



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/84042/>

Version: Accepted Version

Article:

Darwin, Z, McGowan, L and Edozien, LC (2015) Identification of women at risk of depression in pregnancy: using women's accounts to understand the poor specificity of the Whooley and Arroll case finding questions in clinical practice. Archives of Women's Mental Health. ISSN: 1434-1816

<https://doi.org/10.1007/s00737-015-0508-1>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

1 **Title**

2

3 Identification of women at risk of depression in pregnancy: Using women's accounts to understand the poor
4 specificity of the Whooley and Arroll case finding questions in clinical practice

5

6 **Authors**

7 Dr Zoe Darwin^{1*}

8 Professor Linda McGowan¹

9 Dr Leroy C Edozien²

10

11 ¹ School of Healthcare, Faculty of Medicine and Health, University of Leeds, LS2 9JT, UK

12 ² Manchester Academic Health Science Centre, University of Manchester, St Mary's Hospital, Manchester M13
13 9WL, UK

14

15 *Corresponding author

16 Dr Zoe Darwin

17 Address: Research Fellow in Maternal Wellbeing and Women's Health, School of Healthcare,
18 Faculty of Medicine and Health, University of Leeds, Leeds, LS2 9JT

19 Telephone: (+44) 113 343 0549

20 Email: z.j.darwin@leeds.ac.uk

21

22

23

24 **Abstract**

25

26 *Purpose*

27 Antenatal mental health assessment is increasingly common in high-income countries. Despite lacking evidence
28 on validation or acceptability, the Whooley questions (modified PHQ-2) and Arroll ‘help’ question are used in
29 the UK at booking (the first formal antenatal appointment) to identify possible cases of depression. This study
30 investigated validation of the questions and women’s views on assessment.

31

32 *Methods*

33 Women (n=191) booking at an inner-city hospital completed the Whooley and Arroll questions as part of their
34 routine clinical care, then completed a research questionnaire containing the Edinburgh Postnatal Depression
35 Scale (EPDS). A purposive sub-sample (n=22) were subsequently interviewed.

36

37 *Results*

38 The Whooley questions ‘missed’ half the possible cases identified using the EPDS (EPDS threshold ≥ 10 :
39 sensitivity 45.7%, specificity 92.1%; ≥ 13 : sensitivity 47.8%, specificity 86.1%), worsening to nine in ten when
40 adopting the Arroll item (EPDS ≥ 10 : sensitivity 9.1%, specificity 98.2%; ≥ 13 : sensitivity 9.5%, specificity
41 97.1%). Women’s accounts indicated that under-disclosure relates to the context of assessment and perceived
42 relevance of depression to maternity services.

43

44 *Conclusion*

45 Depression symptoms are under-identified in current local practice. Whilst validated tools are needed that can
46 be readily applied in routine maternity care, psychometric properties will be influenced by the context of
47 disclosure when implemented in practice.

48

49

50 **Key words**

51 mixed methods; perinatal mental health; pregnancy; screening; Whooley questions

52

53

54

55

56

57

58

59 **Introduction**

60 Perinatal mental health (PMH) encompasses new onset and pre-existing mental health illness that continues or
61 recurs in the period spanning pregnancy, childbirth and the first postnatal year (Austin 2004; Matthey 2004).
62 This includes severe mental illness (e.g. severe depression, schizophrenia, bipolar disorder, psychosis), which
63 has been implicated in maternal death (Centre for Maternal and Child Enquiries 2011) and more common mild-
64 moderate forms of depression and anxiety, estimated to affect 9-15% women at some stage during or after
65 pregnancy (Bennett et al. 2008; Gavin et al. 2005; Robertson et al. 2004). Clinical guidelines in several
66 countries recommend mental health assessment early in pregnancy to identify women who have or are at risk of
67 having mental health problems (American College of Obstetricians and Gynecologists 2006; American College
68 of Obstetricians and Gynecologists 2010; Austin et al. 2005; Carroll et al. 2005; National Collaborating Centre
69 for Mental Health 2007; Scottish Intercollegiate Guidelines Network 2012). In the UK, Australia and New
70 Zealand, this initial assessment is likely to be undertaken by the midwife as the lead healthcare professional
71 providing maternity care to women. In other areas including North America, assessment is more likely to be
72 considered the remit of medical doctors such as family doctors and obstetricians.

73
74 The Whooley questions (Whooley et al. 1997) have been introduced in England and Wales at booking (the first
75 formal antenatal appointment) and postnatally as an initial “pre-screen” to identify possible cases of depression,
76 based on current symptoms, that warrant further mental health review (National Collaborating Centre for Mental
77 Health 2007). The questions are: During the past month, have you been bothered by: (i) feeling down, depressed
78 or hopeless, (ii) having little interest or pleasure in doing things? (Whooley et al. 1997). Current National
79 Institute for Health and Care Excellence (NICE) clinical guidelines (National Collaborating Centre for Mental
80 Health 2007) advise additionally using the Arroll ‘help’ question (Arroll et al. 2003) to improve specificity, due
81 to concerns that the Whooley questions may over-identify women, resulting in over-burdening of systems and
82 unnecessary negative impact on women falsely identified as possible cases (i.e. false positives). The Arroll
83 question is: Is this something you feel you need/want help with? (Arroll et al. 2003).

84
85 The Whooley questions are a modified version of the PHQ-2 (Kroenke et al. 2003), a two-item version of the
86 PHQ-9 (Kroenke et al. 2001), which is based on the DSM-IV clinical interview. Although addressing the same
87 symptoms, the Whooley questions differ from the PHQ-2 regarding timescale (past four weeks instead of past
88 two weeks) and response format (dichotomous instead of ordinal four-point Likert scale). Published validation
89 studies on the Whooley questions (and the original PHQ-2) are summarised in Table 1.

90
91 No published studies have validated the tool as completed in clinical practice and an evidence synthesis
92 concluded that there was insufficient evidence to justify its clinical use (Hewitt et al. 2009). The NICE
93 guidelines have therefore been criticised (Martin and Redshaw 2009) for rejecting the most commonly used
94 measure of perinatal depression, the Edinburgh Postnatal Depression Scale (EPDS) (Cox et al. 1987; Murray
95 and Cox 1990), because of its sub-optimal positive predictive value and the lack of high-quality randomised
96 controlled trials demonstrating reduction in morbidity accompanying introduction of routine screening, yet
97 advocating an instrument that “probably would not meet the criteria either” of the National Screening
98 Committee (p.117) (National Collaborating Centre for Mental Health 2007). Another criterion of the National

99 Screening Committee is that any procedure is considered acceptable to the population. The EPDS is
100 recommended for routine clinical use in high-income countries including the US and Australia (American
101 College of Obstetricians and Gynecologists 2010; Austin et al. 2011) and whereas an evidence base exists for
102 acceptability of the EPDS, research on the Whooley questions is “urgently needed” given usage in current
103 clinical practice in the UK (Brealey et al. 2010).

104
105 The objectives of this study were to: i) provide the first validation of the Whooley and Arroll questions
106 completed at booking in UK clinical practice, and ii) explore women’s views and experiences of antenatal
107 mental health assessment that uses these questions.

108
109 *[Table 1 (summary of literature) about here]*
110
111

112 **Materials and methods**

113 The study used a mixed methods cohort design with sequential sampling (Teddlie and Yu 2007). In the
114 quantitative component, women attending their booking at an inner-city hospital were invited to take part,
115 regardless of any obstetric or other characteristics; the only exception being those unable to complete English-
116 language questionnaires unassisted. Information about the research accompanied the appointment letter and was
117 additionally provided by the researcher, who attended the antenatal clinic over a six-month period. Women self-
118 completed the Whooley (Whooley et al. 1997) and Arroll (Arroll et al. 2003) questions as part of clinical care
119 before completing a research questionnaire containing several measures, including the Edinburgh Postnatal
120 Depression Scale (EPDS) (Cox et al. 1987). The EPDS is a 10-item self-report measure rating depressive
121 symptoms in the last seven days (e.g. ‘I have felt sad or miserable’) using a 4-point Likert scale (0-3). Despite
122 its name, the EPDS is also validated for use during pregnancy (Murray and Cox 1990). In the absence of a
123 definitive threshold (Matthey et al. 2006) we used thresholds of ≥ 10 and ≥ 13 , respectively indicative of
124 ‘possible’ and ‘probable’ depression, and the more conservative threshold of ≥ 15 which is not commonly
125 reported but has been suggested for antenatal use (Cox and Holden 2003; Matthey et al. 2006). Additional
126 demographic and clinical data were abstracted from health records.

127
128 In the qualitative component, a purposive sub-sample of women were invited to take part in three serial in-depth
129 interviews, on the basis of scoring above threshold on at least one of the measures of psychological distress
130 (EPDS (Cox et al. 1987); State-Trait Anxiety Inventory state scale (Spielberger et al. 1987); GAD-2 (Kroenke et
131 al. 2007)) or psychosocial risk factors for postnatal depression (Antenatal Risk Questionnaire (Austin 2003)).
132 Topics included: women’s experiences of maternal mental health and well-being; its recognition by health
133 professionals, self and others; and available support. Discussion included women’s views and experiences of
134 mental health assessment during pregnancy and the postnatal period, including but not limited to the Whooley
135 and Arroll questions. Interviews were audio-recorded and were conducted twice during pregnancy and once in
136 the postnatal period, either at the participant’s home or the hospital research suite, according to participant
137 preference. Data were collected June 2010 – October 2011. Informed consent was gained prior to participation
138 in each component of the study.

139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178

Agreement between the Whooley and Arroll questions and the EPDS were analysed using standard diagnostic performance measures: sensitivity (the proportion of true positives correctly identified by the test), specificity (the proportion of true negatives correctly identified by the test), positive predictive value (PPV; the proportion of patients with positive test results who are correctly identified) and negative predictive value (NPV; the proportion of patients with negative test results who are correctly identified) (Altman and Bland 1994a; Altman and Bland 1994b). Here, the EPDS was treated as the gold standard against which the ‘test’ was compared; firstly using a positive response to either Whooley item as the criterion for possible caseness and secondly using the Arroll ‘help’ item as the criterion. The EPDS was treated as a ‘gold standard’ because the alternative to using the Whooley questions would be a different self-report measure, not clinical interview.

Qualitative data were transcribed verbatim and analysed using Framework Analysis, as described by Ritchie and colleagues (Ritchie and Spencer 1994). Rigour was promoted through strategies such as member checking with participants and searching for alternative explanations with the supervisory team (Lincoln and Guba 1985). Although the quantitative and qualitative components were primarily designed to answer different research questions, findings were integrated in the analysis stage, with women’s accounts offering insights into the quantitative findings.

Results

Characteristics for the full sample of women returning the research questionnaire (n=191) and the sub-sample interviewed (n=22) are presented in Table 2. These 191 women represented 16.5% of the women booked in the study timeframe, with reasons for non-approach presented in Figure 1. Comparison with local maternity data for the study period provided by the hospital found that White British women and older women were over-represented in the research sample whereas parity was comparable. The full sample and sub-sample did not vary on any characteristics.

[Table 2 (sample characteristics) about here]

Validation of the Whooley and Arroll questions

Responses to the Whooley and Arroll questions were only available via the handheld maternity notes (n=167; see Figure 1) and were uncompleted in five instances. Thirty women (18.5%) endorsed at least one Whooley item; the Arroll ‘help’ item was endorsed by six of these and uncompleted by three.

Using either Whooley item as the criterion for possible caseness (Table 3) had strong specificity (i.e. most women identified as non-cases using the EPDS are identified as non-cases using the Whooley questions) but identified only half the women identified using the commonly adopted EPDS thresholds (EPDS \geq 10: sensitivity 45.7%, specificity 92.1%; EPDS \geq 13: sensitivity 47.8%, specificity 86.1%). Agreement with the EPDS was

179 greater for the Whooley item concerning low mood than the item concerning anhedonia, with the latter leading
180 to more false positives, possibly reflecting somatic aspects of pregnancy.

181
182 Using the Arroll ‘help’ question as the test criterion (Table 4) improved specificity but substantially
183 compromised sensitivity, missing nine in ten possible cases (EPDS ≥ 10 : sensitivity 9.1%, specificity 98.2%;
184 EPDS ≥ 13 : sensitivity 9.5%, specificity 97.1%). Of the six women endorsing the Arroll item, four were
185 identified by the EPDS at the lower of the common thresholds. Details in health records indicated that the
186 responses of the remaining two women reflected somatic aspects (sickness and backache) rather than
187 psychological distress *per se*, suggesting that these were not possible cases ‘missed’ by the EPDS.

188
189 Regardless of test criterion, agreement was greatest at the more conservative EPDS threshold of ≥ 15 , but not
190 substantially so (Whooley as criterion: sensitivity 57.1%, specificity 84.9%; Arroll as criterion: sensitivity
191 16.7%, specificity 97.2%) and, due to the positive predictive value being linked to the prevalence of possible
192 cases in the population, performance was best at the lowest EPDS threshold.

193
194

195 *[Table 3 (Whooley vs EPDS) around here]*

196
197

[Table 4 (Arroll vs EPDS) around here]

198
199

200 *Analysis of women’s views and experiences*

201 Several themes emerged from the analysis; the theme *context of disclosure* is presented here to inform
202 understanding the validation findings and limited disclosure in a clinical context.

203
204 Women’s accounts illustrated that disclosure required women to ‘admit’ symptoms of distress, both to
205 themselves and to others; and that this was influenced by women’s views on the relevance of mental health to
206 maternity services. Such views were shaped by women’s individual understandings of maternal mental health,
207 the context of the appointment and the perceived purpose of assessment.

208
209

209 *Remit of maternity services*

210 Perceived relevance was shaped by perceiving that the emphasis of maternity care was “98% medical physical
211 thing and 2% emotional” (Lena, time 1). Thus, questions such as “How are you?” were interpreted as
212 concerning physical aspects to do with the pregnancy, rather than emotional aspects to do with the woman:

213
214

214 “They’re more interested in you medically ... they’re asking you, “How are you? How are you
215 feeling?” but it’s more, “Have you got any lumps and bumps and pains?” ... they’re not asking you
216 emotionally.” (Anne, time 1)

217

218 Some women felt that their psychological distress was “just personal circumstances” (Jess, time 1), and
219 therefore not a legitimate concern for midwives:

220

221 “I don’t feel I can turn round and go “Yeah, but there’s this that’s gone on and that that’s gone on” and
222 actually it’s unrelated to the pregnancy. I feel like, for them, they need to concentrate on the pregnancy
223 side of things really.” (Emily, time 2)

224

225 *Context of maternity appointments*

226 The context of appointments, both in terms of the nature of busy clinics and in relation to interactions with
227 health professionals, influenced women’s views on relevance of mental health to maternity services. Comments
228 about appointments referred both to the booking appointments, which in this sample took place in a hospital
229 antenatal clinic, and subsequent antenatal appointments either in the community or the hospital; all of which
230 involved consultations with midwives.

231

232 Women’s accounts highlighted a sense that there are too many tasks for the time available, with appointments
233 consequently feeling rushed and potentially limiting disclosure without the “time and space to actually go
234 through those things” (Charlotte, time1):

235

236 “It’s just like a conveyor belt. You’re in and you’re out. They’re just: blood pressure, check your water,
237 check the heartbeat, and then off. There’s no real conversation of how are you? So because I wasn’t
238 really asked, I didn’t speak about it.” (Louise, time 1)

239

240 Alongside the pace of appointments, it was the manner in which they were asked that mattered to women and
241 some felt that factors such as trust and confidence were more important for disclosures concerning mental health
242 and well-being than discussions of physical aspects of health. Although continuity was valued, this was
243 considered less important than skills such as “really listening” (Abbie, time 1) which were contrasted with
244 interactions that felt “a little bit false” (Abbie, time 2), as though they “were going through the motions of it”
245 (Charlotte, time 1) with “bullet type things that they have to ask” (Helen, time 2). Some women also described
246 feeling that staff seemed to lack confidence and felt uncomfortable in discussing mental health.

247

248 *Understandings of maternal mental health*

249 Disclosure of symptoms were also influenced by women’s personal understandings of maternal mental health
250 and several described struggling to determine whether their feelings were “normal” (Louise, time 2); here, some
251 women felt that screening questions helped them to recognise to themselves that they were struggling. Women
252 could however feel deterred from seeking support because assessing symptoms and severity was “so subjective”
253 (Katie, time 1) but also because women needed themselves to be “at the stage where you’ve thought about,
254 “yeah, I could really do with some support” ” (Hannah, time 1).

255

256 *Purpose of assessment and implications of disclosure*

257 Admitting to self and others was influenced by the implications of disclosure. This extended beyond stigma
258 (indicated by terms such as “loony bin”, “bonkers”, “crazy”, “bring branded”) and was more concerned with the
259 perceived purpose of assessment. Women’s accounts suggested great uncertainty around implications:

260
261 “If I tick yes [to the Arroll item], what does that mean, what’s going to happen?” (Emily, time 1)

262
263 Some women held concerns about possible treatment options, both pharmacological and psychological, that
264 could deter them from seeking help. Several women felt that maternity services could, theoretically, be in a
265 position to help with early intervention, most felt that, in reality, the purpose of assessment was to identify risk
266 of harm:

267
268 “The only question that she [health visitor] was more worried about is, would I self-harm or hurt the
269 baby. I went “no”. That’s all she was more worried about, not dealing with the fact that, why am I
270 upset?” (Rebecca, time 1)

271
272 “Unless you’ve been suffering from sort of psychosis, you’re not gonna get any *real*, you know, *service*
273 or support from anywhere anyway. It’s always like “worst case scenario then we will help you”.
274 (Michelle, time 1)”

275
276 Women were sometimes therefore either wary of potential social services involvement or simply cynical about
277 health professionals’ ability to do anything to help them address their underlying causes of distress (Abbie, time
278 3; Katie, time 2).

279
280

281 Discussion

282 This is the first study to offer validation of Whooley questions and Arroll ‘help’ item in UK clinical practice.
283 Contrary to concerns that clinical use of the Whooley questions may unnecessarily over-burden systems through
284 high rates of false positives, they were found to identify only half of women identified by the EPDS completed
285 in a research context. Sensitivity substantially worsened by reliance on the Arroll ‘help’ item, missing nine in
286 ten possible cases identified using the EPDS.

287
288 Performance was far poorer in the current study than reported elsewhere (Bennett et al. 2008; Mann and
289 Gilbody 2011; Smith et al. 2010). The EPDS does not offer diagnosis and is itself therefore vulnerable to issues
290 of sensitivity and specificity; however, this does not explain the poor sensitivity because stronger performance
291 has been found both for validation against diagnostic clinical interview (Mann and Gilbody 2011; Smith et al.
292 2010) and against the EPDS (Bennett et al. 2008). Our finding is also unlikely to be due to gestational age at
293 assessment, as the mean age is similar in the current study and the study by Bennett et al. (2008), as is the
294 percentage of women scoring above threshold (respectively 14.4% and 17.4%, using a threshold of ≥ 13).

295

296 Analysis of women's accounts indicates that a likely explanation for the poor sensitivity found is that women
297 under-disclosed when completing the Whooley and Arroll questions and primarily because of the context of
298 disclosure; which could similarly inhibit disclosure if the EPDS were completed in this manner. In our study the
299 Whooley and Arroll questions were self-completed as part of routine clinical care; the EPDS was also self-
300 completed, but as part of a research study. In contrast, both measures were completed with a physician or nurse
301 in the IMPLICIT network study (Bennett et al. 2008) and both were completed in a research context for the
302 other studies (Mann and Gilbody 2011; Smith et al. 2010).

303

304 Women's accounts conveyed that the manner in which mental health was discussed was considered more
305 important than the exact phrasing used to ask the depression questions, illustrating the need to provide an
306 enabling environment to ensure the process is both acceptable to the population and effective. Thus, rather than
307 endorsing routine use of the EPDS in preference to the Whooley and Arroll questions, this study speaks to the
308 significance of the context of disclosure for mental health assessment which is relevant regardless of the
309 measure used, the setting or healthcare professional involved. The need for enabling environments and
310 challenges around implementation echo those raised when routine enquiry for domestic abuse was introduced
311 (Taket et al. 2003). Ensuring an enabling environment includes addressing consultation-level factors such as
312 time limitations and work pressures that impact patient-centredness (Mead and Bower 2000) and can influence
313 women's help seeking for depression in various maternity settings (Bennett et al. 2009). It is unsurprising that
314 we found parallels with the literature on acceptability of the EPDS. Authors of a review on EPDS acceptability
315 concluded that although the EPDS was "generally acceptable" there could be issues around its administration
316 and they considered the clinic setting "too distracting and uncomfortable for women", instead recommending
317 completion at home, affording more privacy and time (Brealey et al. 2010).

318 Alongside consultation-level factors, staff need the appropriate training and skills in psychological assessment.
319 Low staff confidence in handling perinatal mental health has been reported amongst midwives in the UK (King
320 et al. 2012); similarly, in Australia where mental health assessment is also carried out at booking by midwives,
321 training needs have been identified, including knowledge of perinatal mental health and resources available to
322 women and staff (McCauley et al. 2011). Some women in the current study picked up on staff discomfort and
323 lack of confidence, linked to the perception that mental health is not the remit of maternity services. These
324 findings resonate with a North American study that reported women perceived a lack of mental health expertise
325 amongst obstetricians (i.e. those healthcare professionals who would be expected to undertake mental health
326 assessment during the perinatal period) and that this acted as a potential barrier to depression help seeking
327 (Bennett et al. 2009).

328

329 Perceived relevance also included the purpose of assessment. Women's concerns around implications of
330 disclosure, including others' views of parenting ability and potential involvement of social services, have been
331 raised in relation to the EPDS and women's ability to answer depression screening questions honestly (Brealey
332 et al. 2010). However, unlike the EPDS, the Arroll approach asks women directly about wanting or needing help
333 and our study demonstrated concerns and uncertainty amongst some women about possible implications of
334 reporting this to a healthcare professional; here, a midwife. This is consistent with our quantitative data
335 indicating extremely poor sensitivity of the Arroll question. The finding that reliance on the Arroll 'help' item

336 may risk false negatives is consistent with the findings of Mann and Gilbody (2011), but is considerably more
337 marked in our study. Such findings suggest that, whereas concerns over high false positive rates are consistently
338 raised in respect of ultra-brief screening tools (Mitchell and Coyne 2007) and national screening programmes
339 for postnatal depression have been advised against mainly due to the costs of false positives (Hewitt et al.
340 2009), of equal concern may be high rates of false negatives and the potential for delayed access to interventions
341 (Martin and Redshaw 2009). The guidelines for England and Wales (National Collaborating Centre for Mental
342 Health 2007) position the Whooley and Arroll items as the first assessment stage, to be followed up further
343 assessment. Such two-stage processes require strong sensitivity in the first step to avoid ‘missing’ potential
344 cases (Bennett et al. 2008; Gjerdingden et al. 2009), yet this is compromised by the use of the ‘help’ item in its
345 current format and women’s uncertainty around the purpose of assessment.

346
347 The validation component of this study had two main methodological limitations: i) threats to internal validity
348 by lacking comparison with diagnostic interviews and ii) threats to external validity due to sampling constraints.
349 Analysing the validation data alongside women’s accounts offered alternative perspectives and richer
350 understandings through considering the context of disclosure, illustrating the potential benefit of integrating
351 mixed methods in the analysis stage to provide an end product greater than the constituent parts (Bryman 2007;
352 Moran-Ellis et al. 2006). Although qualitative research does not have the same need for representativeness, it is
353 important to acknowledge the views that are being represented. Interviews were limited to those with high levels
354 of maternal stress as defined by the chosen measures and acceptability may be different in those below and
355 above threshold. In addition, the findings are taken from one local unit and, within the sample, White British
356 women and older women were over-represented; care must therefore be taken in extending the findings beyond
357 the study.

358
359

360 **Conclusion**

361 Contrary to concerns about the numbers of false positives encountered when using ultra-brief mental health
362 assessment, this study suggests that the greater concern when administering the Whooley and Arroll questions in
363 antenatal care is the number of false negatives. A mixed methods approach illustrated the significance of context
364 of disclosure for psychometric properties when measures developed in research settings are adopted in clinical
365 practice. Further research is needed to validate the use of this approach in maternity care and to determine the
366 optimal approach to identifying possible depression in pregnancy; this extends beyond the instrument of choice
367 to include enabling environments and subsequent management. Meanwhile, health professionals and policy
368 makers should be aware that while the Whooley questions offer a simple and quick means of identifying women
369 who need support, they fail to identify a substantial proportion of women.

370
371

372 **Acknowledgements**

373 The lead author was supported by the University of Manchester Strategic Studentship Award which was co-
374 funded by the Medical Research Council and Tommy’s Baby Charity. The work has been presented in the lead
375 author’s doctoral thesis (Darwin 2012). The doctoral thesis was awarded the Annual Doctoral Thesis Award by

376 the Society for Reproductive and Infant Psychology and the work was presented at a prize lecture at the
377 Society's Annual Conference, 2013. We wish to thank the women who took part in the study and acknowledge
378 the support of the clinical and administrative staff.

379

380

381 **Conflict of interest**

382 The authors declare that they have no conflict of interest.

383

384

385 **Ethical standards**

386 The study received favourable ethical opinion from the Greater Manchester East Research Ethics Committee
387 (10/H1013/12) and relevant governance approval from the hospital, and was performed in accordance with the
388 ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.

389

390

391

392

393

394 **References**

- 395 Altman DG, Bland M (1994a) Diagnostic tests 1: sensitivity and specificity. *British Medical Journal* 308:1552
- 396 Altman DG, Bland M (1994b) Diagnostic tests 2: predictive values. *British Medical Journal* 309:102
- 397 American College of Obstetricians and Gynecologists (2006) Psychosocial risk factors: Perinatal screening and
398 intervention. ACOG Committee Opinion Number 343. *Obstetrics and Gynecology* 108:469-477
- 399 American College of Obstetricians and Gynecologists (2010) Screening for depression during and after
400 pregnancy. Committee Opinion No. 453. *Obstetrics and Gynecology* 115:394-395
- 401 Arroll B, Khin N, Kerse N (2003) Screening for depression in primary care with two verbally asked questions:
402 cross sectional study. *British Medical Journal* 327:1144-1146
- 403 Austin MP (2003) Psychosocial assessment and management of depression and anxiety in pregnancy: Key
404 aspects of antenatal care for general practice. *Australian Family Physician* 32:119-126
- 405 Austin MP (2004) Antenatal screening and early intervention for “perinatal” distress, depression and anxiety:
406 Where to from here? *Archives of Women's Mental Health* 7:1-6
- 407 Austin MP, Hadzi-Pavlovic D, Saint K, Parker G (2005) Antenatal screening for the prediction of postnatal
408 depression: validation of a psychosocial Pregnancy Risk Questionnaire. *Acta Psychiatrica Scandinavica*
409 112:310-317
- 410 Austin MP, Hight N, the Guideline Expert Advisory Committee (2011) The beyondblue clinical practice
411 guidelines for depression and related disorders, anxiety, bipolar disorder and puerperal psychosis in the
412 perinatal period. A guideline for primary care health professionals providing care in the perinatal
413 period. Beyond Blue: The National Depression Initiative, Melbourne
- 414 Bennett I et al. (2008) Efficiency of a two-item pre-screen to reduce the burden of depression screening in
415 pregnancy and postpartum: an IMPLICIT Network study. *Journal of the American Board of Family*
416 *Medicine* 21:317-325
- 417 Bennett IM et al. (2009) “One end has nothing to do with the other:” Patient attitudes regarding help seeking
418 intention for depression in gynecologic and obstetric settings. *Archives of Women’s Mental Health*
419 12:301-308
- 420 Brealey SD, Hewitt C, Green JM, Morrell J, Gilbody S (2010) Screening for postnatal depression - is it
421 acceptable to women and healthcare professionals? A systematic review and meta-synthesis *Journal of*
422 *Reproductive and Infant Psychology* 28:328-344 doi:10.1080/02646838.2010.513045
- 423 Bryman A (2007) Barriers to integrating quantitative and qualitative research. *Journal of Mixed Methods*
424 *Research* 1:8-22
- 425 Carroll JC et al. (2005) Effectiveness of the antenatal psychosocial health assessment (ALPHA) form in
426 detecting psychosocial concerns: A randomized controlled trial *Canadian Medical Association Journal*
427 173:253-259
- 428 Centre for Maternal and Child Enquiries (2011) Saving Mothers’ Lives: reviewing maternal deaths to make
429 motherhood safer: 2006–08. The Eighth Report on Confidential Enquiries into Maternal Deaths in the
430 United Kingdom. *British Journal of Obstetrics and Gynaecology* 118 (Supplement 1):1-203
- 431 Cox J, Holden J (2003) Perinatal mental health: a guide to the Edinburgh Postnatal Depression Scale (EPDS).
432 Gaskell, London

433 Cox J, Holden J, Sagovsky R (1987) Detection of postnatal depression: Development of a 10-item Edinburgh
434 Postnatal Depression Scale. *British Journal of Psychiatry* 150:782-786

435 Darwin Z (2012) Assessing and Responding to Maternal Stress (ARMS): Antenatal Psychosocial Assessment in
436 Research and Practice., University of Manchester

437 Gavin NI, Gaynes BN, Meltzer-Brody S, Gartlehner G (2005) Perinatal depression: A systematic review of
438 prevalence and incidence. *Obstetrics and Gynecology* 106:1071-1083

439 Gjerdingden D, Crow S, McGovern P, Miner M, Center B (2009) Postpartum depression screening at well-child
440 visits: Validity of a 2-question screen and the PHQ-9 *Annals of Family Medicine* 7:63-70

441 Hewitt CE et al. (2009) Methods to identify postnatal depression in primary care: An integrated evidence
442 synthesis and value of information analysis. *Health Technology Assessment* 13

443 King L, Pestell S, Farrar S, North N, Brunt C (2012) Screening for antenatal psychological distress. *British*
444 *Journal of Midwifery* 20:396-401

445 Kroenke K, Spitzer RL, Williams JBW (2001) The PHQ-9: validity of a brief depression severity measure.
446 *Journal of General Internal Medicine* 16:606-613

447 Kroenke K, Spitzer RL, Williams JBW (2003) The Patient Health Questionnaire-2: Validity of a two-item
448 depression screener. *Medical Care* 41:1284-1292

449 Kroenke K, Spitzer RL, Williams JBW, Monahan PO, Lowe B (2007) Anxiety disorders in primary care:
450 prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine* 146:317-325

451 Lincoln Y, Guba E (1985) *Naturalistic inquiry*. Sage, Thousand Oaks, CA

452 Mann R, Gilbody S (2011) Validity of two case finding questions to detect postnatal depression: A review of
453 diagnostic test accuracy. *Journal of Affective Disorders* 133:388-397

454 Martin CR, Redshaw M (2009) Carry on screening *Journal of Reproductive and Infant Psychology* 27:327-329

455 Matthey S (2004) Detection and treatment of postnatal depression (perinatal depression or anxiety). *Current*
456 *Opinion in Psychiatry* 17:21-29

457 Matthey S, Henshaw C, Elliott S, Barnett B (2006) Variability in use of cut-off scores and formats on the
458 Edinburgh Postnatal Depression Scale - implications for clinical and research practice. *Archives of*
459 *Women's Mental Health* 9:309-315

460 McCauley K, Elsom S, Muir-Cochrane E, Lyneham J (2011) Midwives and assessment of perinatal mental
461 health. *Journal of Psychiatric and Mental Health Nursing* 18:786-795

462 Mead N, Bower P (2000) Patient centredness: a conceptual framework and review of the empirical literature.
463 *Social Science and Medicine* 51:1087-1110

464 Mitchell AJ, Coyne JC (2007) Do ultra-short screening instruments accurately detect depression in primary
465 care? A pooled analysis and meta-analysis of 22 studies. *British Journal of General Practice* 57:144-
466 151

467 Moran-Ellis J, Alexander VD, Cronin A, Dickinson M, Fielding J, Slaney J, Thomas H (2006) Triangulation
468 and integration: Processes, claims and implications. *Qualitative Research* 6:45-59

469 Murray D, Cox JL (1990) Screening for depression during pregnancy with the Edinburgh Depression Scale
470 (EPDS). *Journal of Reproductive and Infant Psychology* 8:99-107

471 National Collaborating Centre for Mental Health (2007) Antenatal and postnatal mental health. The NICE
472 guideline on clinical management and service guidance. The British Psychological Society and The
473 Royal College of Psychiatrists, Leicester

474 Ritchie J, Spencer L (1994) Qualitative data analysis for applied policy research. In: Bryman A, Burgess R (eds)
475 Analysing qualitative data. Routledge, London, pp 173-194

476 Robertson E, Grace S, Wallington T, Stewart DE (2004) Antenatal risk factors for postpartum depression: A
477 synthesis of recent literature. *General Hospital Psychiatry and Clinical Neurosciences* 26:289-295

478 Scottish Intercollegiate Guidelines Network (2012) Management of perinatal mood disorders. Scottish
479 Intercollegiate Guidelines Network, Edinburgh

480 Smith MV, Gotman N, Lin H, Yonkers KA (2010) Do the PHQ-8 and the PHQ-2 accurately screen for
481 depressive disorders in a sample of pregnant women? *General Hospital Psychiatry* 32:544-548

482 Spielberg CD, Gorusch RL, Lushene RE (1987) *The State-Trait Anxiety Inventory: Test Manual*. Consulting
483 Psychological Press, Palo Alto

484 Taket A et al. (2003) Routinely asking women about domestic violence in health settings. *British Medical*
485 *Journal* 327:673-676

486 Teddlie C, Yu F (2007) Mixed methods sampling: A typology with examples. *Journal of Mixed Methods*
487 *Research* 1:77-100

488 Whooley MA, Avins AL, Miranda J (1997) Case-finding instruments for depression. Two questions are as good
489 as many. *Journal of General Internal Medicine* 12:439-445

490
491
492
493

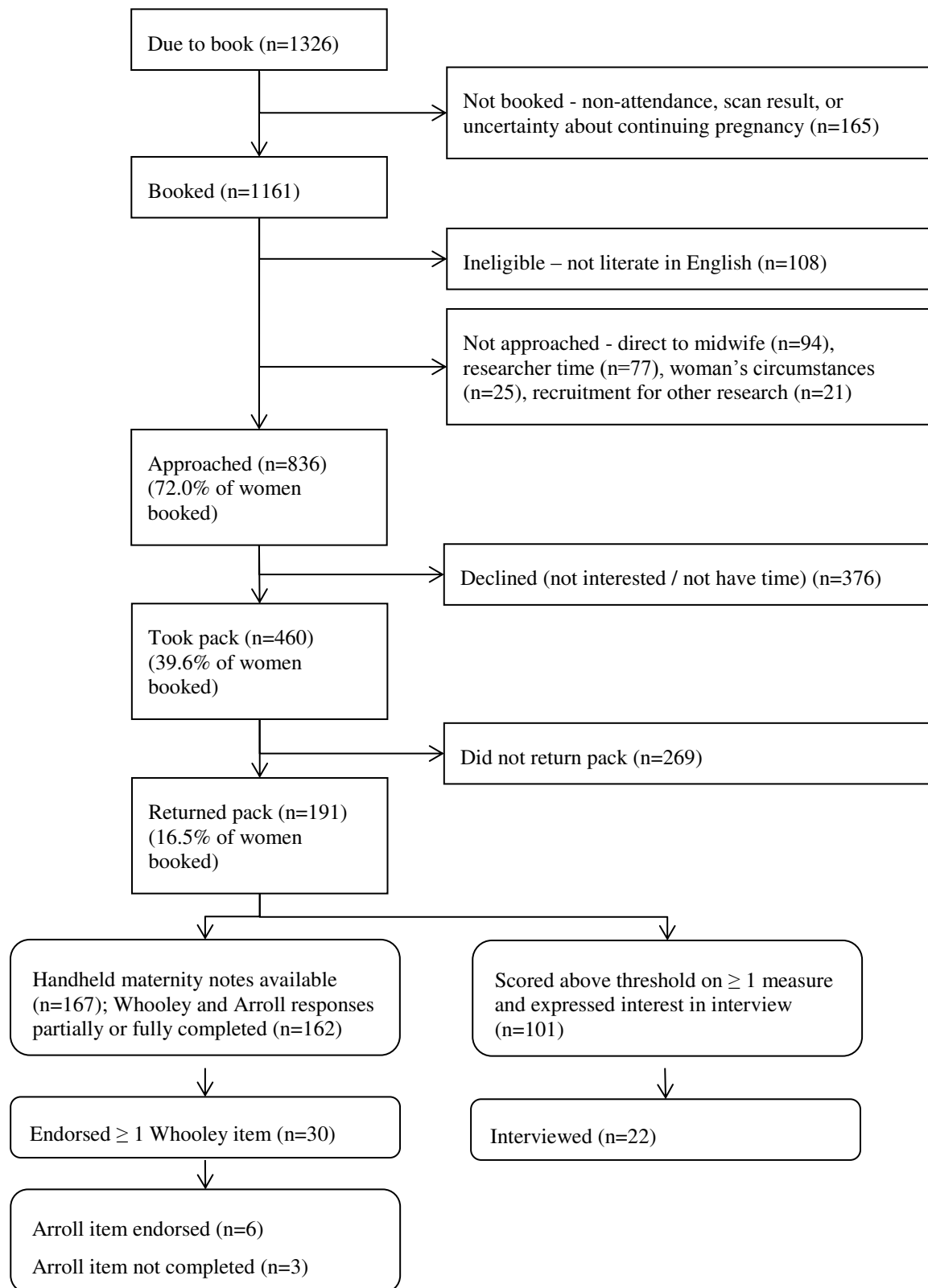


Table 1 Summary of published literature on validation of the Whooley questions and original PHQ-2 in antenatal populations

Study	Details	Measures of performance			
		sensitivity	specificity	PPV	NPV
- Author - Country	- Whooley 'test' criterion - 'gold standard' comparison - gestational age - sample size				
Bennett et al 2008 USA	Whooley (yes to either item) EPDS \geq 13 15 weeks n=414	93	75	44	98
Bennett et al 2008 USA	Whooley (yes to either item) EPDS \geq 13 30 weeks n=334	82	80	24	91
Smith et al 2010 USA	PHQ-2 (\geq 3) Diagnostic interview Before 17 weeks n=214	59	77	n/r	n/r
Smith et al 2010 USA	PHQ-2 (\geq 4) Diagnostic interview Before 17 weeks n=214	62	79	n/r	n/r
Mann et al 2012 UK	Whooley (yes to either item) Diagnostic interview 26-28 weeks n=126	100	68	33	99
Mann et al 2012 UK	Arroll (yes) Diagnostic interview 26-28 weeks	58	91	77	82

	n=126				
--	-------	--	--	--	--

501 Notes: sensitivity = proportion of women who are possible cases (based on the EPDS) who are identified as
502 possible cases (using Whooley/Arroll); specificity = proportion of women who are non-cases (based on the
503 EPDS) who are identified as being non-cases (using Whooley/Arroll); NPV (Negative Predictive Value) =
504 proportion of women with negative test result (on Whooley/Arroll) who are correctly classified as non-cases;
505 PPV (Positive Predictive Value) = proportion of women with positive test result (on Whooley/Arroll) who are
506 correctly classified as possible cases; n/r = not reported; diagnostic interviews were completed approximately
507 two weeks after completion of the Whooley questions and original PHQ-2; sample size is the number for which
508 both data sets were available, not the number recruited
509
510

511 **Table 2 Sample characteristics**

	Full sample completing research questionnaire (n=191)	Sub-sample interviewed (n=22)
Age (years)	mean 31.1 sd 5.3 (19-46)	mean 31.7 sd 4.2 (26-39)
Ethnicity	129 (67.9%) White British	17 (77.3%) White British
In a relationship	174 (91.1%)	20 (90.9%)
Primigravida (first pregnancy)	71 (37.2%)	7 (31.2%)
Primipara (first birth)	111 (58.1%)	9 (40.9%)
Gestation (weeks) at booking	mean 13 sd 5.4 (8-38) 144 (75.4%) 1st trimester	mean 13 sd 2.8 (8-20) 15 (68.2%) 1st trimester
Timing of interviews (weeks)	not applicable	<i>Antenatal</i> Time 1: mean 16 sd 2.8 (10-22) Time 2: mean 33 sd 1.7 (28-36) <i>Postnatal</i> Time 3: mean 10 sd 1.4 (7-13)

Table 3 Validation of the Whooley questions against the EPDS, using yes to either item as case criterion (n=160)

EPDS threshold	Whooley (either item)		Measures of performance			
	No (n=130)	Yes (n=30)	Sensitivity	Specificity	PPV	NPV
< 10 (n=114)	105 (65.6)	9 (5.6)	21/46 (45.7)	105/114 (92.1)	21/30 (70.0)	105/130 (80.8)
≥ 10 (n=46)	25 (15.6)	21 (13.1)				
< 13 (n=137)	118 (73.8)	19 (11.9)	11/23 (47.8)	118/137 (86.1)	11/30 (36.7)	118/130 (90.8)
≥ 13 (n=23)	12 (7.5)	11 (6.9)				
< 15 (n=146)	124 (77.5)	22 (13.8)	8/14 (57.1)	124/146 (84.9)	8/30 (26.7)	124/130 (95.4)
≥ 15 (n=14)	6 (3.8)	8 (5.0)				

Notes: NPV = Negative Predictive Value; PPV = Positive Predictive Value

Table 4 Validation of the Whooley questions against the EPDS, using Arroll ‘help’ item as case criterion (n=157)*

EPDS threshold	Arroll ‘help’ item		Measures of performance			
	No (n=151)	Yes (n=6)	Sensitivity	Specificity	PPV	NPV
< 10 (n=113)	111 (70.7)	2 (1.3)	4/44 (9.1)	111/113 (98.2)	4/6 (66.7)	111/151 (73.5)
≥ 10 (n=44)	40 (25.5)	4 (2.5)				
< 13 (n=136)	132 (84.1)	4 (2.5)	2/21 (9.5)	132/136 (97.1)	2/6 (33.3)	132/151 (87.5)
≥ 13 (n=21)	19 (12.1)	2 (1.3)				
< 15 (n=145)	141 (89.8)	4 (2.5)	2/12 (16.7)	141/145 (97.2)	2/6 (33.3)	141/151 (93.4)
≥ 15 (n=12)	10 (6.4)	2 (1.3)				

Notes: NPV = Negative Predictive Value; PPV = Positive Predictive Value

* EPDS scores were not available for two women