected from the Acorn evaluation was used to inform a health economic evaluation of improvements in maternal depression, one of its key indicators of client wellbeing. The Patient Health Questionnaire [PHQ-9] is a 9-item screen for mood disorders that is widely used in primary care (Kroenke, Spitzer et al. 2001, Hanusa, Scholle et al. 2008). Each item is rated on how frequently each symptom has occurred over the past 2 weeks and has four possible responses: not at all (scored as 0), several days (1), more than half the days (2), and nearly every day (3). A total score (out of 27) of:

- 5-9 indicates 'minimal symptoms' (referred to as 'mild depression' in this report)
- 10-14 indicates 'minor depression / major depression mild' (referred to as 'moderate depression')
- 15-19 indicates 'major depression moderately severe' (referred to as 'moderately severe depression'
- 20 or more indicates 'major depression severe' (referred to as 'severe depression')

3.2 Wait-list control group

While the Acorn program did not include a control group, a number of randomised controlled trials of psycho-social interventions for PND have previously used wait-list controls (Meager and Milgrom 1996, O'Hara, Stuart et al. 2000, Clark, Tluczek et al. 2003, Clark, Tluczek et al. 2008). For example, a mother-infant psychotherapy group and interpersonal psychotherapy were shown to be superior to a waiting-list comparison group in reducing maternal depressive symptoms (O'Hara, Stuart et al. 2000). Following a search of Pubmed and of the reference lists of recently conducted systematic reviews, a dataset from wait-list control studies was extracted. This data from wait-list control groups of relevant published evaluations was used to estimate a background change in depressive symptoms for mothers were they not receiving the Acorn intervention, while acknowledging a number of limitations in this approach (see section 4.6)

3.3 Acorn program costs

Costs of delivery of Acorn were provided by Anglicare, the Acorn service provider. Annual costs included:

- Acorn support team / multi-disciplinary team staffing
- Management and coordination
- Facilities and materials
- Indirect costs.

In addition to actual costs to deliver the program during its development since 2015, a modified cost to future delivery based on expected efficiencies have been used to estimate costs of a future Acorn program.

3.4 Estimates of effect of Acorn

There were two methods used to estimate the effect of Acorn on PND to inform an economic analysis.

3.4.1 Method 1 – Clinically significant change in PND

"Clinically significant change" is defined in psychotherapy as the extent to which therapy moves someone outside the range of the dysfunctional population or within the range of the functional population (Jacobson, Follette et al. 1984, Jacobson and Truax 1991). These authors have also proposed a reliable change (RC) index to determine whether the magnitude for a given client is statistically reliable. These methods have become extremely influential in defining treatment outcome in studies of depression. The original validation study for the PHQ-9 identified a clinically significant change (ie an improvement from a clinical (or dysfunctional population) score to a non-clinical (functional population)) as a post-treatment score of ≤ 9 combined with an improvement of 50% or more (Kroenke, Spitzer et al. 2001). This has been shown to have acceptable agreement with a gold standard diagnostic interview (McMillan, Gilbody et al. 2010), however, it does not account for deterioration of depression status. Calculation of the reliable change index, however, allows for estimation of both clinically significant improvement and deterioration (Jacobson and Truax 1991, McMillan, Gilbody et al. 2010). The reliable change index reports the number of individuals who experience a reliable change over the course of an intervention while accounting for the reliability of the instrument used to assess change (Jacobson and Truax 1991). It determines cut-off points based on the central location and distribution of scores for a clinical and non-clinical group, incorporating the test-retest reliability (Chronbach's alpha) of the tool.

Reliable change indices allow the measurement of number of participants whose degree of change across therapy represents a clinically significant level (i.e. one that would be clearly recognised by clinicians) (Jacobson and Truax 1991, Brock, O'Hara et al. 2017). The reliable change index for the PHQ-9 has been validated against the gold standard diagnostic interview in a study in which the original criteria proposed by Kroenke (a post-treatment score of ≤ 9 combined with an improvement of 50% or more) was also shown to be acceptable (McMillan, Gilbody et al. 2010). This study compared two definitions of clinically significant change: the reliable change index (RCSC criterion c from (McMillan, Gilbody et al. 2010) and Kroenke's standard definition (Kroenke, Spitzer et al. 2001).The operational definitions for improvement for the two measures are defined as:

- Reliable change index (RCSC criterion c from (McMillan, Gilbody et al. 2010)): a clinically significant change as defined by the reliable change index (Jacobson and Truax 1991) AND the pre-treatment score must be 10 or more PHQ-9 units, post-treatment score must be 9 or less and the improvement in score must be 5 or more.
- Kroenke standard definition: clinically significant change is a pre-treatment score must be 10 or more, post-treatment score must be 9 or less AND the improvement in score must be at least 50% of pre-treatment score. For example, for a pre-Acorn measure of 16 on the PHQ, a score of 8 or lower post-Acorn would constitute a clinically significant change.

The kappa for the reliable change index was 0.62, and for the Kroenke standard definition was 0.58 (McMillan, Gilbody et al. 2010).

The number of reliable and clinically significant improvements in the PHQ-9 was estimated to generate a rate of improvement (ie the number who improved reliably and clinically significantly divided by the eligible number) for both sets of criteria (RCSC criterion c and Kroenke), which was compared against data

for wait-list control groups from published studies. Excess cases of improvement (ie assuming that the rate of improvement is greater in the Acorn group than the wait-list control group) were combined with the cost savings to society per case of prevented PND. Estimates from three studies were used to derive the cost savings to society per case of prevented PND (Diaz and Chase 2010, Bauer, Parsonage et al. 2014, PwC Consulting Australia 2019). Greatest weighting was applied to the most recent, Australian study of the societal costs of perinatal depression (PwC Consulting Australia 2019).

3.4.2 Method 2 - QALYs

A recent study has generated algorithms to convert PHQ-9 scores into Euro-Qol five dimensions (EQ-5D) scores (Furukawa, Levine et al. 2021). The EQ-5D is the generic, preference-based measure of health that permits evaluation of the quality adjusted life year (QALY), which are standard measures for impact of disease on the quality and quantity of life. One QALY is equal to one year of life in perfect health, ranging down to 0 (death) and below zero (worse than death). The QALY is used to measure the burden of disease, impact of interventions and in setting priorities in resource allocation (Whitehead and Ali 2010). If the costs of an intervention and its impacts on QALYs (ie effectiveness) are known, it is possible to estimate its cost effectiveness (ie the additional cost to achieve an incremental health benefit) against a counterfactual (ie control group).

PHQ-9 scores were converted into Euro-Qol five dimensions (EQ-5D) scores using recently developed algorithms derived from seven randomised controlled trials of internet cognitive-behavioural therapies (iCBT) for depression (Furukawa, Levine et al. 2021). These relate a change in PHQ-9 score (from pre- to post-Acorn program) to a concomitant change in EQ-5D score (Figure 1). Separate algorithms (curves) are provided for both baseline (pre-intervention) and endpoint (post-intervention) time periods. This result is explained as understandable “because a patient would rate their health status less satisfactory if they stayed equally symptomatic as before after the treatment and also because it means that they continued to suffer from depression for longer” (p. 99). In other words, when a patient experiences the same level of depression (PHQ-9) after an intervention as before, they rate their quality of life lower. Therefore, to demonstrate a substantial improvement in quality of life (as shown by an increase in the EQ-5D) from baseline to endpoint, the PHQ-9 measure would need to decrease by at least 3-4 points to account for this effect. To demonstrate an effect on quality of life from an intervention, this decrease in PHQ-9 in the intervention group would need to be in excess of any changes to a control group (Furukawa, Levine et al. 2021). Calculation of changes in the PHQ-9/EQ-5D over time between intervention and control groups allows for an estimate of the gain in QALY per year due to the intervention.

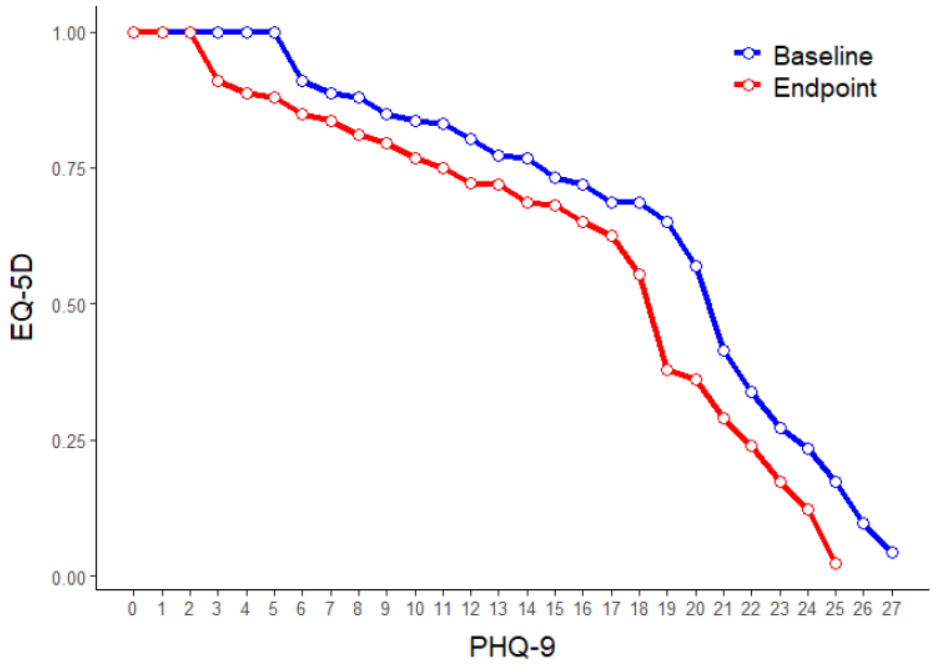

Figure 1 PHQ-9 and EQ-5D total scores calculated at baseline (pre) and endpoint (post). EQ-5D, Euro-Qol Five Dimensions; PHQ-9, Patient Health Questionnaire-9 (Furukawa, Levine et al. 2021).

The conversion from PHQ-9 to EQ-5D is based on seven RCTs of internet cognitive behaviour therapies (n=2457) administering validated depression scales (three studies used PHQ-9, three studies used Beck Depression Inventory and one used the CES-D) and the EQ-5D both at baseline and endpoint (Furukawa, Levine et al. 2021). BDI and CES-D scores were converted into the PHQ-9 scores. For included studies contributing to the conversion estimates, the mean age was 41.8 years, 66% were female and they scored 14.0 and 0.74 on PHQ-9 and EQ-5D, respectively, at baseline, and 9.1 and 0.79, respectively, at endpoint. At baseline, based on the PHQ-9: 2.4% had no depression, 20.2% had mild depression (5-9), 33.45% had moderate depression (10-14), 26.5% had moderately severe depression (15-19), 17.3% had severe depression (20+).

3.5 Ethics

The existing ethics approval (Bellberry Human Research Ethics Committee, 2014-09-509-PRE-6) was updated. Only de-identified data was used in this study.

4 Results

4.1 Program participants

There was a total of 493 mothers involved in one of 11 Acorn waves since 2015. This analysis is based on the 344 mothers who completed both a PHQ_{pre} and $PHQ\text{-}9_{post}$; 142 completed PHQ_{pre} only, three completed $PHQ\text{-}9_{post}$ only and four completed neither. The characteristics of the participants are described in Table 1, and the flow of patients through the Acorn program and their initial depression status (as measured by the $PHQ\text{-}9_{pre}$) is presented in Figure 1.

Table 1 Characteristics of Acorn participants (n=344) who completed both a PHQpre and PHQ-9post

Characteristic	Category	n	%
How many of the group sessions have you attended?	All	44	13.0
	Most	257	76.0
	Half	35	10.4
	Less than half	2	0.6
Depression status (pre-ACORN)	Non	45	13.1
	Mild	109	31.7
	Moderate	93	27.0
	Moderate-severe	60	17.4
	Severe	37	10.8
Parent age group	<30	135	39.5
	30 or older	207	60.5
Child age groups	0-5 months	140	40.7
	6-11 months	128	37.2
	12-17 months	47	13.7
	18+ months	29	8.4
Other children	Yes	113	32.8
	No	86	25.0
	Unknown	145	42.2
Place of birth	Australia	298	87.4
	Other	43	12.6
Speak language other than English	Yes	55	16.0
	No	288	84.0
Aboriginal or Torres Strait Islander	Yes	7	2.0
	No	335	98.0
Education	School including Yr 12	87	25.4
	Cert Dip Tafe or College	133	38.8
	Uni degree	116	33.8
	Other	7	2.0
Employment	F/T parent or on leave from work	198	62.1
	Currently working	62	19.4
	Not working but may be studying	59	18.5
Income	Wages earned by you or partner	233	67.9
	Gov Benefit, pension, allowance	110	32.1
Marital status	Married or defacto	265	77.7
	Single or separated/divorced or other	76	22.3

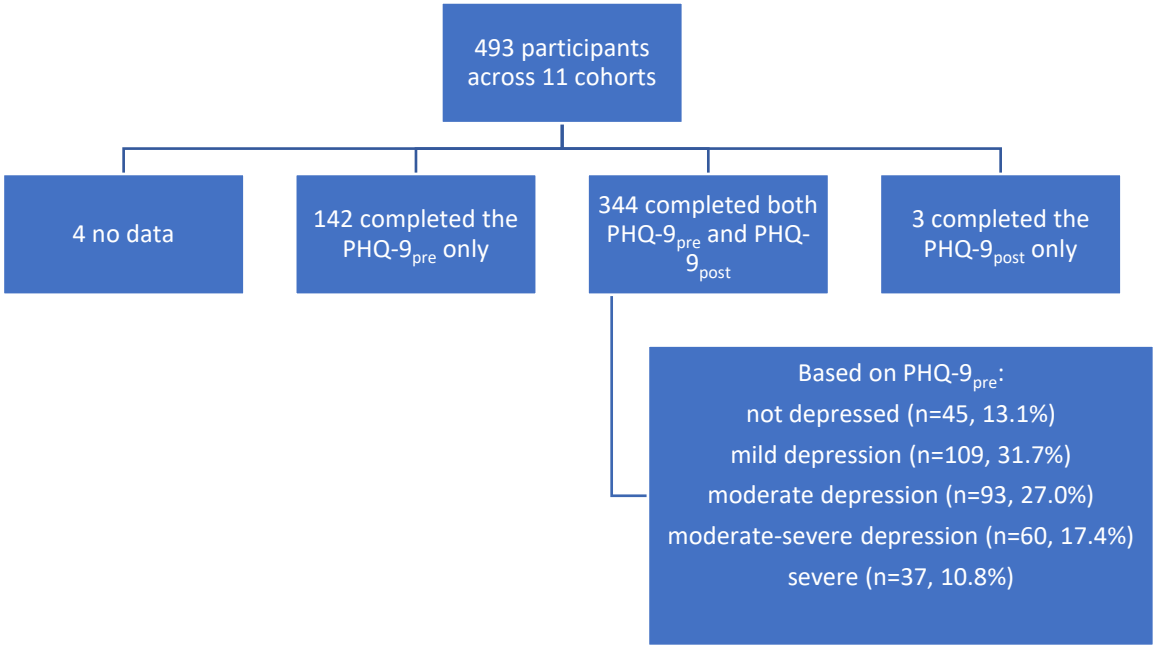

Figure 2 Flow of patients through the Acorn program, including their initial depression status (as measured by the PHQ-9_{pre})

4.2 Estimates of a control effect

To try and estimate what may have happened to mothers in the Acorn study in the absence of an intervention, data from wait-list control studies were used to estimate reductions in depression levels that may have otherwise occurred in the background without any formal intervention.

4.2.1 Wait-list control studies

A search of Pubmed and the included studies of recent systematic reviews (Stephens, Ford et al. 2016, Cook, Ayers et al. 2018, Huang, Zhao et al. 2018, Carter, Bastounis et al. 2019, Rees, Channon et al. 2019, Sambrook Smith, Lawrence et al. 2019, Sinesi, Maxwell et al. 2019) identified nine studies evaluating the effectiveness of interventions for postpartum depression using a wait-list control (Meager and Milgrom 1996, O'Hara, Stuart et al. 2000, Clark, Tluczek et al. 2003, Clark, Tluczek et al. 2008, Buttner, Brock et al. 2015, Segre, Brock et al. 2015, Fonseca, Monteiro et al. 2019, Fonseca, Alves et al. 2020, Van Lieshout, Layton et al. 2021). The characteristics of these studies are summarised in Table 9. One study was not available in full text format (Meager and Milgrom 1996). Studies were conducted on women with post-partum depression, testing interventions such as cognitive behavioural therapy, mother-infant therapy groups, yoga, and listening visits, over the course of 8-12 weeks. Total sample size ranged from 20 (Meager and Milgrom 1996) to 403 (Van Lieshout, Layton et al. 2021). The most common outcomes that were measured were the Beck Depression Inventory (BDI-II, n=4 studies), Edinburgh Postnatal Depression Scale (EPDS, n=4 studies) and the Hamilton Rating Scale for Depression (HRSD, n=3 studies). No studies measured the PHQ-9. Two studies used a sequence to allocate patients to either of the intervention or wait-list control groups, all other studies were randomised.

Outcomes from these studies were reported in two ways: either as means and standard deviation pre- and post-interventions for continuous outcomes, such as the HRSD; or as a dichotomised outcome, sometimes articulated as a "clinically significant change" (e.g. as specified above, a pre-treatments score of at least 10, a post-treatment score of ≤ 9 combined with an improvement of 50% or more (Kroenke,

Spitzer et al. 2001)), or as “reliable change indices”, the number of individuals who experience a reliable change over the course of an intervention (Jacobson and Truax 1991).

Three studies do not report data as a dichotomous outcome (Clark, Tluczek et al. 2003, Clark, Tluczek et al. 2008, Fonseca, Monteiro et al. 2019). Five studies reported reliable change indices indicating a clinically significant improvement for intervention and control groups using a range of outcome measures, including the HRSD, BDI-II, IDAS-GD, EPDS, and GAD-7; although it was not possible to extract findings from one study that only reported odds ratios, numbers needed to treat and risk differences (Van Lieshout, Layton et al. 2021). Clinically significant results ranged from 31.7-78.2% of participants in the intervention group and 14-59.2% for wait-list controls (Table 2). Calculation of a mean % clinical significant improvement in the wait-list control studies across four outcome measures provides the following estimates: BDI-II (1 study) 18.3%, EPDS (2 studies) 30.7%, HRSD (3 studies) 29.4%, IDAS-GD (1 study) 29%. One study had a very high rate of clinically significant improvement: 78.2% in the intervention and 59.2% in the control group (Buttner, Brock et al. 2015). This study utilised five measurement time periods (pre, week 2, week 4, week 6 and post) that may have contributed to the effect. Given this outlier, it is most appropriate to use the median estimate of 18.3% (interquartile range 16.7 to 36.0) for clinically significant improvement in wait-list control studies across the different outcome measures, noting that this includes some studies on more than one occasion. Additionally, it should be noted that relatively few of these studies provided detail about how the clinically significant change was calculated.

Table 2 Clinically significant results from five controlled studies between pre- and post-measures for intervention and wait-list control groups.

		Intervention		Control	
Outcome	**Citation**	**n**	**%**	**n**	**%**
HRSD	(O'Hara, Stuart et al. 2000)	51	31.7	51	15.0
BDI-II	(O'Hara, Stuart et al. 2000)	51	38.3	51	18.3
HRSD	(Buttner, Brock et al. 2015)	23	78.2	27	59.2
HRSD	(Segre, Brock et al. 2015)	21	36[@]	21	14[@]
IDAS-GD	(Segre, Brock et al. 2015)	21	69	21	29
EPDS	(Segre, Brock et al. 2015)	21	64	21	43
EPDS	(Fonseca, Alves et al. 2020)[$]	64	42.2	82	18.3
EPDS and GAD-7	(Van Lieshout, Layton et al. 2021)	201	[&]	202	[&]

HRSD Hamilton Rating Scale for Depression
BDI-II Beck Depression Inventory II
IDAS-GD Inventory of Depression and Anxiety Symptoms
EPDS Edinburgh Postnatal Depression Scale
GAD-7 Generalised Anxiety Disorder Questionnaire
@ 5% of women in the WLC group experienced an increase in symptoms and surpassed the RCI threshold; none of the women in the treatment group deteriorated over time
$ recovery rates defined as “clinical levels of symptoms at T1 and absence of clinical symptoms at T2”
& - reported odds ratios, number needed to treat and risk difference but cannot derive proportions: eg EPDS indicated greater odds of exhibiting a clinically significant change in scores relative to control participants (odds ratio [OR], 4.15; 95%CI, 2.66-6.46, number needed to treat, 2.9; risk difference, 33.9%).

4.3 Acorn Program costs

The total direct cost per group was $28,887 and the indirect costs per group was $4,319 for a total cost of $33,206 per group. There were 83 groups trained across 11 cohorts, therefore the total cost to deliver Acorn was $2,756,098. The cost per Acorn mother (based on 493 mothers) is $5,590.

Discussions with Acorn team members have indicated that some of the indirect costs (eg general manager oversight, and audit and evaluation) may be removed in a sustainable delivery of the program in the future, and that the 'Management and coordination' component may be reduced given other efficiencies. This would reduce the direct cost to $24,863 and indirect cost to $3,019, for a total of $27,882 per group, or $2,314,206 for a total of 493 mothers. This would cost $4,694 per Acorn mother in a future program delivery.

4.4 Method 1 – Clinically significant change in Acorn cohorts

Comparison of depression status (as determined by PHQ-9) before and after the Acorn program identified that at the start of the program the largest number of mothers (n=109) had mild depression, followed by moderate (n=93), moderate-severe (n=60), none (n=45) and severe (n=37). After the program the largest number of mothers had mild depression (n=122), followed by non (n=90), moderate (n=72), moderate-severe (n=39) and severe (n=21) (Table 3). Comparing pre- and post-program scores for the same mothers indicated that the largest number (n=48) had mild depression pre- and post-program, with the next largest groups moving from mild to non (n=38) and from moderate to mild (n=34). 61 of 299 (20.4%) mothers who were symptomatic initially became asymptomatic after Acorn.

Table 3 Initial and final depression status for 343 Acorn participants.

Final status	Non		Mild		Moderate		Moderate-severe		Severe		Total	
Initial status	n	%	n	%	n	%	n	%	n	%	n	%
Non	28	63.6	16	36.4	0	0.0	0	0.0	0	0.0	44	100
Mild	38	34.9	48	44.0	18	16.5	5	4.6	0	0.0	109	100
Moderate	15	16.1	34	36.6	25	26.9	11	11.8	8	8.6	93	100
Moderate-severe	6	10.0	21	35.0	18	30.0	12	20.0	3	5.0	60	100
Severe	2	5.4	3	8.1	11	29.7	11	29.7	10	27.0	37	100
Total	89	25.9	122	35.6	72	21.0	39	11.4	21	6.1	343	100

4.4.1 Reliable and clinically significant change

The pre- and post-Acorn PHQ-9 measures were compared to determine whether there was a 'clinically significant change' in the PHQ for each mother.

There were 190 mothers who were eligible for calculation of the RCSC c and Kroenke standard definition for clinically significant change (ie their initial PHQ-9 was 10 or more) and 154 mothers with no depression (4 or less on PHQ-9) or mild depression (9 or less on PHQ-9). Using the RCSC c criteria, 57 mothers (30% of eligible) experienced a clinically significant change, and 133 (70% did not experience an improvement (Table 4). Using the Kroenke standard definition, 50 mothers (26.3% of eligible) experienced a clinically significant change, whereas 140 (73.7%) did not (Table 5).

Table 4 Determination of reliable and clinically significant change for 343 Acorn participants based on the RCSC criterion c (McMillan, Gilbody et al. 2010)

Clinically significant change	Ineligible		Improvement		No improvement		Total	
Initial status	n	%	n	%	n	%	n	%
Non	44	100	-		-		44	100
Mild	109	100	-		-		109	100
Moderate	-		25	26.9	68	73.1	93	100
Moderate-severe	-		27	45.0	33	55.0	60	100
Severe	-		5	13.5	32	86.5	37	100
Total	153	44.6	57	16.6	133	38.8	343	100

Table 5 Determination of reliable and clinically significant change for 343 Acorn participants based on based on the Kroenke criteria

Clinically significant change	Ineligible		Improvement		No improvement		Total	
Initial status	n	%	n	%	n	%	n	%
Non	44	100	-		-		44	100
Mild	109	100	-		-		109	100
Moderate	-		22	23.7	71	76.3	93	100
Moderate-severe	-		23	38.3	37	61.7	60	100
Severe	-		5	13.5	32	86.5	37	100
Total	153	44.6	50	14.6	140	40.8	343	100

These two estimates provide an upper and lower bound for the % improvement in depression from mothers attending Acorn. This ranged from:

- 30.0% of eligible mothers (RCSC c criteria).
- 26.3% of eligible mothers (Kroenke standard definition), to

4.4.2 Effect of Acorn on maternal depression

Based on estimates of the uncontrolled effects of Acorn on the intervention group (range: 26.3-30.0% of eligible mothers) and a wait-list control group (18.3%), the overall effect of Acorn on maternal depression is estimated as ranging from 8.0% (for Kroenke standard definition) to 11.7% (for RCSC c criteria) of eligible mothers experienced reliable and clinically significant reduced depression (Equation 1).

Equation 1

$$Acorn\ effect\ (\%\ mothers\ with\ reduced\ depression) = \%\ Acorn\ effect\ - \%\ waitlist\ control\ effect$$

Use of the Kroenke standard definition and the RCSC c criteria provides a lower and upper limits, respectively, for the Acorn effect.

Lower Acorn effect (% mothers experiencing less depression according to Kroenke standard definition) = 26.3% – 18.3%=8.0%

Upper Acorn effect (% mothers experiencing less depression RCSC c criteria) = 30.0% – 18.3%=11.7%

Across the 190 eligible Acorn participants, this equates to a **reduction of 15.2 to 22.2 mothers with depression**.

4.4.3 Costs per case of PND

US, UK and Australian studies have reported costs from perinatal depression and anxiety but have not separated costs into antenatal and postnatal time periods. The estimated costs for perinatal depression, adjusted for inflation and currency, range from AU$44,270 in the US study, at AU$82,734 in the Australian study, and AU$171,600 in the UK study:

- The estimated two-generational annual economic cost to society of not treating one mother with peripartum depression is US$22,647 (Diaz and Chase 2010). About one thirds (31.8%) of the costs are attributed to poorer maternal outcomes, and the remaining 68% are attributed to poorer outcomes for their babies. Adjusted for 3% inflation and converted into Australian dollars (1USD=1.41 AUD), this is equivalent to AU$44,270 in 2021 terms.
- The cost of untreated perinatal depression in the UK is £74,000, of which £23,000 (31.1%) relates to the mother and £51,000 (68.9%) relates to impacts on the child (Bauer, Parsonage et al. 2014) (Appendix 3). Adjusted for 3% inflation and converted into Australian dollars (1UK£=AU$1.88), this is equivalent to AU$171,600 in 2021 terms.
- A recent Australian study estimated that perinatal anxiety and depression costs AU$121,667 across the family unit, including maternal, paternal and child impacts (PwC Consulting Australia 2019) (Appendix 4). The UK study estimated that the costs of depression (alone) were 68% of the total costs of depression and anxiety combined (Bauer, Parsonage et al. 2014). On this basis, the estimated cost to society of perinatal depression is estimated at AU$82,734.

Based on its recency, specificity to Australian healthcare system and as a mid-point for the three national estimates, we will use the Australian estimate of AU$82,734 per untreated case of perinatal depression. It is difficult to estimate the breakdown of these costs between antenatal and postnatal depression; however, given the extended duration of postnatal depression and the proximity of the effects of postnatal depression on the infant, which accounts for some 68% of the total costs of perinatal depression (Diaz and Chase 2010, Bauer, Parsonage et al. 2014), it seems reasonable to assume that postnatal depression accounts for at least half of the total costs of perinatal depression, and potentially as much as three-quarters of the total costs. These two estimates will be used in subsequent analyses as lower and upper ranges for the cost of postnatal depression. Therefore, the estimated cost to society of an untreated case of postnatal depression is estimated to range from:

- AU$41,367 at the lower range, and
- AU$62,051 at the upper range.

4.4.4 Costs and benefits

The projected benefits were calculated as $628,778 – $943,175 for the lower range of Acorn effectiveness (8.0% of eligible mothers) and $918,347-$1,377,532 for the upper range of Acorn effectiveness (11.7% of eligible mothers) (Table 6). Comparison of the actual costs and projected benefits demonstrate that between 22.8% and 50.1% of the costs of Acorn are realised as a benefit from reduced postnatal depression. These projected benefits should be viewed in light of the fact that nearly 45% of the cohort of mothers were either diagnosed with no depression, or mild depression only, and were therefore not eligible to demonstrate any improvement.

Table 6 Costs and benefits for the Acorn program –total costs of program for 493 participants, benefits may be accrued for 273 mothers with pre-Acorn PHQ-9 of 10 or more. The lower ($41,367) and upper ($62,051) estimated benefits from treating a case of PND were derived from recent Australian study (PwC Consulting Australia 2019)

Item	Description	Cost	Item	Description	Benefit (lower) @$41,367 per treated case		Benefit (upper) @$62051 per treated case	
					$	%	$	%
Program costs	Deliver Acorn to 493 participants across 11 cohorts	2,756,098	Acorn effect – lower range (Kroenke) (8.0% of eligible mothers)	Significantly reduced PND in 15.2 mothers	628,778	22.8	943,175	34.2
			Acorn effect – upper range (RCSC c) (11.7% of eligible mothers)	Significantly reduced PND in 22.2 mothers	918,347	33.3	1,377,532	50.1

% - percent of costs to deliver program (2,922,128)

4.5 Method 2 – QALYs

4.5.1 Acorn results

The PHQ-9 and converted EQ-5D results are presented for pre- and post-ACORN and delta (Δ, calculated as pre-post) in Table 7. The mean PHQ-9 decreased from 11.24 (within the moderate depression range) to 9.98 (within the mild depression range). The finding of virtually no change (0.02 units) in EQ-5D between pre- and post-ACORN measures suggests that quality of life was in general consistent between measurement time periods, reflecting the previously mentioned finding of a stable EQ-5D in the presence of only small changes (~2-3 units) in PHQ-9 from baseline to endpoint (Furukawa, Levine et al. 2021).

Table 7 Pre-, post-, and delta (Δ)scores for the PHQ-9 (measured) and EQ-5D (converted) for 344 ACORN participants.

		Pre-		Post-		Delta (Δ)	
	n	mean	SD	mean	SD	mean	SD
PHQ-9	344	11.24	5.87	8.98	5.57	-2.26	5.32
EQ-5D	344	0.79	0.18	0.78	0.17	0.02	0.17

There was a significant relationship between $EQ\text{-}5D_{pre}$ and Δ EQ-5D (Figure 3) (R^2=0.267, $F_{1,338}$=123.2, P<0.001) (Appendix 2). As the $EQ\text{-}5D_{pre}$ decreases (indicating lower quality of life), the Δ EQ-5D (ie $EQ\text{-}5D_{pre}$ - $EQ\text{-}5D_{post}$) also decreases, which indicates a higher quality of life in the post-ACORN measures than pre-ACORN (Figure 3). Therefore, the effectiveness of ACORN on improving quality of life was greater in those with an initial lower quality of life. This impact is also demonstrated in Figure 4, where the median Δ EQ-5D decreases as the severity of depression (based on PHQ_{pre} scores) increases. Examination of outliers and extreme results identified in Figure 4 did not identify any obvious characteristics in either the mild, moderate, or moderate-severe groups.

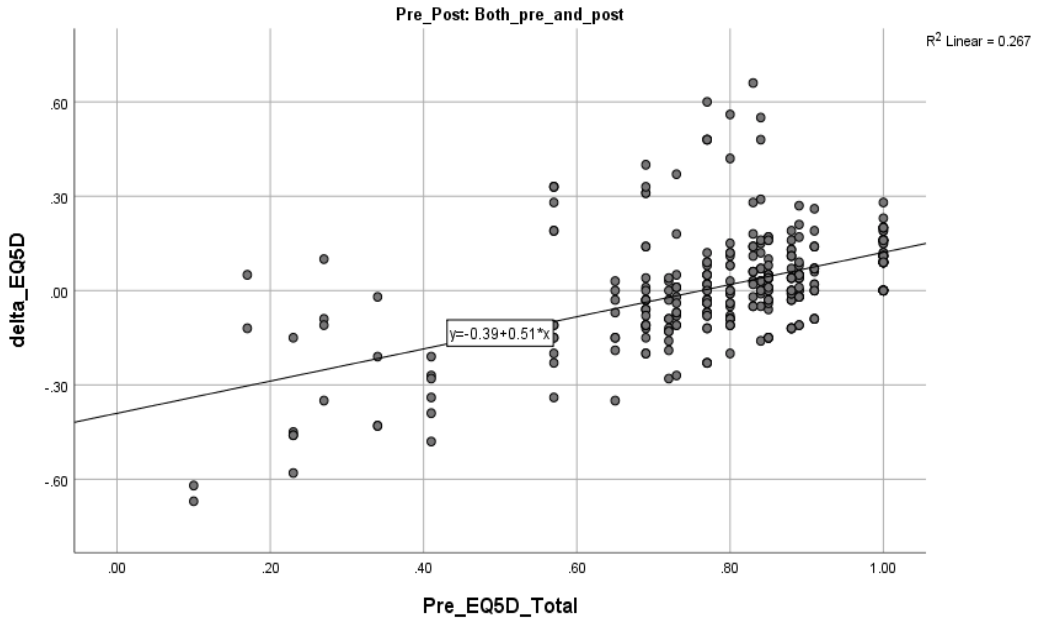

Figure 3 Scatter plot of EQ-5Dpre and Δ EQ-5D. A smaller (ie negative) Δ EQ-5D indicates improved quality of life post-ACORN.

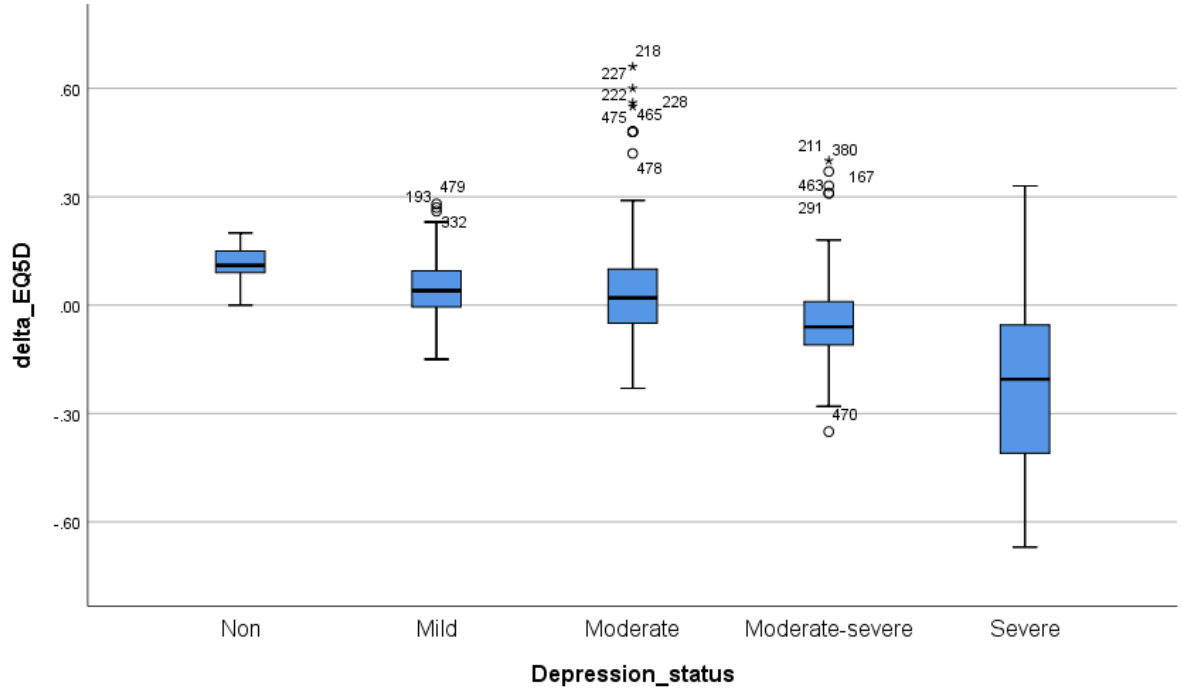

Figure 4 Box-plots showing median scores and 25th and 75th percentiles (upper and lower limits of the box), and outliers (O) and extremes (*) based on 1.5 and 3 x the interquartile range, respectively

The results for estimates of Δ EQ-5D are presented for participant characteristics, including pre-ACORN depression status, in Table 8. These results demonstrate that – apart from pre-ACORN depression status - the Δ EQ-5D does not vary by participant characteristics. For example, there was no obvious trend in Δ EQ-5D between attendance rates (ignoring the 'Less than half' category, n=2), age of parent or child, presence of other children, place of birth, language spoken at home, Aboriginality, education,

employment, income or marital status. Similarly, there were no apparent trends comparing Δ EQ-5D by wave of ACORN (Figure 5).

Table 8 Δ EQ-5D by participant characteristics. A smaller (ie negative) Δ EQ-5D indicates improved quality of life post-ACORN.

Characteristic	Category	n	Δ EQ-5D Mean	SD
How many of the group sessions have you attended?	All	44	0.02	0.16
	Most	257	0.00	0.17
	Half	35	0.08	0.21
	Less than half	2	0.19	0.19
Depression status (pre-ACORN)	Non	45	0.10	0.06
	Mild	109	0.04	0.10
	Moderate	93	0.06	0.18
	Moderate-severe	60	-0.04	0.15
	Severe	37	-0.18	0.27
Parent age group	<30	135	0.00	0.18
	30 or older	207	0.03	0.17
Child age groups	0-5 months	140	0.02	0.19
	6-11 months	128	0.02	0.16
	12-17 months	47	0.03	0.19
	18+ months	29	-0.04	0.17
Other children	Yes	113	0.03	0.17
	No	86	0.03	0.22
	Unknown	145	0.00	0.15
Place of birth	Australia	298	0.01	0.17
	Other	43	0.06	0.21
Speak language other than English	Yes	55	0.04	0.19
	No	288	0.01	0.17
Aboriginal or Torres Strait Islander	Yes	7	0.10	0.12
	No	335	0.01	0.18
Education	School including Yr 12	87	0.02	0.18
	Cert Dip Tafe or College	133	0.02	0.16
	Uni degree	116	0.00	0.19
	Other	7	-0.03	0.11
Employment	F/T parent or on leave from work	198	0.02	0.17
	Currently working	62	0.00	0.12
	Not working but may be studying	59	0.02	0.20
Income	Wages earned by you or partner	233	0.02	0.17
	Gov Benefit, pension, allowance	110	0.01	0.19
Marital status	Married or defacto	265	0.02	0.17
	Single or separated/divorced or other	76	0.01	0.20

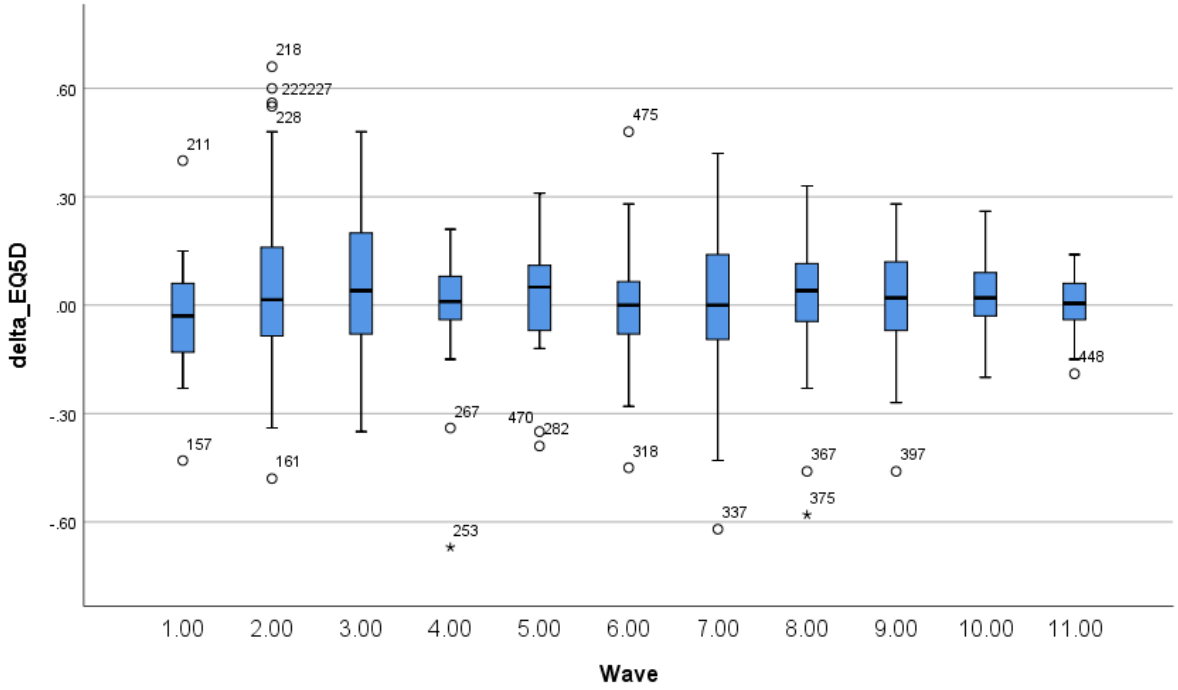

Figure 5 Δ EQ-5D by ACORN wave

4.5.2 Wait-list controls

Data for intervention and control outcomes were extracted from the wait-list control studies using the three most common outcomes (BDI-II, EPDS and HRSD) in Appendix 1. A review of the literature for these outcomes identified conversions to the PHQ-9 from both the BDI-II (Hawley, Gale et al. 2013) and HRSD (Gerbasi, Eldar-Lissai et al. 2020), however, no conversion was identified for the EPDS. Based on data extraction presented in Appendix 1, the converted data for control groups in wait-list control studies are presented in Table 10. Estimates of the change (Δ) in PHQ-9 ($PHQ\text{-}9_{pre}$ – $PHQ\text{-}9_{post}$) over the course of the wait-list control time period (8-12 weeks) ranged from 7.0 to 0.6, with a mean of 2.7 (median 2.0, interquartile range 1.6 to 2.9), demonstrating a slight improvement in mental health. Given that the majority of studies were conducted on postpartum women with major depression, this would be equivalent to a PHQ_{pre} from 15-19 (correlating to moderately severe depression on the PHQ scale). Taking a mid-point for PHQ_{pre} of 17.5, this would equate to a PHQ_{post} of 14.3. In EQ-5D terms, this would equate to an $EQ\text{-}5D_{pre}$ of 0.69 and an $EQ\text{-}5D_{post}$ of 0.69 (Furukawa, Levine et al. 2021). Therefore, the Δ EQ-5D for wait-list controls is 0, indicating no change in quality of life.

4.5.3 Effect of Acorn on quality of life

The results for wait-list controls indicates substantial differences in the mental health of wait-list control studies and Acorn participants. For example, at baseline, the mean PHQ-9 of Acorn participants, was 11.2, whereas the above estimates for wait-list controls are a mean of 17.5. Examining the curve generated for calculating EQ-5D from PHQ-9 scores (Figure 1) indicates that there is a clear risk to the validity of this conversion in the presence of substantial variation in PHQ-9 because the relationship is non-linear. The relationship is very steep (indicating a large change in EQ-5D for a small change in PHQ-9) between PHQ-9 scores of 16-20 (equivalent to control studies), whereas the curve is shallower between PHQ-9 scores of 10-15 (equivalent to Acorn participants). Therefore, these differences do suggest caution in the following PHQ-9-EQ-5D conversions.

$EQ\text{-}5D_{pre}$ (month 0) and $EQ\text{-}5D_{post}$ (month 3) scores for wait-list control and Acorn groups are presented in Figure 6. Four groups are presented: wait-list control estimates, all Acorn participants (n=343); and results

are also presented for Acorn participants with severe depression prior to commencing Acorn (n=37) and those with either non, mild, moderate or moderate-severe depression ("not severely", n=306). After three months, all groups are extrapolated to a hypothetical point of an EQ-5D of 0.85, which is equivalent to a PHQ-9 of 6 (Furukawa, Levine et al. 2021), indicating mild depression.

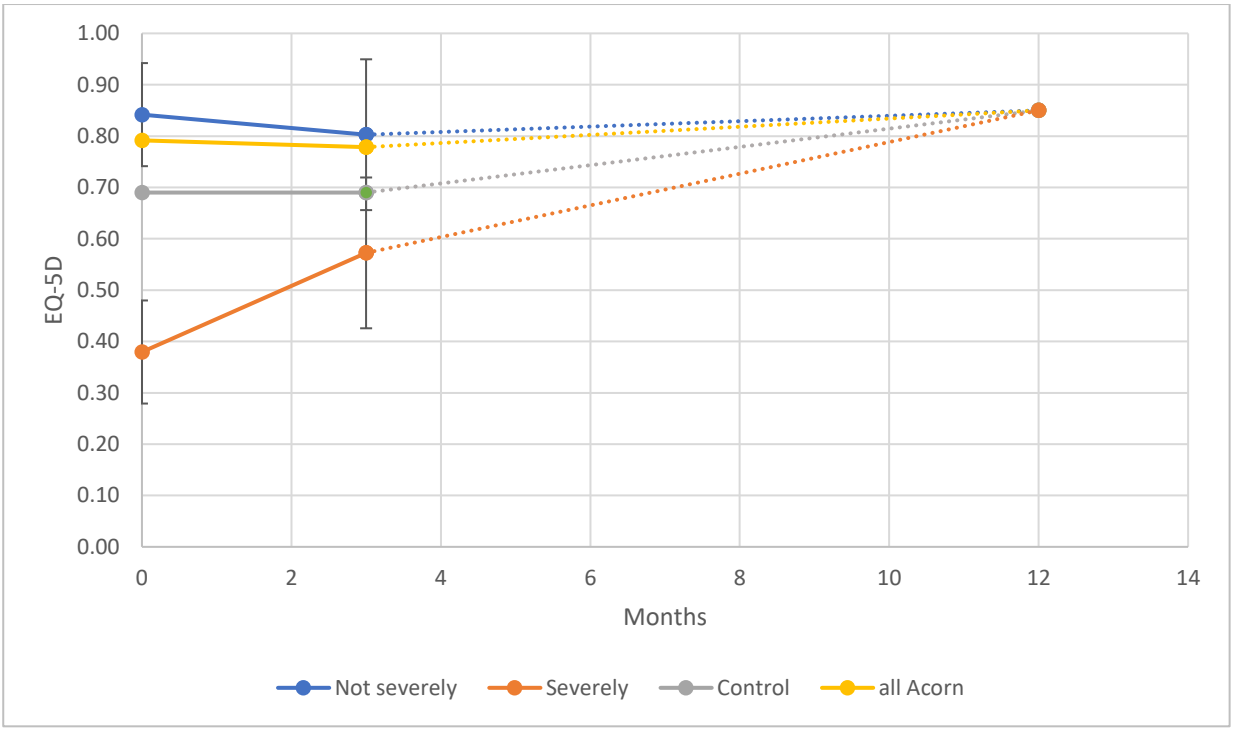

Figure 6 Schematic figure of change in EQ-5D scores pre-intervention (month 0) and post-intervention (month 3) for control (grey) and all Acorn participants (n=343, yellow); results are also presented for Acorn participants with severe depression prior to commencing Acorn (n=37, orange) and those with either non, mild, moderate or moderate-severe depression (n=306, blue).

Examination of Figure 6 identified two key findings which precluded further analysis of the PHQ-9/EQ-5D results. First, both control populations (grey) and all Acorn participants (yellow) demonstrated no apparent change in EQ-5D between pre- and post-time periods, despite changes in PHQ-9 of 2.7 units, and 2.3 units, respectively. This is in contrast to the worked example provided in (Furukawa, Levine et al. 2021), in which a change in PHQ-9 of 2 units between intervention and control groups led to estimates of a change in EQ-5D ranging between 0.08 and 0.13. This is due to the different curves between 'baseline' and 'endpoint' curves previously described in Figure 1, in which the same level of depression (as measured by the PHQ-9), is rated as a lower quality of life in measures that are made after an intervention compared to before the intervention. As mentioned previously, due to this effect, the PHQ-9 would need to change by 3-4 units - rather than the 2-3 units demonstrated here – to impact the EQ-5D quality of life measure. Email discussion with Prof Furukawa highlighted that the lack of a direct control group, and the requirement to use the different curves corresponding to baseline (pre-) and endpoint (post-), suggested that the method was too imprecise to identify any effect.

The second problematic finding relates to the shape of the two curves presented in Figure 1, and the earlier discussion about caution in the PHQ-9-EQ-5D conversions. As demonstrated by the 'severely depressed' group in Figure 6, there is a much greater gain in EQ-5D at higher PHQ-9 scores, especially between scores of 17-22. This may represent a true relationship, however, the higher levels of depression

in the wait-list control participants compared to the Acorn mothers suggests a lack of validity of using the PHQ-9/EQ-5D conversion to compare results between these two groups. In other words, because the gain in EQ-5D per PHQ-9 unit is greater within the range 15-19 there will be a greater gain in EQ-5D from control participants because of their higher levels of depression than Acorn mothers. For these reasons, the results from this method are not discussed further.

4.6 Assumptions and limitations

While the use of a wait-list control population from other studies of PND is novel, there are implicit assumptions and limitations in its use that impact the strength of this study's findings. For example, the wait-list control participants tended to have higher levels of depression, and were exposed to 8, 10, and 12 week interventions, and were sourced entirely from North America (US and Canada). Additionally, the majority of the control participants were experiencing PND during the early post-partum period (first 2-6 months of motherhood), whereas more than 200 of the Acorn mothers (n=204, 59%) had infants aged 6 months or older. This difference in the timing and age of babies is important, because many women who experience early PND improve within the early four-month period and remain in remission with an estimated 38% continuing to be chronically depressed mothers (Vliegen, Casalin et al. 2014). Given that the program addressed mothers with children up to the age of 3 years, it is likely that the Acorn group contained more mothers experiencing chronic depression than the control groups. This effect would tend to underestimate the true effect of Acorn in terms of reducing maternal depression.

It is also the case that the Acorn program included mothers with a broad range of comorbidities which were not identified in the wait-list controls, including around one third who indicated borderline personality disorder. These mothers are known to create a higher economic burden on society due to their extensive use of treatment services (Meuldijk, McCarthy et al. 2017) and in the case of Acorn derived significant wellbeing benefits from the program across a number of dimensions including improved parental stress, perceived parental self-efficacy, improved quality of interactions with their children, in addition to their significant improvements in depression (Aylward 2022). Cost benefits derived from these improvements for these mothers are therefore potentially higher than for those without BPD and the cost savings of improvements in depression for these mothers are likely to have been underestimated in this study. We have focussed on one quantifiable aspect of wellbeing for the purposes of this study and it is important to acknowledge that an array of other significant clinical improvements in wellbeing measures for Acorn mothers cited in the broader evaluation are not included in this economic evaluation which subsequently underestimates the economic value of the program.

There was a wide range of estimates of clinically significant improvement in the control groups of the wait-list control studies, ranging from 14%-59.2%. The study with the highest rate measured participants' depression levels on five occasions over 8 weeks, suggesting that the repeated measurements may have contributed to a potential placebo effect. While the use of a median (rather than a mean) to estimate the percent clinically significant improvement in the control studies helps to reduce the impact of such outliers, such findings do raise questions about the applicability of wait-list control studies that vary significantly from the Acorn population and study design. Finally, there is undoubtedly variation in the measurement of 'clinically significant change' within the wait-list control studies. Given that this study has measured a clinically significant change ranging from 26.3% to 30.0% in the same dataset using slightly different approaches, small changes in the method of calculation are likely to lead to differences in estimates within the wait-list control studies and in comparing wait-list control results with Acorn.

Additionally, there are a number of estimates from various sources used in the calculations above. For example, due to a lack of comparable Australian data, we have used UK estimates that the costs to society from perinatal depression are 68% of the total costs of depression and anxiety combined. Similarly, we

have estimated that the total societal costs of PND are between 50% and 75% of the total societal costs of perinatal depression.

Finally, the population used in the key study informing Method 2 – QALYs (Furukawa, Levine et al. 2021) was different from Acorn (eg 66% female, mean age 41.8, with higher levels of depression), and were undergoing internet cognitive behavioural therapies. The Furuawa et al. (2021) study also relied on the inclusion of four studies (out of seven) that had used other scales than PHQ-9 (eg Beck Depression Inventory), which were converted to PHQ-9 scores using algorithms (Furukawa, Levine et al. 2021). Although the results were consistent between the three PHQ-9 studies and the four converted studies (Furukawa, Levine et al. 2021) the equipercentile method for mapping between symptom severity scales such as the PHQ-9 and utility scales such as the EQ-5D was criticised as lacking rigour (Franklin and Young 2021).

5 References

Aylward, P. (2022). Evaluation of Acorn Parenting Program Waves 1-11. Adelaide, Action Research Partnerships and the Hopscotch Foundation.

Bauer, A., M. Parsonage, M. Knapp, V. Iemmi, B. Adelaja and S. Hogg (2014). The costs of perinatal mental health problems. London, Centre for Mental Health and London School of Economics**:** 44.

Brock, R. L., M. W. O'Hara and L. S. Segre (2017). "Depression Treatment by Non-Mental-Health Providers: Incremental Evidence for the Effectiveness of Listening Visits." Am J Community Psychol **59**(1-2): 172-183.

Buttner, M. M., R. L. Brock, M. W. O'Hara and S. Stuart (2015). "Efficacy of yoga for depressed postpartum women: A randomized controlled trial." Complement Ther Clin Pract **21**(2): 94-100.

Carter, T., A. Bastounis, B. Guo and C. Jane Morrell (2019). "The effectiveness of exercise-based interventions for preventing or treating postpartum depression: a systematic review and meta-analysis." Archives of Women's Mental Health **22**(1): 37-53.

Clark, R., A. Tluczek and R. Brown (2008). "A mother–infant therapy group model for postpartum depression." Infant Mental Health Journal **29**(5): 514-536.

Clark, R., A. Tluczek and A. Wenzel (2003). "Psychotherapy for postpartum depression: a preliminary report." Am J Orthopsychiatry **73**(4): 441-454.

Cook, N., S. Ayers and A. Horsch (2018). "Maternal posttraumatic stress disorder during the perinatal period and child outcomes: A systematic review." Journal of Affective Disorders **225**: 18-31.

Cox, E. Q., N. A. Sowa, S. E. Meltzer-Brody and B. N. Gaynes (2016). "The Perinatal Depression Treatment Cascade: Baby Steps Toward Improving Outcomes." J Clin Psychiatry **77**(9): 1189-1200.

Dennis, C. L. and E. Hodnett (2007). "Psychosocial and psychological interventions for treating postpartum depression." Cochrane Database Syst Rev(4): Cd006116.

Diaz, J. Y. and R. Chase (2010). The cost of untreated maternal depression. St Paul, Wilder Research.

Fitelson, E., S. Kim, A. S. Baker and K. Leight (2010). "Treatment of postpartum depression: clinical, psychological and pharmacological options." International journal of women's health **3**: 1-14.

Fonseca, A., S. Alves, F. Monteiro, R. Gorayeb and M. C. Canavarro (2020). "Be a Mom, a Web-Based Intervention to Prevent Postpartum Depression: Results From a Pilot Randomized Controlled Trial." Behav Ther **51**(4): 616-633.

Fonseca, A., F. Monteiro, S. Alves, R. Gorayeb and M. C. Canavarro (2019). "Be a Mom, a Web-Based Intervention to Prevent Postpartum Depression: The Enhancement of Self-Regulatory Skills and Its Association With Postpartum Depressive Symptoms." Front Psychol **10**: 265.

Franklin, M. and T. Young (2021). "Correspondence on "How can we estimate QALYs based on PHQ-9 scores? Equipercentile linking analysis of PHQ-9 and EQ-5D" by Furukawa *et al*." Evidence Based Mental Health **24**(4): e5.

Furukawa, T. A., S. Z. Levine, C. Buntrock, D. D. Ebert, S. Gilbody, S. Brabyn, D. Kessler, C. Björkelund, M. Eriksson, A. Kleiboer, A. van Straten, H. Riper, J. Montero-Marin, J. Garcia-Campayo, R. Phillips, J. Schneider, P. Cuijpers and E. Karyotaki (2021). "How can we estimate QALYs based on PHQ-9 scores? Equipercentile linking analysis of PHQ-9 and EQ-5D." Evidence Based Mental Health **24**(3): 97.

Gerbasi, M. E., A. Eldar-Lissai, S. Acaster, M. Fridman, V. Bonthapally, P. Hodgkins, S. J. Kanes and S. Meltzer-Brody (2020). "Associations between commonly used patient-reported outcome tools in postpartum depression clinical practice and the Hamilton Rating Scale for Depression." Archives of Women's Mental Health **23**(5): 727-735.

Hall, P. (2012). "Current considerations of the effects of untreated maternal perinatal depression and the National Perinatal Depression Initiative." Journal of Developmental Origins of Health and Disease **3**(4): 293-295.

Hanusa, B. H., S. H. Scholle, R. F. Haskett, K. Spadaro and K. L. Wisner (2008). "Screening for depression in the postpartum period: a comparison of three instruments." J Womens Health (Larchmt) **17**(4): 585-596.

Hawley, C. J., T. M. Gale, P. S. Smith, S. Jain, A. Farag, R. Kondan, C. Avent and J. Graham (2013). "Equations for converting scores between depression scales (MÅDRS, SRS, PHQ-9 and BDI-II): good statistical, but weak idiographic, validity." Hum Psychopharmacol **28**(6): 544-551.
Henderson, C., S. Dixon, A. Bauer, M. Knapp, C. J. Morrell, P. Slade, S. J. Walters and T. Brugha (2019). "Cost-effectiveness of PoNDER health visitor training for mothers at lower risk of depression: findings on prevention of postnatal depression from a cluster-randomised controlled trial." Psychol Med **49**(8): 1324-1334.
Huang, L., Y. Zhao, C. Qiang and B. Fan (2018). "Is cognitive behavioral therapy a better choice for women with postnatal depression? A systematic review and meta-analysis." PLOS ONE **13**(10): e0205243.
Jacobson, N. S., W. C. Follette and D. Revenstorf (1984). "Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance." Behavior Therapy **15**(4): 336-352.
Jacobson, N. S. and P. Truax (1991). "Clinical significance: a statistical approach to defining meaningful change in psychotherapy research." J Consult Clin Psychol **59**(1): 12-19.
Kroenke, K., R. L. Spitzer and J. B. Williams (2001). "The PHQ-9: validity of a brief depression severity measure." J Gen Intern Med **16**(9): 606-613.
McMillan, D., S. Gilbody and D. Richards (2010). "Defining successful treatment outcome in depression using the PHQ-9: A comparison of methods." Journal of Affective Disorders **127**(1): 122-129.
Meager, I. and J. Milgrom (1996). "Group Treatment for Postpartum Depression: A Pilot Study." Australian & New Zealand Journal of Psychiatry **30**(6): 852-860.
Meuldijk, D., A. McCarthy, M. E. Bourke and B. F. S. Grenyer (2017). "The value of psychological treatment for borderline personality disorder: Systematic review and cost offset analysis of economic evaluations." PLoS ONE **12**(3).
Morrell, C. J., P. Sutcliffe, A. Booth, J. Stevens, A. Scope, M. Stevenson, R. Harvey, A. Bessey, A. Cantrell, C. L. Dennis, S. Ren, M. Ragonesi, M. Barkham, D. Churchill, C. Henshaw, J. Newstead, P. Slade, H. Spiby and S. Stewart-Brown (2016). "A systematic review, evidence synthesis and meta-analysis of quantitative and qualitative studies evaluating the clinical effectiveness, the cost-effectiveness, safety and acceptability of interventions to prevent postnatal depression." Health Technol Assess **20**(37): 1-414.
Morrell, C. J., R. Warner, P. Slade, S. Dixon, S. Walters, G. Paley and T. Brugha (2009). "Psychological interventions for postnatal depression: cluster randomised trial and economic evaluation. The PoNDER trial." Health Technol Assess **13**(30): iii-iv, xi-xiii, 1-153.
O'Hara, M. W., S. Stuart, L. L. Gorman and A. Wenzel (2000). "Efficacy of Interpersonal Psychotherapy for Postpartum Depression." Archives of General Psychiatry **57**(11): 1039-1045.
Poobalan, A. S., L. S. Aucott, L. Ross, W. C. S. Smith, P. J. Helms and J. H. G. Williams (2007). "Effects of treating postnatal depression on mother-infant interaction and child development: Systematic review." British Journal of Psychiatry **191**(5): 378-386.
PwC Consulting Australia (2019). The cost of perinatal depression and anxiety in Australia, Gidget Foundation Australia.
Rees, S., S. Channon and C. S. Waters (2019). "The impact of maternal prenatal and postnatal anxiety on children's emotional problems: a systematic review." European Child & Adolescent Psychiatry **28**(2): 257-280.
Sambrook Smith, M., V. Lawrence, E. Sadler and A. Easter (2019). "Barriers to accessing mental health services for women with perinatal mental illness: systematic review and meta-synthesis of qualitative studies in the UK." BMJ Open **9**(1): e024803.
Segre, L. S., R. L. Brock and M. W. O'Hara (2015). "Depression treatment for impoverished mothers by point-of-care providers: A randomized controlled trial." Journal of Consulting and Clinical Psychology **83**(2): 314-324.
Sinesi, A., M. Maxwell, R. O'Carroll and H. Cheyne (2019). "Anxiety scales used in pregnancy: systematic review." BJPsych Open **5**(1): e5.

Steffen, A., J. Nübel, F. Jacobi, J. Bätzing and J. Holstiege (2020). "Mental and somatic comorbidity of depression: a comprehensive cross-sectional analysis of 202 diagnosis groups using German nationwide ambulatory claims data." BMC Psychiatry **20**(1): 142.
Stein, A., R. M. Pearson, S. H. Goodman, E. Rapa, A. Rahman, M. McCallum, L. M. Howard and C. M. Pariante (2014). "Effects of perinatal mental disorders on the fetus and child." Lancet **384**(9956): 1800-1819.
Stephens, S., E. Ford, P. Paudyal and H. Smith (2016). "Effectiveness of Psychological Interventions for Postnatal Depression in Primary Care: A Meta-Analysis." The Annals of Family Medicine **14**(5): 463.
Underwood, L., K. Waldie, S. D'Souza, E. R. Peterson and S. Morton (2016). "A review of longitudinal studies on antenatal and postnatal depression." Archives of Women's Mental Health **19**(5): 711-720.
Van Lieshout, R. J., H. Layton, C. D. Savoy, J. S. L. Brown, M. A. Ferro, D. L. Streiner, P. J. Bieling, A. Feller and S. Hanna (2021). "Effect of Online 1-Day Cognitive Behavioral Therapy-Based Workshops Plus Usual Care vs Usual Care Alone for Postpartum Depression: A Randomized Clinical Trial." JAMA Psychiatry **78**(11): 1200-1207.
Vliegen, N., S. Casalin and P. Luyten (2014). "The Course of Postpartum Depression: A Review of Longitudinal Studies." Harvard Review of Psychiatry **22**(1): 1-22.
Whitehead, S. J. and S. Ali (2010). "Health outcomes in economic evaluation: the QALY and utilities." British Medical Bulletin **96**(1): 5-21.
Wisner, K. L., D. K. Y. Sit, M. C. McShea, D. M. Rizzo, R. A. Zoretich, C. L. Hughes, H. F. Eng, J. F. Luther, S. R. Wisniewski, M. L. Costantino, A. L. Confer, E. L. Moses-Kolko, C. S. Famy and B. H. Hanusa (2013). "Onset timing, thoughts of self-harm, and diagnoses in postpartum women with screen-positive depression findings." JAMA Psychiatry **70**(5): 490-498.

Appendix 1. Wait-list controls

A search of Pubmed and the included studies of recent systematic reviews (Stephens, Ford et al. 2016, Cook, Ayers et al. 2018, Huang, Zhao et al. 2018, Carter, Bastounis et al. 2019, Rees, Channon et al. 2019, Sambrook Smith, Lawrence et al. 2019, Sinesi, Maxwell et al. 2019) identified nine studies evaluating the effectiveness of interventions for postpartum depression using a wait-list control (Meager and Milgrom 1996, O'Hara, Stuart et al. 2000, Clark, Tluczek et al. 2003, Clark, Tluczek et al. 2008, Buttner, Brock et al. 2015, Segre, Brock et al. 2015, Fonseca, Monteiro et al. 2019, Fonseca, Alves et al. 2020, Van Lieshout, Layton et al. 2021). The characteristics of these studies are summarised in Table 9. One study was not available in full text format (Meager and Milgrom 1996). Studies were conducted on women with post-partum depression, testing interventions such as cognitive behavioural therapy, mother-infant therapy groups, yoga, and listening visits, over the course of 8-12 weeks. Total sample size ranged from 20 (Meager and Milgrom 1996) to 403 (Van Lieshout, Layton et al. 2021). The most common outcomes that were measured were the Beck Depression Inventory (BDI-II, n=4 studies), Edinburgh Postnatal Depression Scale (EPDS, n=4 studies) and the Hamilton Rating Scale for Depression (HRSD, n=3 studies). No studies measured PHQ-9. Two studies used a sequence to allocate patients to either of the intervention or wait-list control groups, all other studies were randomised.

Data for intervention and control outcomes were extracted from the wait-list control studies using the three most common outcomes (BDI-II, EPDS and HRSD). A review of the literature for these outcomes identified conversions to the PHQ-9 from both the BDI-II (Hawley, Gale et al. 2013) and HRSD (Gerbasi, Eldar-Lissai et al. 2020), however, no conversion was identified for the EPDS. Based on data extraction presented in Appendix 1, the converted data for control groups in wait-list control studies are presented in

Table 10. Estimates of the change (Δ) in PHQ-9 (PHQ_{pre} – PHQ_{post}) over the course of the wait-list control time period (8-12 weeks) ranged from 7.0 to 0.6, with a mean of 2.7, demonstrating a slight improvement in mental health. Given that the majority of studies were conducted on postpartum women with major depression, this would be equivalent to a PHQ_{pre} from 15-19 (correlating to moderately severe depression on the PHQ scale). Taking a mid-point of 17, this would equate to a PHQ_{post} of 14.3.

Table 9 Characteristics of nine studies evaluating the effectiveness of interventions for postpartum depression using a wait-list control

Author	Population	Intervention	Outcomes	Notes
(Meager and Milgrom 1996)	Women with post-partum depression	A 10-week treatment program with educational, social support and cognitive-behavioural components.	BDI-II, EPDS	"Women with PPD receiving group therapy intervention reported more significantly reduced depressive symptoms than wait-list control" (from Clark et al 2003, no full text available)
(O'Hara, Stuart et al. 2000)	Post-partum mothers meeting DSM-IV criteria for major depression	12 weeks of interpersonal psychotherapy	BDI-II, HRSD	Random assignment
(Clark, Tluczek et al. 2003)	Post-partum mothers meeting DSM-IV criteria for major depression	Two interventions of 12 week duration: (i) a relationally focused mother-infant group model that addresses the needs of the mother, infant, and spouse, and (ii) an individual therapy approach	BDI-II, CES-D	Sequential assignment
(Clark, Tluczek et al. 2008)	Women with moderate to severe depressive symptoms during the postpartum period	12 week mother-infant therapy group	BDI-II	Report estimated marginal means; Sequential assignment
(Buttner, Brock et al. 2015)	Postpartum women with scores of 12 or more on the HDRS	8 week yoga intervention, twice per week	GD-IDAS, HSRD	Random assignment
(Segre, Brock et al. 2015)	Low-income, ethnic-minority pregnant women or mothers of young children with an EPDS score of greater than 12	8 week program of listening visits (LV) delivered at a woman's usual point-of-care, including home-visits or an ob-gyn office	EPDS, GD-IDAS, HSRD	Random assignment
(Fonseca, Monteiro et al. 2019, Fonseca, Alves et al. 2020)	At-risk post-partum women and/or women presenting with early-onset postpartum depressive symptoms	"Be a Mom": self-guided web-based intervention, grounded in cognitive behavioral therapy	EPDS, HADS	Report estimated marginal means
(Van Lieshout, Layton et al. 2021)	Mothers at least 18 years and EPDS at least 10, and infant younger than 12 months	1 day interactive cognitive behavioural therapy workshop with pre-post measures over 12 weeks	EPDS, GAD-7	Random assignment

BDI	Beck Depression Inventory (n=4 studies)	GD-IDAS	Inventory of Depression and Anxiety Symptoms (n=2 studies)
CES-D	Centre for Epidemiological Studies Depression Scale (n=1 study)	HADS	Hospital Anxiety and Depression Scale (n=1) study)
EPDS	Edinburgh Postnatal Depression Scale (n=4 studies)	HRSD	Hamilton Rating Scale for Depression (n=3) studies)
GAD-7	Generalised Anxiety Disorder Questionnaire (n=1 study)		

Table 10 Results from seven controlled studies evaluating the effectiveness of interventions for postpartum depression using a wait-list control before (pre) and after (post) the intervention. Outcomes have been converted to the PHQ-9 scale using conversion studies for the BDI-II (Hawley, Gale et al. 2013) and HRSD (Gerbasi, Eldar-Lissai et al. 2020).

Outcome	Citation	**Pre** Control			**Post** Control			
		n	mean	SD	n	mean	SD	Δ PHQ
BDI-II→PHQ-9	(Meager and Milgrom 1996)	10			10			-
	(O'Hara, Stuart et al. 2000)	51	12.0	5.2	51	10.4	6.0	1.6
	(Clark, Tluczek et al. 2003)	11	12.6	5.0	11	11.0	6.2	1.6
	(Clark, Tluczek et al. 2008)	14	11.6	5.4	14	11.0	5.4	0.6
HRSD→PHQ-9	(O'Hara, Stuart et al. 2000)	51	16.3	1.4		13.2	4.5	3.1
	(Buttner, Brock et al. 2015)	29	11.7	-0.9	29	4.7	1.5	7.0
	(Segre, Brock et al. 2015)	21	13.0	2.6	21	10.6	4.3	2.4

BDI-II Beck Depression Inventory II
HRSD Hamilton Rating Scale for Depression
PHQ-9 Patient Health Questionnaire

Table 11 Data extraction from seven wait-list control studies including comparison of pre- and post- measures for Intervention and Control groups

Outcome	Citation	**Pre** Intervention			Control			**Post** Intervention			Control		
		n	mean	SD	n	mean	SD	n	mean	SD	n	mean	SD
BDI	(Meager and Milgrom 1996)	10			10								
	(O'Hara, Stuart et al. 2000)	48	23.6	7.2	51	23.0	6.9	48	10.6	6.8	51	19.2	8.7
	(Clark, Tluczek et al. 2003)	9	26.9	7.3	11	24.5	6.4	9	15.9	8.5	11	20.6	9.2
	(Clark, Tluczek et al. 2008)	18	22.6	9.4	14	22.0	7.4	18	12.4	7.1	14	20.5	7.3
EPDS	(Meager and Milgrom 1996)	10			10								
	(Segre, Brock et al. 2015)	39	17.2	4.0	21	15.1	4.4	39	10.3	6.0	21	11.1	6.0

	(Fonseca, Monteiro et al. 2019, Fonseca, Alves et al. 2020)	98	9.3	5.0	96	8.1	4.5	65	6.9	3.6	82	6.9	3.7
	(Van Lieshout, Layton et al. 2021)	202	16.5	4.4	201	15.9	4.5	165	11.7	4.8	192	14.0	4.5
HRSD	(O'Hara, Stuart et al. 2000)	48	19.4	4.6	51	19.8	5.3		8.3	5.3		16.8	8.4
	(Buttner, Brock et al. 2015)	28	17.3	5.1	29	15.3	3.1	28	5.9	6.0	29	8.5	5.4
	(Segre, Brock et al. 2015)	39	18.4	6.5	21	16.6	6.6	39	11.0	7.3	21	14.3	8.2

BDI-II: Beck Depression Inventory II

EPDS: Edinburgh Postnatal Depression Scale

HRSD: Hamilton Rating Scale for Depression

Appendix 2. Regression between EQ-5Dpre and Δ EQ-5D

Regression

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Pre_EQ5D_Total[b]	.	Enter

a. Dependent Variable: delta_EQ5D

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.517[a]	.267	.265	.14959

a. Predictors: (Constant), Pre_EQ5D_Total

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2.759	1	2.759	123.286	.000[b]
	Residual	7.563	338	.022		
	Total	10.322	339			

a. Dependent Variable: delta_EQ5D

b. Predictors: (Constant), Pre_EQ5D_Total

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.390	.037		-10.413	.000
	Pre_EQ5D_Total	.512	.046	.517	11.103	.000

a. Dependent Variable: delta_EQ5D

Appendix 3. UK study of costs from perinatal depression and anxiety

A UK study has estimated that the average cost to society of one case of perinatal depression is around £74,000, of which £23,000 relates to the mother (Table 12) and £51,000 relates to impacts on the child (Table 13) (Bauer, Parsonage et al. 2014). These estimates include antenatal and postnatal depression, but exclude the costs of perinatal anxiety (estimated as £35,000 per case, comprising £21,000 to the mother and £14,000 to the child) and perinatal psychosis (estimated as £53,000 per case, comprising £47,000 to the mother and £6,000 to the child) (Bauer, Parsonage et al. 2014).

Table 12 Costs of perinatal depression, impact on mothers, £ per case (Table 1 from (Bauer, Parsonage et al. 2014))

Public sector	Wider society			Total
Health and social care	QALY losses	Productivity losses	Other	
1,688	18,158	2,514		**22,360**

Table 13 Costs of perinatal depression, impact on children, £ per case (Table 2 from (Bauer, Parsonage et al. 2014))

	Public sector			Wider society			Total
	Health and social care	Education	Criminal Justice	QALY losses	Productivity losses	Other	
Pre-term birth	974	-	-	418	22	14	
Infant death	-	-	-	22,157	-	-	
Emotional problems	1,020	-	-	4,609	2,169	-	
Conduct problems	837	-	1,974	3,396	1,797	7,446	
Special educational needs	-	3,166	-	-	-	-	
Leaving school without qualifications	-	-	-	-	1,463	-	
Total	2,831	3,166	1,974	30,580	5,451	7,460	**51,462**

Adjusting for an inflation rate of 3%, the cost to society of one case of perinatal depression (£74,000 in 2014 terms) amount to a total of £91,000 (28,300 attributed to the mother and 62,700 attributed to the child) in 2021 terms. In Australian dollars, this amount equals a total of $171,600 ($53,400 and $118,200, respectively).

Appendix 4. Australian study of costs from perinatal depression and anxiety

A recent Australian study documented the costs of perinatal depression and anxiety, including maternal and paternal impacts mainly in the short and medium term, and impacts to the child in the short, medium and longer term (Figure 7). The first year impacts were estimated as $877m and longer terms impacts were AU$5.2b (total $7.3b) for the 600,000 births per year in Australia and the estimated 60,000 mother and 30,000 fathers with perinatal depression and anxiety (Figure 8) (PwC Consulting Australia 2019). This equates to AU$121,667 per case of perinatal depression and anxiety.

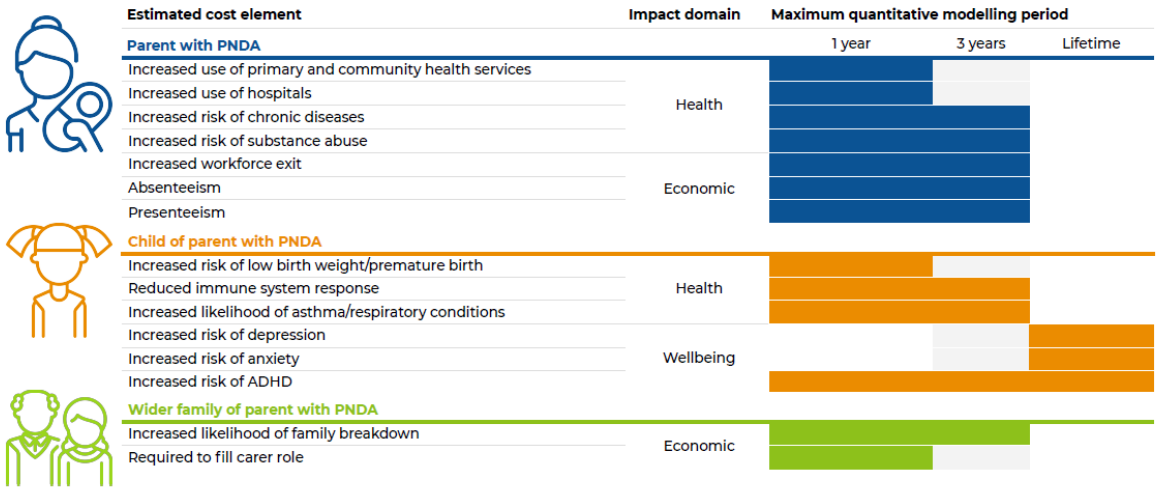

Figure 7 Time periods for different cost elements for estimates associated with perinatal depression and anxiety (Figure 4 from (PwC Consulting Australia 2019))

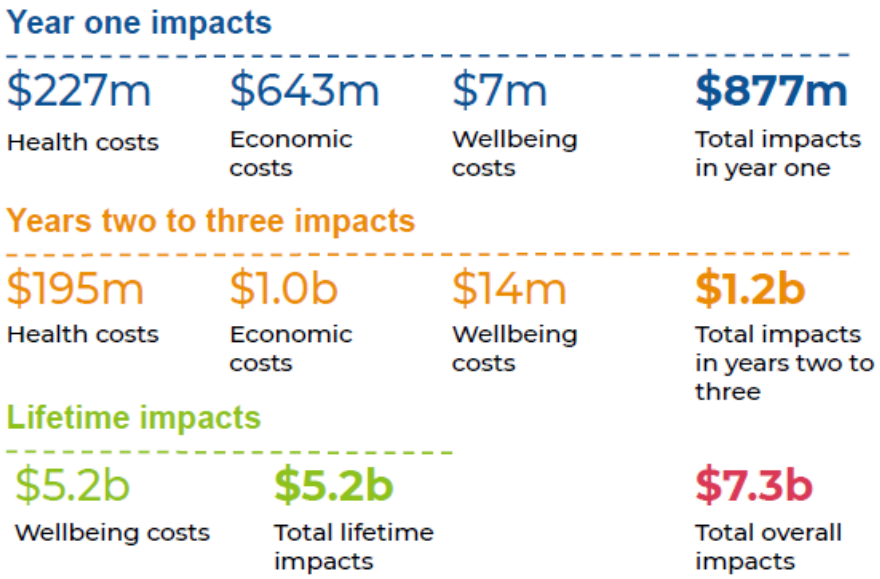

Figure 8 Summary of costs associated with perinatal depression and anxiety (Figure 6 from(PwC Consulting Australia 2019)

Assuming that the UK study can be used to estimate the attribution of costs to PND alone (excluding anxiety), it is estimated that the costs of depression (alone) are 68% of the total costs of depression and anxiety combined (Bauer, Parsonage et al. 2014). On that basis, the Australian estimate for the cost to society from perinatal depression is AU$82,734 (i.e. 0.68 x AU$121,667). This is a little under half of the estimate for the UK (AU$171,600), adjusted from 2014 to 2021 with a 3% inflation rate and a conversion of GBP1£ = AU$1.88.

Note that both estimates are for the perinatal period, and apart from 'pre-term birth' (total £1428 in the UK study), which accounts for the higher probability of pre-term birth in antenatal depression, it is difficult to allocate the costs attributed to either pre- or post-natal depression separately (Bauer, Parsonage et al. 2014). Up to 33% of postnatal depression begins in pregnancy and 27% begins in pre-pregnancy (Wisner, Sit et al. 2013). However, given the extended duration of postnatal depression and the proximity of the effects of postnatal depression on the infant (in comparison to antenatal depression), which accounts for some 68% of the total costs of perinatal depression, it seems reasonable to assume that postnatal depression accounts for at least half of the total costs of perinatal depression, and potentially as much as three-quarters of the total costs. These estimates will be used in subsequent analyses.

Therefore, the cost to Australian society from one case of postnatal depression is estimated as ranging from AU$41,367 to $62,051.

Appendix 5. PHQ-9 measures pre- and post-Acorn

The PHQ-9 results are presented for pre- and post-ACORN and delta (Δ, calculated as pre-post) in Table 14. The mean PHQ-9 for all participants who completed a pre- and post-Acorn evaluation decreased from 11.24 ± 5.87 (within the moderate depression range) to 8.98 (within the mild depression range). The Δ was -2.26 ± 5.32.

A subset of 190 participants with moderate, moderate-severe, or severe depression pre-Acorn had higher PHQ-9 scores pre-Acorn (15.48 ± 4.15) and a greater Δ (-4.02 ± 5.77) (Table 14). On average, these participants moved from the moderate severe category (a PHQ-9 score ranging from 15-19 to moderate depression (a PHQ-9 of 10-14).

Table 14 Pre-, post-, and delta (Δ)scores for the PHQ-9 (measured) for 343 ACORN participants who completed both pre- and post- measurements, and for the subset of participants (n=190) with moderate, moderate-severe, or severe depression pre-Acorn.

		Pre-		Post-		Delta (Δ)	
Group	n	mean	SD	mean	SD	mean	SD
Pre&Post	343	11.24	5.87	8.98	5.57	-2.26	5.32
≥ Moderately depressed	190	15.48	4.15	11.46	5.63	-4.02	5.77

Mean PHQ-9 scores are presented for participant characteristics in Table 15. Smaller (ie more negative) Δ PHQ-9 score, indicating a bigger effect on depression, were observed for participants attending all or most group sessions compared to half or less than half. Additionally, bigger effects were seen in mothers with more severe depression pre-Acorn (Table 15). Other characteristics were not, apparently related to Δ PHQ-9 score.

Table 15 Pre-, post-, and delta (Δ)scores for the PHQ-9 (measured) by participant characteristics for 343 ACORN participants

			Pre-		Post-		Delta (Δ)	
Characteristic	Category	n	mean	SD	mean	SD	mean	SD
How many of the group sessions have you attended?	All	44	10.6	4.8	8.3	4.9	-2.2	5.8
	Most	257	11.5	6.2	9.0	5.7	-2.5	5.0
	Half	35	10.1	4.9	9.3	5.8	-0.9	6.9
	Less than half	2	15.5	6.4	16.0	8.5	0.5	2.1
Depression status (pre-Acorn)	Non	45	2.8	1.3	4.3	2.3	1.5	2.4
	Mild	109	7.4	1.4	6.6	3.9	-0.8	3.9
	Moderate	93	12.0	1.3	10.2	5.4	-1.8	5.7
	Moderate-severe	60	16.9	1.4	11.0	4.8	-5.9	4.8
	Severe	37	22.0	2.1	15.5	5.6	-6.5	5.6
Parent age group	<30	135	12.0	6.2	9.3	5.6	-2.7	5.6
	30 or older	207	10.7	5.6	8.8	5.5	-1.9	5.1
Child age groups	0-5 months	140	11.3	6.0	9.1	6.0	-2.2	5.3
	6-11 months	128	10.7	5.7	8.4	5.3	-2.3	5.1
	12-17 months	47	12.4	6.0	11.0	4.9	-1.4	5.6

	18+ months	29	11.7	5.7	7.7	4.7	-4.0	6.0
Other children	Yes	113	10.8	5.7	9.1	5.7	-1.7	5.3
	No	86	11.8	6.0	9.5	6.2	-2.2	6.1
	Unknown	145	11.3	5.9	8.6	5.1	-2.7	4.8
Place of birth	Australia	298	11.4	5.8	8.9	5.5	-2.5	5.3
	Other	43	10.4	6.4	9.7	6.3	-0.7	5.0
Speak language other than English	Yes	55	10.8	6.2	9.1	6.1	-1.6	5.4
	No	288	11.3	5.8	9.0	5.5	-2.4	5.3
Aboriginal or Torres Strait Islander	Yes	7	5.7	2.1	7.0	5.5	1.3	4.8
	No	335	11.4	5.9	9.0	5.6	-2.3	5.3
Education	School including Yr 12	87	11.5	6.0	9.4	5.5	-2.1	5.4
	Cert Dip Tafe or College	133	11.5	5.8	9.5	5.7	-2.0	4.9
	Uni degree	116	10.8	5.9	8.0	5.4	-2.8	5.8
	Other	7	11.0	5.2	7.9	3.7	-3.1	2.3
Employment	F/T parent or on leave from work	198	11.3	5.7	8.8	5.3	-2.4	5.2
	Currently working	62	9.8	5.4	7.2	4.6	-2.6	4.9
	Not working but may be studying	59	12.6	6.6	10.9	6.5	-1.7	6.0
Marital status	Married or defacto	265	11.1	5.7	8.8	5.4	-2.3	5.3
	Single or separated/divorced or other	76	11.7	6.5	9.5	6.0	-2.2	5.5