Women and Birth 35 (2022) e133–e141

Contents lists available at ScienceDirect

Women and Birth

journal homepage: www.elsevier.com/locate/wombi

The clinical performance and cost-effectiveness of two psychosocial assessment models in maternity care: The Perinatal Integrated Psychosocial Assessment study

Georgina M. Chambers[a], Willings Botha[a], Nicole Reilly[b,c,d], Emma Black[e,f,g], Dawn Kingston[h], Marie-Paule Austin[c,d,i,*]

[a] National Perinatal Epidemiology and Statistics Unit (NPESU), Centre for Big Data Research in Health and School of Women's and Children's Health, University of New South Wales, Sydney, NSW, 2052 Australia
[b] Centre for Health Service Development, Australian Health Services Research Institute, University of Wollongong, NSW 2522, Australia
[c] Perinatal and Women's Mental Health Unit, St John of God Health Care, 23 Grantham St, Burwood NSW 2134, Australia
[d] School of Psychiatry, University of New South Wales, Sydney, NSW 2052, Australia
[e] Drug and Alcohol Services, South Eastern Sydney Local Health District, 591 South Dowling Street, Surry Hills, NSW 2010, Australia
[f] Discipline of Addiction Medicine, University of Sydney, Camperdown, NSW 2006, Australia
[g] School of Public Health and Community Medicine, Faculty of Medicine, University of NSW, Sydney 2052, Australia
[h] Faculty of Nursing, University of Calgary, 2500 University Drive, NW, University of Calgary, Calgary, AB T2N 1N4, Canada
[i] Royal Hospital for Women, Sydney, Barker Street, Randwick, NSW 2031, Australia

ARTICLE INFO

Article history:
Received 6 March 2021
Received in revised form 20 May 2021
Accepted 21 May 2021

Keywords:
Pregnancy
Psychosocial assessment
Depression screening
Cost-effectiveness

ABSTRACT

Problem: Although perinatal universal depression and psychosocial assessment is recommended in Australia, its clinical performance and cost-effectiveness remain uncertain.
Aim: To compare the performance and cost-effectiveness of two models of psychosocial assessment: Usual-Care and Perinatal Integrated Psychosocial Assessment (PIPA).
Methods: Women attending their first antenatal visit were prospectively recruited to this cohort study. Endorsement of significant depressive symptoms or psychosocial risk generated an 'at-risk' flag identifying those needing referral to the Triage Committee. Based on its detailed algorithm, a higher threshold of risk was required to trigger the 'at-risk' flag for PIPA than for Usual-Care. Each model's performance was evaluated using the midwife's agreement with the 'at-risk' flag as the reference standard. Cost-effectiveness was limited to the identification of True Positive and False Positive cases. Staffing costs associated with administering each screening model were quantified using a bottom-up time-in-motion approach.
Findings: Both models performed well at identifying 'at-risk' women (sensitivity: Usual-Care 0.82 versus PIPA 0.78). However, the PIPA model was more effective at eliminating False Positives and correctly identifying 'at-risk' women (Positive Predictive Value: PIPA 0.69 versus Usual Care 0.41). PIPA was associated with small incremental savings for both True Positives detected and False Positives averted.
Discussion: Overall PIPA performed better than Usual-Care as a psychosocial screening model and was a cost-saving and relatively effective approach for detecting True Positives and averting False Positives. These initial findings warrant evaluation of longer-term costs and outcomes of women identified by the models as 'at-risk' and 'not at-risk' of perinatal psychosocial morbidity.

Statement of significance

Problem
Depressive and anxiety disorders in the perinatal period are associated with a range of salient risk factors, which may lead to poorer outcomes for both mother and infant. The

* Corresponding author at: Perinatal and Women's Mental Health Unit, St John of God Health Care and University of New South Wales, PO Box 261, Burwood, NSW 1805, Australia.
E-mail address: m.austin@unsw.edu.au (M.-P. Austin).

http://dx.doi.org/10.1016/j.wombi.2021.05.007

most effective approach to identifying risk, initiating referral, and the economic costs associated with psychosocial and depression screening, remain uncertain.

What is already known

Routine psychosocial assessment is becoming more widespread in the Australian maternity setting and psychosocial assessment in pregnancy is associated with greater referral to psychosocial and mental health services.

What this paper adds

This paper provides initial evidence for the effectiveness of the Perinatal Integrated Psychosocial Assessment (PIPA) model of care compared to Usual-Care for psychosocial screening in hospital maternity settings. Assuming the slightly lower sensitivity of PIPA is clinically acceptable, this study indicates that the PIPA model would be cost-saving and a relatively effective approach for detecting the True Positive cases and averting False Positive cases of women requiring further assessment of psychosocial risk.

1. Introduction

The perinatal period – pregnancy through to the first postnatal year - is a time of increased risk for recurrent and new onset mental health morbidity [1,2], with significant impact on the outcomes for women, their offspring and families [3]. Mental health and severe psychosocial risk associated with suicide, account for some of the highest causes of maternal mortality [4,5]. Depression and anxiety disorders arising within the perinatal period are recognised as an international public health problem, affecting around 15% women [6–8]. The majority of these disorders are pre-existing [9] and up to 40% of women with symptoms of postnatal depression also experienced these in pregnancy, highlighting the importance of beginning psychosocial and depression assessment during pregnancy [10]. Furthermore, the financial costs to society and the healthcare system of maternal mental health conditions are substantial [11–13]. Recent Australian data indicates that the total healthcare costs to mothers alone of untreated perinatal depression and anxiety to be $70 million in the first postnatal year, increasing to over $500 million after additional healthcare and productivity costs for fathers and children are taken into consideration [14].

Over the last two decades, routine integrated psychosocial assessment (inclusive of depression screening) has been examined in terms of its possible value in the early identification of women at greater risk of mental health morbidity or poor adjustment to parenting [15]. Integrated psychosocial assessment is a broader approach to assessment that also enquires about issues such as domestic violence, substance misuse, social support and stressors, mental health history and history of childhood trauma, in addition to current symptoms. Current Australian national clinical guidelines recommend the routine use of the Edinburgh Depression Scale (EPDS) [16] along with psychosocial questions using tools such as the Antenatal Risk Questionnaire (ANRQ) [17] to flag women at significant psychosocial risk needing further monitoring, referral or support. This approach is also reflected in a number of state-based initiatives and is embedded into routine maternity care in a range of maternity hospitals in Australia. In New South Wales, for example, the SAFE START policy directive [18] and clinical practice guidelines [19] have underpinned midwife-led antenatal psychosocial care at the study site (Royal Hospital for Women [RHW]) since 2010.

A review of the effectiveness of integrated psychosocial assessment in perinatal women concluded that ongoing research is critical if the ideal model of routine depression screening in conjunction with integrated psychosocial assessment is to be established, and that availability of integrated, on-site assessment and intervention services improves accessibility to those identifed as being 'at-risk' [20]. Small randomised controlled trials examining the value of an integrated psychosocial assessment in pregnant women [21,22] have found more women with psychosocial risk were detected when administered a structured psychosocial health assessment [23] than those who were not. Two Australian survey-based studies have shown that the various perinatal mental health initiatives of the last two decades have been associated with a steady increase in perinatal integrated psychosocial assessment [24] and that integrated psychosocial assessment in pregnancy is associated with greater referral to psychosocial and mental health services [25].

However, there remains considerable controversy and uncertainty around the clinical and cost-effectiveness of routine depression screening in the perinatal period [26,27]. For example, a study which used modelled data to examine the value of routine perinatal depression screening using validated tools such as the Edinburgh Depression Scale (EPDS) [16] (but without broader psychosocial assessment), coupled with best practice depression treatment in a primary care setting, concluded that formal identification methods for postnatal depression *did not* represent value for money [28]. However, the authors noted that a key determinant of cost-effectiveness was the likely additional costs associated with 'False Positive' cases, that is managing women incorrectly diagnosed as depressed, highlighting the importance of minimising false positives if integrated psychosocial assessment and depression screening in pregnant women is to be cost-effective. Similarly, an internal review of the SAFESTART policy implementation at RHW found that a significant number of women were unnecessarily referred for psychosocial services leading to an undue burden on the hospital's psychosocial services (A. Datta, personal communication, 1st January–31st October, 2012). These initial findings, and the need to examine the performance and cost-effectiveness of an integrated psychosocial assessment in the perinatal period as articulated in the Australian national guideline [29], provided the impetus for the development of an alternative model of care – the Perinatal Integrated Psychosocial Assessment (PIPA) model [30].

The aim of this study was to compare the clinical and cost-effectiveness of two models of routine integrated psychosocial assessment and referral – Usual-Care and PIPA – in a real-world maternity setting using contemporary data, with the key outcome and reference standard being midwife agreement with the 'at-risk' flag generated in each model.

2. Methods

2.1. Study design, participants and data sources

This was a prospective cohort study that consecutively recruited women attending their first antenatal visit at RHW. The data collection period for Usual-Care was from January to December 2015, and for the PIPA model was from August 2017 to July 2018. For this study, administrative data were extracted from the hospital administrative database for all eligible women [30] who completed a routine integrated psychosocial assessment as part of their first midwife visit: data was drawn from ObstetriX for Usual-Care and from eMaternity for PIPA (as eMaternity was implemented across NSW from 2017).

Sociodemographic and clinical information extracted included maternal age, gestation, country of birth, interpreter required, parity, partner status (available for PIPA only), EPDS scores and responses to the psychosocial assessment questions.

The EPDS is a 10-item measure and is arguably the most widely used depression screening tool in perinatal care worldwide [16].

An EPDS score of 13 or more has been shown to have moderate sensitivity (0.66; 95% CI 0.58–0.74) and high specificity (0.95; 95% CI 0.92–0.96) for detecting possible major depression in pregnant and postpartum women [31].

The psychosocial questions used in Usual-Care were those provided in the SAFE START policy and guideline, whereas the PIPA model used the Antenatal Risk Questionnaire-Revised (ANRQ-R; see below). The ANRQ-R has been shown to have an acceptable area under the curve (AUC = 0.79−0.80) and to correctly classify 72–76% of current 'cases' and 'non-cases' when used to detect current depression or anxiety in the perinatal period (sensitivity = 0.70−0.74, specificity = 0.72−0.76) [32]. No corresponding psychometric information is available for the SAFE START psychosocial questions.

Information relating to whether a woman was flagged as 'at risk' was also extracted. Key differences in how this 'at risk' flag is defined in the Usual-Care and PIPA models is described in more detail in the following section.

Resource consumption and unit resource cost estimates were collected by survey and time in motion observational data collected during the study period. These data included time required for the midwife to complete the psychosocial assessment, administrative tasks associated with the initiation of referrals and discussion at the Triage meeting.

The study protocol, including more detailed information relating to data collection procedures, is available elsewhere [30].

2.2. Ethics

The project was approved by South Eastern Sydney Local Health District Human Research Ethics Committee (SESLHD HREC; Ref. 14/117).

2.3. Key differences between Usual-Care and PIPA models of integrated psychosocial assessment and referral

Women completed a series of questions during their first antenatal visit with a midwife in both the Usual-Care and PIPA models. Unlike Usual-Care, the PIPA model used a scored psychosocial questionnaire (the ANRQ-R, a revised version of ANRQ) which measures cumulative risk and generates both item level and total scores; auto-scores both the EPDS and ANRQ-R (eliminating manual scoring errors); includes structured questions to explore self-harm on the EPDS; and enables the midwife to document specific clinical concerns not elicited by the questionnaires (e.g., the woman's presentation at interview). In addition, PIPA generates clear psychosocial risk levels and referral pathways tailored to the woman's psychosocial risk level (see Table 1 for detail). While each model included an 'at-risk' flag, indicating that referral to the hospital Triage Committee was appropriate, a critical difference between the Usual-Care and PIPA model flags was that just one risk factor could trigger referral in Usual-Care, while PIPA required a higher threshold of risk for referral to the Triage Committee.

However, based on her opinion during the antenatal visit, the midwife undertaking the assessment, acting as the 'reference standard', could either override or agree with a positive or negative 'at-risk' flag generated by the Usual-Care or PIPA models resulting in the four possible outcomes for each woman: True Positive, False Positive, True Negative or False Negative. This approach acknowledges that unlike symptom-based screening, there is no 'gold standard' comparison measure available for the more comprehensive psychosocial assessment central to this study. That is, use of traditional gold standards instruments (such as the SCID [33]) is inappropriate in the context of a study that focuses on the clinical and cost-effectiveness of broader psychosocial assessment rather than the effectiveness of symptom-based mental health screening per se. Instead, the approach we have applied is in line with similar studies that have effectively used midwives as a 'reference standard' when determining the true psychosocial risk status of patients relative to women being identified as such through an electronic algorithm [34]. We hypothesised that the more detailed and precise PIPA risk and referral flag might be expected to reduce False Positives, and in turn reduce unnecessary referral to the Triage Committee meeting and costs associated with this.

Midwives have been undertaking routine integrated psychosocial assessment at the study site for many years and have regular in services on integrated psychosocial assessment, while new

Table 1
Comparison of key features of the Usual-Care and PIPA models of integrated psychosocial assessment.

	Model of integrated psychosocial assessment	
	SAFE START model (Usual-Care)	PIPA model (Alternative model)
Psychosocial assessment measures	EPDS, SAFE START psychosocial questions	EPDS, ANRQ-R (psychosocial questions); clinician concerns
Psychosocial risk levels	*Three levels of psychosocial risk, defined as:* *Level 1*: no specific vulnerabilities or risk *Level 2*: *one or more risk factors* of variable severity and significance including, but not limited to, low supports, multiple birth, financial stress, isolation, 'mild-moderate' depression or anxiety, history of mental health problem, young age. *Level 3*: *one or more of four complex risk factors* (domestic violence, involvement with child protection services, substance misuse, severe mental illness)	*Six levels of psychosocial risk, defined as:* *No risk*: ANRQ-R = 0; EPDS < 13 (Q10 = 0); no clinician concerns. *No risk on ANRQ-R* (ANRQ-R = 0) *but clinician concerns and/or EPDS = 13 or 14.* *Low risk*: ANRQ-R = 1–24 (excluding e.g., significant mental health history[a]) and EPDS < 15 (Q10 = 0). *Medium risk*: ANRQ-R = 1–24 (including e.g., significant mental health history[a]) and EPDS < 15 (Q10 = 0). *Medium-high risk*: ANRQ-R ≥25 (excluding any 'complex' risk factors[b]) or combination of 'social' risk factors[c] or EPDS ≥ 15 (Q10 = 0) or childhood trauma and neglect. *High risk*: ANRQ-R >25 and other 'social' risk factors[c] or any 'complex' risk factor (s)[b] or EPDS Q10 ≥ 1.
Triage meeting referral threshold	Levels 2 and 3	High, Medium- High, Medium risk

MCD; multidisciplinary case discussion meeting, ANRQ-R; Antenatal Risk Questionnaire-Revised, EPDS; Edinburgh Postnatal Depression Scale.
Reproduced with permission from Reilly et al. 2017.

[a] 'Significant' mental health history: involving professional help and/or had functional impact.

[b] 'Complex' risk factors: homelessness or housing instability; domestic violence; substance misuse; contact with child protection services.

[c] 'Social' risk factors: young maternal age (less than 20 years); no partner; booking-in appointment at >20 weeks gestation.

midwives are trained in integrated psychosocial assessment when they commence working in the antenatal clinics [35]. There were no differences in the two timeframes in terms of midwife training.

2.4. Descriptive statistics and clinical performance analysis

Descriptive statistics were used to compare the demographic characteristics of women who received care under the Usual-Care and PIPA models. The outcome measure for the performance of each model of care was whether the midwife undertaking the assessment (henceforth referred to as the screening midwife) agreed with the 'at-risk' flag or not. Measures of model performance included sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), likelihood ratios and Youden's Index [36], using midwife agreement as the reference standard.

2.5. Cost-effectiveness analysis

The cost analysis was performed from a healthcare perspective, and included the time commitment of the screening midwives, administrative staff and Triage Committee clinicians. Costs were quantified using a bottom-up approach using the staff time in minutes for the antenatal booking-in visits, administrative tasks and the Triage Committee meetings [37,38]. Staff time was costed using New South Wales Public Hospital pay rates, including on-costs [39–41]. Overhead costs, such as office space, depreciation and consumables, were not included as they were assumed to be the same for both models. All salaries were costed at 2017 pay rates in Australian dollars (AUD).

A cost-effectiveness analysis was performed for the cases flagged as 'at-risk' (according to the PIPA or Usual-Care flag) in each model plus whether the screening midwife agreed with the 'at- risk' flag or not. Effectiveness was measured in terms of True Positives (that is, the screening midwife agreed with the positive 'at-risk' flag and the woman was referred to the Triage Committee), and False Positives (that is, the screening midwife did not agree with the positive 'at-risk flag and the woman was not referred to the Triage Committee). These specific measures of performance were chosen because of the perceived high False Positive rate under the Usual-Care model and in other psychosocial screening tools [28]. The time horizon for the economic evaluation was set to one year, therefore discounting was inconsequential.

A decision tree was used as a vehicle for conducting the cost-effectiveness analysis [42,43] reflecting the screening, assessment and referral for psychosocial care during pregnancy for the Usual-Care and PIPA models.

Overall uncertainty of costs and effectiveness was quantified using the percentile method which estimated the 95% confidence intervals from 5,000 Monte Carlo simulations of a probabilistic sensitivity analysis (PSA) for costs and a non-parametric bootstrapping the case outcomes. A Gamma probability distribution was assigned to the cost parameters computed from the variation of individual costs activities. The 5,000 replications were plotted on a cost-effectiveness plane to visualise the uncertainty around the parameter estimates.

Analyses of the study were performed using SPSS v24.0 [44] and Microsoft Excel 2019 [45,46].

3. Results

Between January and December 2015, 3673 women were recruited to the Usual-Care model, and between August 2017 and July 2018, 3132 women were recruited to the PIPA model. Only 4.9% of women birthed twice or more in the total study period (January 2015 to September 2018). While there were statistically significant differences between the cohorts on key sociodemographic and clinical characteristics, including age (mean 32.2 years vs 32.6 years, $p < .001$) and gestational age at booking-in visit (15.3 weeks vs 15.6 weeks, $p < .001$), the effect size differences for all variables compared were small or not clinically significant (Table 2). The first antenatal visit occurred at 15.4 weeks gestation on average.

3.1. Clinical performance of Usual-Care and PIPA models of integrated psychosocial assessment

Fig. 1 outlines the results of the cohorts of women who received care under the Usual-Care model and the PIPA model. The decision tree was used to track women based on the results of the model's integrated psychosocial assessment at their first antenatal visit (positive or negative 'at-risk' flag indicating whether a woman should be referred to the Triage Committee), and whether the screening midwife agreed with the 'at-risk' flag and sent the women to the Triage Committee or not. Of the 3673 women who were assessed under the Usual-Care model, 1343 (37%) were identified as being at psychosocial risk by the Usual-Care 'at-risk' flag. By contrast, of the 3132 women assessed under the PIPA model, 850 (27%) were identified as being at psychosocial risk by the PIPA 'at-risk' flag (Table 3).

The probability of detecting a True Positive case (sensitivity) was marginally higher in the Usual-Care model than the PIPA

Table 2
Sociodemographic and clinical characteristics, Usual-Care and PIPA.

	Usual-Care (N = 3673) N (%)	PIPA (N = 3132) N (%)	p-Value	Effect size[a]
Maternal age (mean years and SD)	32.2 (4.7)	32.6 (4.5)	$p < 0.001$	$\eta^2 = 0.002$
Gestation at first antenatal visit (mean weeks and SD)	15.3 (4.4)	15.6 (4.1)	$p < 0.001$	$r = -0.108$
Country of birth				
Australia	1552 (42.3%)	1252 (40.0%)	Ns	
Born overseas (English speaking background	780 (21.3%)	729 (23.3%)	$p = 0.047$	$Phi = 0.024$
Born overseas (non-English speaking background)	1337 (36.4%)	1151 (36.7%)	Ns	
Interpreter Required	175 (4.8%)	106 (3.4%)	$p = 0.005$	$Phi = -0.034$
Parity 1 or more	1490 (40.6%)	1308 (41.8%)	$p < 0.001$	$Phi = 0.010$
Partnered (married or de facto)[b]	Not available	3065 (97.9%)	–	–
EPDS Total Score 13 or more	153 (4.3%)	101 (3.2%)	$p = 0.029$	$Phi = -0.027$
EPDS Q10 Score 1 or more	58 (1.6%)	31 (1.0%)	$p = 0.032$	$Phi = -0.027$

Ns = not significant.
[a] Interpretation of effect sizes: η^2 (eta-squared): 0.01 = small; 0.06 = moderate; 0.14 large; Cohen's r and phi: 0.1 = small; 0.3 = moderate; 0.5 = large.
[b] Informtation relating to partner status was extracted from eMaternity for PIPA, but was not available in ObstetriX for Usual-Care.

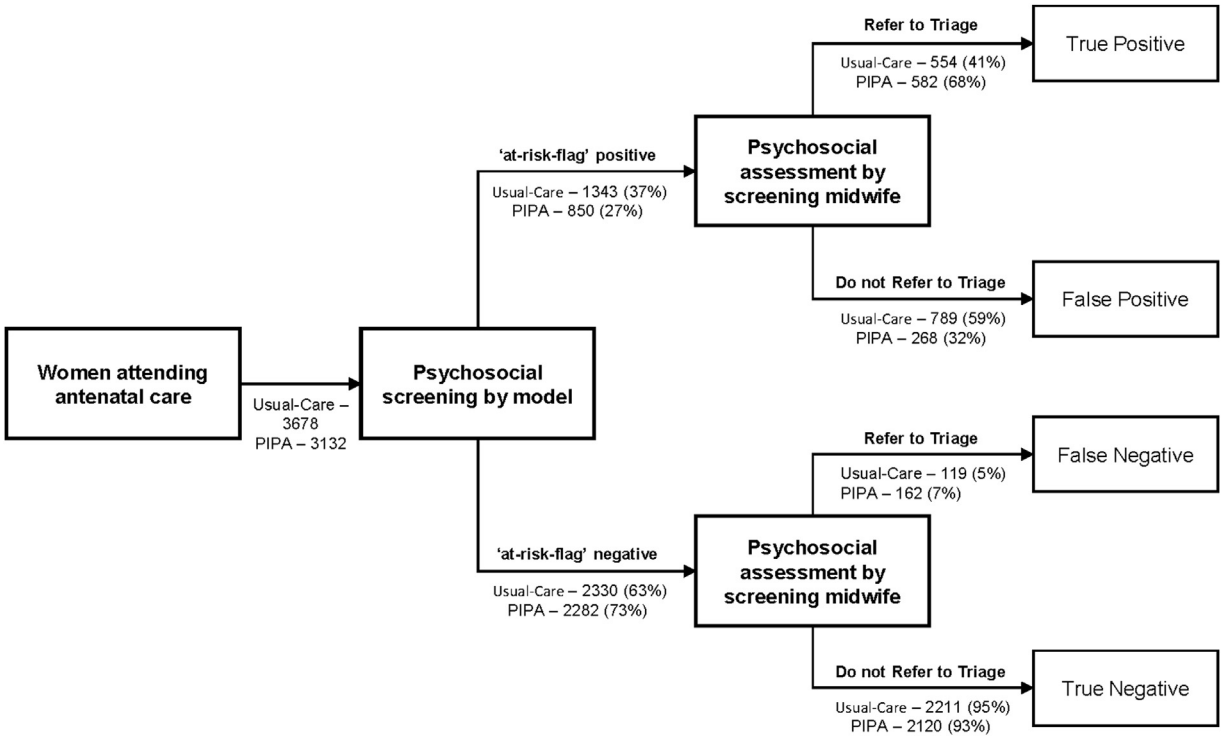

Fig. 1. Schematic diagram of study results Usual-Care versus PIPA models of care.

Table 3
Psychosocial 'at risk' flag: test performance and 95% CIs, Usual-Care versus PIPA (using the screening midwife as the reference standard).

Clinical performance characteristic	Definition	Usual-Care (n = 3673)	PIPA (n = 3132)
True Positives (TP)	Number of women where the midwife agreed with the positive 'at risk' flag.	554	582
False Positives (FP)	Number of women where the midwife did not agree with the positive 'at risk' flag.	789	268
True Negatives (TN)	Number of women where the midwife agreed with the negative 'at risk' flag.	2211	2120
False Negatives (FN)	Number of women where the midwife did not agree with the negative 'at risk' flag.	119	162
'Prevalence'	Proportion of women that the screening midwife considered to be at psychosocial risk (TP + FN)	0.18 (0.17–0.20)	0.24 (0.22–0.25)
Sensitivity (TP/TP + FN) ('at risk' flag is correct)	Probability of a *positive 'at risk' flag* in women with significant psychosocial risk (as identified by the screening midwife)	0.82 (0.79–0.85)	0.78 (0.75–0.81)
Specificity (TN/TN + FP) ('at risk' flag is correct)	Probability of a *negative 'at risk' flag* in women without significant psychosocial risk (as identified by the screening midwife)	0.74 (0.72–0.75)	0.89 (0.87–0.90)
Positive predictive value (PPV) (TP/TP + FP)	Proportion of *positive 'at risk' flags* that are True Positives according to the screening midwife	0.41 (0.39–0.44)	0.69 (0.65–0.72)
Negative Predictive Value (NPV) (TN/TN + FN)	Proportion of *negative 'at risk' flags* that are True Negatives according to the screening midwife	0.95 (0.94–0.96)	0.93 (0.92–0.94)
Youden's Index (J statistic) Sensitivity + (specificity −1)	Summary performance of the 'at risk' flag. A value of 1 indicates that there are no False Positives or False Negatives	0.56	0.67

TP = True Positive; TN = True Negative; FP = False Positive; FN = False Negative.

model (0.82 and 0.78, respectively). That is, the PIPA model identified a slightly smaller proportion of women who the midwife thought needed to be referred to the Triage Committee. Conversely, the probability of detecting a True Negative case (specificity) was lower in the Usual-Care model than the PIPA model (0.78 and 0.89, respectively). The Positive Predictive Value (PPV), that is, the probability of the midwife agreeing that a woman was at psychosocial risk in those with a positive 'at-risk' flag, was lower in the Usual-Care model than the PIPA model (0.41 and 0.69, respectively). The PPV for the PIPA model was in the 'moderate' range (0.69) and significantly superior to that seen for the Usual-Care model (which was in the 'low' range at 0.41). The Negative Predictive Values (NPV), that is the probability of the midwife agreeing that a woman was not at risk, in those with a negative 'at-risk' flag, were very similar (0.95 and 0.93, respectively).

The PIPA model was also better at excluding women who the midwife did not agree were 'at' risk'. That is, it was over twice as effective at eliminating False Positive cases (False Positive Ratio of 26% and 11%).

The Youden index provides an overall measure of test performance, by balancing sensitivity (detecting women who the midwives agree are at psychosocial risk) and specificity (detecting women who the midwives agree are not at psychosocial risk). Values range from 0 to 1, with a value of one representing no False Positives. Generally, a cut-off point of above 0.5 represents an acceptable clinical test. The PIPA model performed better than Usual-Care (0.67 and 0.56, respectively).

3.2. Cost-effectiveness of Usual-Care and PIPA

The mean staff time to screen and refer women as needed, along with the salary rates and resulting costs are presented in Table 4. The mean cost of assessment and screening was marginally lower in the PIPA model ($13.63) than Usual-Care ($14.38), with almost identical costs for the Triage Committee meeting under both models ($11.32 and $11.34 respectively).

For True Positives (that is the screening midwife agreed with the 'at-risk' flag and referred the women to the Triage Committee), the mean cost for the PIPA model was $16.32 (95% CI: $14.53–$18.30), while the mean cost of the Usual-Care model was $16.46 (95% CI: $14.04–$19.12). See Table 5.

The overall proportion of women that were identified as True Positives was 18.6% (582/3132) for PIPA and 15.1% (554/3678) for the Usual-Care model. Thus, the incremental cost-effectiveness ratio (ICER) of per True Positive women detected as a proportion of all women screen was −$3.97.

For False Positives (that is where the screening midwife over-rode the positive 'at-risk' flag and did not refer the case to the Triage Committee), the mean cost (the cost of assessment) for the PIPA model was $13.63 (95% CI: $12.11–$15.23) while that of the Usual-Care model was $14.38 (95% CI: $12.15–$16.93). The overall proportion of screened women that were identified as False Positives was 8.6% (268/3132) for PIPA and 21.4% (789/3678) for Usual-Care, while the overall cost per woman assessed was $16.32 for PIPA and $16.45 for Usual-Care. Thus, the ICER results showed that the PIPA model was more effective at averting a False Positive case detected at a cost saving of $5.80/woman.

Overall, the results indicated that these two measures of model performance are cost-saving using the PIPA model to assess for psychosocial risk. However, the wide confidence intervals indicate a high degree of uncertainty. This is demonstrated by the Monte Carlo simulation plotted on the cost-effectiveness planes for the two models of care in Fig. 2. The screening effectiveness of the PIPA model was better than Usual-Care for both True Positive and False Positives with all simulations in the favourable direction (in the right-hand quadrants for True Positives cases and the left-hand quadrants for False Positives cases). However, a significant number of simulations fall above and below the upper and lower quadrants indicating a great degree of uncertainty around the cost-effectiveness.

The effectiveness of the PIPA model at identifying women as being at psychosocial risk was better than Usual-Care for both True Positives and False Positives with all simulations in the favourable direction (in the right-hand quadrants for True Positive cases and the left-hand quadrants for False Positive cases).

4. Discussion

This study showed that both the Usual-Care and PIPA models performed well in terms of appropriately identifying those women 'at-risk' (in terms of True Positives and False Positives) and requiring referral to the Triage Committee, with the Usual-Care model having a slightly better sensitivity (82%) compared to PIPA (78%). The study also found that the PIPA model had a higher probability than Usual Care of correctly identifying women who the midwife agreed were at sufficient psychosocial risk to refer to Triage (PPV: 69% and 41%, respectively). This better PPV for the PIPA model likely reflects the ability of the PIPA model algorithm, underpinned by the cumulative risk factor score calculation generated by the ANRQ-R, to set a more refined referral threshold and thus increase the proportion of True Positive cases. The PIPA model was also over twice as effective at eliminating False Positive cases at excluding women who the midwife did not agree were at sufficient psychosocial risk to refer to the Triage Committee.

The lower threshold that is set for the Usual-Care for women to be flagged as 'at-risk' places a greater demand on midwives to decide whether psychosocial referral is appropriate and as such relies more heavily on the presence of staff who are highly skilled in integrated psychosocial assessment. The midwives at the participating site were trained and highly confident in undertaking routine psychosocial assessment, resulting at least in part from workplace culture that has prioritise routine perinatal mental health care consistently over the last 20 years. However, we acknowledge that this is not necessarily the case in all maternity settings nationally and internationally, thus the more directive referral prompts that are central to the PIPA model may be useful for less well-trained midwives or in less resourced maternity settings. Furthermore, the impact of clinical competencies on the effectiveness of the model remains to be determined [35].

The mean cost of integrated psychosocial assessment and depression screening was only marginally lower in the PIPA models ($13.63) than the Usual-Care model ($14.38), but the PIPA model resulted in more correct referrals (True Positive cases) than the Usual-Care model and less incorrect referrals (False Positive cases) than the Usual-Care model. The cost savings per correct referral under the PIPA model compared to the Usual-Care model was $3.97, and per incorrect referral avoided was $5.80. Future

Table 4
Staff times and costs associated with Usual-Care vs PIPA, AUD 2017.

Screening	n	Mean time (minutes)	SD	mean cost[a]	SD	Distribution
Usual-Care	3673	20.6	10	$14.38	$1.19	Gamma
PIPA	3132	20.2	9.2	$13.63	$0.79	Gamma
Triage Committee						
Usual-Care	673	3.7	17.3	$11.34	$2.98	Gamma
PIPA	744	3.7	16.2	$11.32	$2.38	Gamma
All women						
Usual-Care	3673	–	–	$16.45	$1.31	Gamma
PIPA	3132	–	–	$16.32	$0.98	Gamma

[a] Staff salary rates sourced from New South Wales public health system awards. Available at: https://www.health.nsw.gov.au/careers/conditions/Pages/awards.aspx. Hourly salary rates; Level 3 Admin officer: $32.34, Midwife - Junior: $37.40, Senior Midwife: $44.99, Data Manager: $44.99, Senior Social Worker: $51.84, Clinical Midwife Consultant - Level 2: $64.93.

Table 5
Expected outcomes and mean costs PIPA versus Usual-Care for women identified as 'at risk'.

	Expected outcome (95%CI)	Mean cost, AUD 2017 (95%CI)	ICER, AUD 2017 (95%CI)
Correct referrals (True Positives)[a]			
Usual-Care	0.151 (0.139–0.162)	$16.45 ($14.04–$19.12)	
PIPA	0.186 (0.172–0.199)	$16.32 ($14.53–$18.30)	−$3.97 (−$201.97–56.58)
Incorrect referrals (False Positives)[b]			
Usual-Care	0.214 (0.201–0.228)	$14.38 ($12.15–$16.93)	
PIPA	0.086 (0.076–0.096)	$13.63 ($12.11–$15.23)	$5.80 (−$17.18– $25.09.33)

ICER; Incremental Cost-effectiveness Ratio, Usual-Care and PIPA; perinatal integrated psychosocial assessment.
[a] Expected outcomes for Correct referral (True Positives) is the probability of a screened women being a True Positive.
[b] Expected outcomes for Incorrect referral (False Positives) is the probability of a screened women being a False Positive.

Fig. 2. Cost-effectiveness plane for True Positive ('correct' referrals) and False Positives ('incorrect' referrals avoided), Usual-Care versus PIPA.
True Positive ICER replication in blue (upper right quadrant). False Positive replication in grey (lower left quadrant).
Note: The screening effectiveness of the PIPA model was better than Usual-Care for both True Positives and False Positives with all simulations in the favourable direction (in the right-hand quadrants for True Positive cases and the left-hand quadrants for False Positive cases).

studies that examine whether these results are an underestimate of the true cost savings are required, as this study only captured costs from the initial antenatal booking-in appointment to the Triage Committee meeting and did not capture costs associated with any subsequent or unnecessary engagements with mental health and social work clinicians.

The Usual-Care and PIPA models flagged 37% and 27% of women as 'at-risk', respectively, which largely reflect the range of values reported in other perinatal universal integrated psychosocial assessment studies. Most studies, like ours, have used a combined depression screener and structured psychosocial questionnaire to identify which women warrant a referral for further assessment. In a relatively disadvantaged Australian sample, Matthey et al. [47] found 52% of women endorsed one or more psychosocial risk factors, while our group in a more socioeconomically advantaged sample [48], found a 28.1% rate of significant psychosocial risk. Blackmore et al. [22] in Canada reported 26% and 36% respectively of their pregnant sample had significant psychosocial risk depending on how psychosocial risk was identified by maternity staff: at routine consult or using a structured assessment tool. Honikman et al. [49], in a South African sample referred 32% of their sample for counselling after routine integrated psychosocial assessment. Clearly variation of prevalence rates for 'at-risk' women assessed in this way, is to be expected depending on sample characteristics, how psychosocial risk is defined and the detection approach used (e.g., structured questionnaire versus clinical interview, and whether an additional 'filter' such as the clinical judgement of the midwife, is included in protocols for generating an 'at-risk' flag). While the rates of psychosocial risk found in our study appear to be in line with other populations, the performance of clinical tests is partially dependent on prevalence of the condition in question, and thus the generalisability of our findings to other antenatal populations should be confirmed before adoption of this approach more broadly, is planned.

There have been calls for studies which establish an optimal model of administering routine perinatal depression screening in conjunction with broader psychosocial assessment and which rigorously examine the effectiveness of electronic algorithms to support the early identification of possible depression or elevated psychosocial risk [20,50]. This study responds to these evidence gaps through its dual examination of clinical performance and cost-effectiveness of two alternative models of integrated psychosocial care. Its use of 'real world' psychosocial assessment, electronic decision aids and referral data overcomes some of the limitations associated with the use of simulated data in previous studies [28].

4.1. Limitations

This study has a number of limitations. It was a cohort study rather than a randomised controlled trial. While randomised trials may represent a superior design, the risk of contamination between groups in the real world setting central to this study was high. There is also the possibility that the sequential nature of the data collection (2015 and 2017) may have resulted in differences in midwives' experience with integrated psychosocial assessment. Midwives were used as the reference standard against which the performance of the 'at-risk' flag was assessed. The study was conducted at one site only, restricting the generalisability of our findings to other settings.

In this study, only an intermediate measure of cost-effectiveness of the PIPA and Usual-Care models were performed. That is, an assessment of the performance of the models according to midwives' opinion, and the cost-effectiveness of False Positive and True Positives results. This was in accordance with the published protocol [30] and motivated by the perceived high False Positive rates from the Usual-Care model and other studies [28], but does highlight that decision models can oversimplify clinical pathways [51,52]. It should be also noted that both models included clear care pathways for women with subclinical issues, including attendance at antenatal psychoeducational groups, midwife monitoring and repeat administration of the EPDS for women with elevated but subthreshold scores at the booking-visit. To perform a comprehensive cost-effectiveness analysis, quantification of health-related quality of life, generally measured by Quality Adjusted Life Years, would need to be estimated for the complete cohort of women who completed an integrated psychosocial assessment, either through a series of multi attribute utility instrument surveys at different time points, or using estimates from the literature which are few and not directly relevant to this study.

Finally the cost-effectiveness analysis focused solely on screening and triage activities and did not account for other costs associated with the two models. In particular False Negatives have a cost in terms of untreated morbidity, and that cost could not be accounted for in this study. Ideally a replication study that includes longer term follow-up data would allow additional examination of performance of the PIPA model in terms of clinical outcomes.

5. Conclusion

This study has extended the existing evidence base by examining both the clinical and cost-effectiveness of two integrated models of integrated psychosocial assessment and depression screening embedded within a primary care maternity setting. Assuming the slightly lower sensitivity of PIPA is clinically acceptable, the PIPA model of care would be a cost-saving and effective approach for detecting the True Positive cases and averting False Positive cases for integrated psychosocial assessments and depression screening by midwives in hospital maternity settings. However, because the midwives' assessment of risk during screening was used as the reference standard, meaning that True and False Negatives could not be determined, the overall cost-effectiveness and health outcomes of women screening by the PIPA model and Usual-Care could not be calculated.

Studies which address overall cost-effectiveness of perinatal integrated psychosocial assessment programs, that incorporate longer-term health outcomes, are warranted.

Conflict of interest

The authors have no financial or other conflict of interest to declare.

Ethical statement

The research on which this manuscript is based involved human research. We used administrative data routinely collected by midwives when women first present for antenatal care for this study and thus individual consent was not required. Permission to use this administrative data for our study was granted by the South Eastern Sydney Local Health District Human Research Ethics Committee and Research Governance Office (SESLHD HREC Ref: 14/117; SSA Ref: 14/336) on the 23 October 2014. The scientific value, methodological value and safety of the research has been reviewed and approved by the SESLHD HREC, Scientific Review Sub-Committee.

Funding

No direct research funding was received for this study. MPA, NR and EB thank St John of God Health Care for infrastructure funding. NR thanks Australian Rotary Health and the University of Newcastle (2018–2020) and the University of Wollongong (current) for postdoctoral fellowship funding support. DK holds the Lois Hole Hospital for Women Cross-Provincial Chair in Perinatal Mental Health and a Canadian Institutes of Health Research New Investigator Award. The funding bodies had no involvement in the design of this study, analysis or interpretation of data, writing of the manuscript or decision to submit the article for publication.

Author contributions

Marie-Paule V. Austin: Conceptualization; funding acquisition, supervision, writing - review & editing. **Georgina Chambers:** Methodology, formal analysis, original draft, writing - review & editing. **Willings Botha:** Formal analysis, writing - review & editing. **Nicole M. Reilly:** Conceptualization, methodology, writing - review & editing. **Emma Black:** Project administration, Writing - review & editing. **Dawn Kingston:** Writing - review & editing.

Acknowledgements

We gratefully acknowledge the contributions of Virginia Spear in the weekly extraction of routinely collected administrative and clinical data and for her ongoing support in the implementation of the PIPA Project. We sincerely thank Catriona Andronicus and Dinaz Kaithakulathil for preparing the scripts and contributing to the testing of the PIPA algorithm. We also thank Christine Marsh and the team at Meridian Health Informatics for their support in embedding the PIPA model in the NSW Health eMaternity database. Sincere thanks to RHW management, in particular Helen Jarman, Vanessa Medunic and Alyce Finch, and Nursing Unit Managers Jo Arkwright, Michele Pezzuti and Rebecca Moore, for their support of the PIPA Project. We thank the midwives and other clinicians and administrative staff who have generously given their time and expertise to this project. We also extend thanks to Drs Matthew Holt and Rohini Vasudevan for their contribution to data collection for the economic component, and Victoria Mule for her thorough editing of this manuscript. Finally, we gratefully acknowledge the co- investigators on the larger PIPA project (Professor Virginia Schmied, A/Professor Stephen Matthey and Dr Andrew Bisits).

References

[1] L.M. Howard, E. Molyneaux, C.-L. Dennis, T. Rochat, A. Stein, J. Milgrom, Non-psychotic mental disorders in the perinatal period, Lancet 384 (9956) (2014) 1775–1788.

[2] K.L. Wisner, D. Sit, M.C. McShea, et al., Onset timing, thoughts of self-harm, and diagnoses in postpartum women with screen-positive depression findings, J. Am. Med. Assoc. Psychiatry 70 (5) (2013) 490–498.
[3] A. Stein, R.M. Pearson, S.H. Goodman, et al., Effects of perinatal mental disorders on the fetus and child, Lancet 384 (9956) (2014) 1800–1819.
[4] AIHW, Maternal Deaths in Australia 2015–2017. Cat. No. PER 106, Australian Institute of Health and Welfare, Canberra, 2020.
[5] AIHW, M.D. Humphrey, M.R. Bonello, et al., Maternal Deaths in Australia 2008–2012. Maternal Deaths Series no. 5. Cat. No. PER 70, Australian Institute of Health and Welfare, Canberra, 2015.
[6] N.I. Gavin, B.N. Gaynes, K.N. Lohr, S. Meltzer-Brody, G. Gartlehner, T. Swinson, Perinatal depression: a systematic review of prevalence and incidence, Obstet. Gynecol. 106 (5(1)) (2005) 1071–1083.
[7] M.-P. Austin, D. Hadzi-Pavlovic, S. Priest, et al., Depressive and anxiety disorders in the postpartum period: how prevalent are they and can we improve their detection? Arch. Womens Ment. Health 13 (5) (2010) 395–401.
[8] P. Almond, Postnatal depression: a global public health perspective, Perspect. Public Health 129 (5) (2009) 221–227.
[9] T. Munk-Olsen, M.L. Maegbaek, B.M. Johannsen, et al., Perinatal psychiatric episodes: a population-based study on treatment incidence and prevalence, Transl. Psychiatry 6 (10) (2016) e919.
[10] J. Green, D. Murray, The use of the Edinburgh postnatal depression scale in research to explore the relationship between antenatal and postnatal dysphoria, in: J. Cox, J. Holden (Eds.), Perinatal Psychiatry, Gaskell, London, 1994 p. 180.
[11] A. Bauer, M. Parsonage, M. Knapp, V. Iemmi, B. Adelaja, The Costs of Perinatal Mental Health Problems, London School of Econmics, Personal Social Services Unit & Centre for Mental Health, London, 2014.
[12] G.M. Chambers, S. Randall, V.P. Hoang, et al., The national perinatal depression initiative: an evaluation of access to general practitioners, psychologists and psychiatrists through the Medicare Benefits schedule, Aust. N. Z. J. Psychiatry 50 (3) (2016) 264–274.
[13] G.M. Chambers, S. Randall, C. Mihalopoulos, et al., Mental health consultations in the perinatal period: a cost-analysis of Medicare services provided to women during a period of intense mental health reform in Australia, Aust. Health Rev. 42 (5) (2017) 514–521.
[14] COPE, Valuing Perinatal Mental Health: the Consequences of Not Treating Perinatal Depression and Anxiety Melbourne, Centre of Perinatal Excellence (COPE) & PricewaterhouseCoopers Australia, 2014.
[15] M.-P. Austin, D. Kingston, Psychosocial assessment and depression screening in the perinatal period: benefits, challenges and implementation, in: A.L. Sutter-Dallay, N. Glangeaud-Freudenthal, A. Guedeney, A. Riecher-Rössler (Eds.), Joint Care of Parents and Infants in Perinatal Psychiatry, Springer International Publishing, Cham, 2016, pp. 167–195.
[16] J. Cox, J. Holden, R. Sagovsky, Detection of postnatal depression: development of the 10 item Edinburgh Postnatal Depression Scale, Br. J. Psychiatry 150 (1987) 782–786.
[17] M.P. Austin, J. Colton, S. Priest, N. Reilly, D. Hadzi-Pavlovic, The Antenatal Risk Questionnaire (ANRQ): acceptabilty and use for psychosocial risk assessment in the maternity setting, Women Birth 26 (2013) 17–25.
[18] NSW Department of Health, NSW Health/Families NSW Supporting Families Early Package – SAFE START Strategic Policy, NSW Department of Health, 2010.
[19] NSW Department of Health, NSW Health/Families NSW Supporting Families Early Package – SAFE START Guidelines: Improving Mental Health Outcomes for Parents and Infants, NSW Department of Health, 2010.
[20] E. Felice, A. Agius, R. Sultana, E.M. Felice, J. Calleja-Agius, The effectiveness of psychosocial assessment in the detection and management of postpartum depression: a systematic review, Minerva Ginecol. 70 (3) (2018) 323–345.
[21] L. Willinck, R. Schubert, Antenatal psychosocial risk assessment project, Aust. Coll. Midwives Inc. J. 13 (3) (2000) 7–12.
[22] E.R. Blackmore, J.C. Carroll, A.J. Reid, et al., The use of the antenatal psychosocial health assessment (ALPHA) tool in the detection of psychosocial risk factors for postpartum depression: a randomized controlled trial, JOCG 28 (2006).
[23] A.J. Reid, A. Biringer, J.D. Carroll, et al., Using the ALPHA form in practice to assess antenatal psychosocial health. Antenatal Psychosocial Health Assessment, CMAJ 159 (6) (1998) 677–684.
[24] K. Moss, N. Reilly, A. Dobson, D. Loxton, L. Tooth, G. Mishra, How rates of perinatal mental health screening in Australia have changed over time and which women are missing out, Aust. N. Z. J. Public Health 44 (4) (2020) 301–306.
[25] N. Reilly, S. Harris, D. Loxton, et al., Referral for management of emotional health issues during the perinatal period: does mental health assessment make a difference? Birth 40 (4) (2013) 297–306.
[26] L.H. Chaudron, K.L. Wisner, Perinatal depression screening: let's not throw the baby out with the bath water!, J. Psychosom. Res. 76 (6) (2014) 489–491.
[27] Bd Thombs, E. Arthurs, S. Coronado-Montoya, et al., Depression screening and patient outcomes in pregnancy or postpartum: a systematic review, J. Psychosom. Res. 73 (2014) 433–446.
[28] M. Paulden, S. Palmer, C. Hewitt, S. Gilbody, Screening for postnatal depression in primary care: cost effectiveness analysis, Br. Med. J. 339 (b5203) (2009).
[29] M.-P. Austin, N. Highet, the Expert Working Group, Mental Health Care in the Perinatal Period: Australian Clinical Practice Guideline, COPE: Centre of Perinatal Excellence, Melbourne, 2017.
[30] N. Reilly, E. Black, G.M. Chambers, et al., Study protocol for a comparative effectiveness trial of two models of perinatal integrated psychosocial assessment: the PIPA project, BMC Pregnancy Childbirth 17 (1) (2017) 236.
[31] B. Levis, Z. Negeri, Y. Sun, A. Benedetti, B.D. Thombs, Accuracy of the Edinburgh Postnatal Depression Scale (EPDS) for screening to detect major depression among pregnant and postpartum women: systematic review and meta-analysis of individual participant data, BMJ 371 (2020) m4022.
[32] N. Reilly, D. Hadzi-Pavlovic, D. Loxton, E. Black, V. Mule, M.-P. Austin, Supporting routine psychosocial assessment in the perinatal period: the concurrent and predictive validity of the Antenatal Risk Questionnaire-Revised, Women Birth (2021) (in press).
[33] M.B. First, R.L. Spitzer, M. Gibbon, J.B. Williams, User's Guide for the Structural Clinical Interview for DSM-IV Axis Disorders, American Psychiatric Press, Washington, D.C, 1997.
[34] C. Quispel, T.A. Schneider, G.J. Bonsel, M.P. Lambregtse-van den Berg, An innovative screen-and-advice model for psychopathology and psychosocial problems among urban pregnant women: an exploratory study, J. Psychosom. Obstet. Gynecol. 33 (1) (2012) 7–14.
[35] V. Schmied, N. Reilly, E. Black, et al., Opening the door: midwives' perceptions of two models of psychosocial assessment in pregnancy- a mixed methods study, BMC Pregnancy Childbirth 20 (1) (2020).
[36] M. Hunink, M. Weinstein, E. Wittenberg, et al., Decision Making in Health and Medicine, 2nd edition, Cambridge University Press, UK, 2014.
[37] Z. Mogyorosy, P. Smith, The Main Methodological Issues in Costing Health Care Services: A Literature Review, (2005) .
[38] E. McIntosh, P. Clarke, E. Frew, J. Louviere, Applied Methods of Cost-Benefit Analysis in Health Care, OUP Oxford, 2010.
[39] NSW Health, Public Health System Nurses And Midwives (State) Award Sydney, NSW Department of Health, 2017.
[40] NSW Health, Health Employees Administrative Staff (State) Award Sydney, NSW Department of Health, 2017.
[41] NSW Health, NSW Health Service Health Professionals (State) Award, NSW Department of Health, Sydney, 2017.
[42] R. Edlin, C. McCabe, C. Hulme, P. Hall, J. Wright, Cost Effectiveness Modelling for Health Technology Assessment, Springer, 2015.
[43] A. Briggs, M. Sculpher, K. Claxton, Decision Modelling for Health Economic Evaluation, OUP Oxford, 2006.
[44] IBM Corp, IBM SPSS Statistics for Windows, Version 24.0, IBM Corp, Armonk, NY, 2016.
[45] A.M. Gray, P.M. Clarke, J.L. Wolstenholme, S. Wordsworth, Applied Methods of Cost-effectiveness Analysis in Healthcare, OUP Oxford, 2010.
[46] J. Lord, M.A. Asante, Estimating uncertainty ranges for costs by the bootstrap procedure combined with probabilistic sensitivity analysis 1, Health Econ. 8 (4) (1999) 323–333.
[47] S. Matthey, J. Phillips, T. White, et al., Routine psychosocial assessment of women in the antenatal period: frequency of risk factors and implications for clinical services, Arch. Womens Ment. Health 7 (4) (2004) 223–229.
[48] S.R. Priest, M.P. Austin, B.B. Barnett, A. Buist, A psychosocial risk assessment model (PRAM) for use with pregnant and postpartum women in primary care settings, Arch. Womens Ment. Health 11 (5) (2008) 307.
[49] S. Honikman, T. van Heyningen, S. Field, E. Baron, M. Tomlinson, Stepped care for maternal mental health: a case study of the perinatal mental health project in South Africa, PLoS Med. 9 (5) (2012)e1001222.
[50] N. Reilly, D. Kingston, D. Loxton, K. Talcevska, M.-P. Austin, A narrative review of studies addressing the clinical effectiveness of perinatal depression screening programs, Women Birth 33 (1) (2020) 51–59.
[51] H.F. Ryder, C. McDonough, A.N.A. Tosteson, J.D. Lurie, Decision analysis and cost-effectiveness analysis, Semin. Spine Surg. 21 (4) (2009) 216–222.
[52] Y. Yang, L. Abel, J. Buchanan, T. Fanshawe, B. Shinkins, Use of decision modelling in economic evaluations of diagnostic tests: an appraisal and review of health technology assessments in the UK, PharmacoEconomics Open (2019) 1–11.