

27 Abstract

28 Adult cochlear implant (CI) candidacy is assessed in part by the use of speech perception
29 measures. In the United Kingdom the current cut-off point to fall within the CI candidacy
30 range is a score of less than 50% on the BKB sentences presented in quiet (presented at 70
31 dB SPL).

32 The specific goal of this article was to review the benefit of adding the AB word test to the
33 assessment test battery for candidacy. The AB word test scores showed good sensitivity and
34 specificity when calculated based on both word and phoneme scores. The word score
35 equivalent for 50% correct on the BKB sentences was 18.5% and it was 34.5% when the
36 phoneme score was calculated; these scores are in line with those used in centres in Wales.
37 The goal of the British Cochlear Implant Group (BCIG) service evaluation was to determine
38 if the pre-implant assessment measures are appropriate and set at the correct level for
39 determining candidacy, the future analyses will determine whether the speech perception
40 cutoff point for candidacy should be adjusted and whether other non-speech based measures
41 should be used in the candidacy evaluation.

42

43 Introduction

44 Current practice in the United Kingdom (UK) for assessing candidacy for cochlear implants
45 (CIs) was determined based on research conducted over ten years ago (UK study group
46 2004). The project involved the collaboration of 13 UK CI centres working with academics
47 from the Medical Research Council Institute of Hearing Research to determine the
48 appropriate limits for candidacy for CIs in adults. All study centres used the same testing
49 apparatus, and the same standard measures for assessing CI candidacy. This included the
50 Bamford, Kowal and Bench (BKB) sentences presented in quiet and without any visual
51 information from the talker's face or lips.

52

53 The vocabulary used in the development of the BKB sentences was taken from the utterances
54 of eight to fifteen year old partially hearing children when describing everyday scenarios. The
55 sentences could therefore be used with the majority of implant users pre-and post-implant.
56 The scoring of BKBs was based on the percentage of keywords correct; the recommended
57 cut-off criterion for falling within the CI candidacy range was set at <50% correct. This
58 criterion was adopted by the National Institute for Health and Care Excellence (NICE) in
59 2009 (NICE, 2009) and is in place today.

60

61 The use of monosyllabic word tests or speech in noise assessments to assess outcomes post-
62 implantation was uncommon in the UK in 2004, because the post-operative performance of
63 implantees listening to sentences in quiet was rarely, if ever, constrained by ceiling effects.
64 With improvements in patient outcomes over the years, the assessment of adult patients has
65 become more sophisticated and sensitive to changes in post-operative performance levels.
66 These changes in outcome over the years have been attributed to changes in candidacy,
67 surgical practice, clinical intervention and device development (Blamey et al., 2013).

68

69 In the UK, CI teams struggle with the current candidacy criteria because they do not permit
70 all individuals whom they currently consider as likely candidates to be implanted. In
71 particular, two issues arise from the use of BKB sentences alone for assessing candidacy. The
72 first issue is that they use highly predictable materials. Some candidates may only have
73 minimal access to speech cues but because they have refined cognitive skills, they are able to
74 work out the content of the sentence in spite of the severe loss of auditory information. While
75 this skill is unlikely to benefit their speech perception in everyday complex listening
76 situations, it may result in too high a test score for them to be eligible for implantation.

77

78 The second issue is that BKB sentence scores do not reflect the hearing abilities of some
79 patients at the lower end of the performance scale. Individuals who do not speak English as
80 their primary language may not have the same abilities as a native English speaker to fill in
81 gaps in predictable sentences because their linguistic knowledge is poorer. The same issue
82 arises in pre-lingually deaf adults who may not have a high levels of language sufficient to be
83 able to get a representative score on the BKB sentences. Craddock et al. (2016) suggested
84 that the CUNY audio-visual sentences might be more appropriate for this group.

85

86 Doran and Jenkinson (2016) reported on the additional criteria used by the Welsh CI teams.
87 Individuals who have borderline BKB scores (<60%) are evaluated with a second measure,
88 the Arthur Boothroyd (AB) monosyllabic word test. They based their cut-off levels on the
89 10th percentile of post-operative scores for both BKB sentences (50%) and AB words (15%)
90 and used these as their criteria. The advantages of the AB words are that they are a better
91 measure of speech audibility and not subject to the predictability effects associated with the
92 BKB sentences. However, just like the BKB sentences, the confidence intervals for AB
93 words scores are also extremely wide and dependent upon the number of items used
94 (Thornton and Raffin, 1978) and setting an absolute single value for candidacy may not be
95 appropriate. For example, a patient whose 'true' performance level is at 50% on the BKB
96 sentences may actually achieve scores between approximately 40-60%. In Wales, the
97 additional rule is applied for individuals who score greater than 50% on the BKB sentences,
98 if their AB word score is less than 15% then they are considered to be potential candidates for
99 CI.

100

101 Currently, the recommendation for scoring the AB words is to use the “word” score; i.e. the
102 proportion of whole words reported correctly. However, a score based on the proportion of
103 phonemes reported correctly (“phoneme” score) may give a finer resolution of performance
104 ability. This approach to scoring may also be more appropriate for those patients with English
105 as a second language or pre-lingually deafened patients in whom the capacity to report whole
106 words may not relate well to audibility.

107

108 It is known that there is a strong linear relationship between BKB sentence and the AB
109 word scores (Vickers et al., 2009). The use of both assessments to verify candidacy has great
110 potential for improving the quality and reliability of the candidacy assessments. An additional
111 assessment provides greater clinical confidence in the eligibility of an individual because of
112 this “second look” at performance. Exploratory analyses were therefore conducted to assess
113 whether the 15% AB word score described by Doran and Jenkinson (2016) determines
114 eligibility in a similar way to the 50% BKB sentences score.

115

116 Ten UK CI centres have been engaged in collecting clinical data to determine if the BKB
117 sentence test is still the most appropriate pre-operative measure for determining candidacy for
118 CIs in the UK. The ultimate goal of the data collection is to determine if the actuarial model
119 for candidacy for CIs in adults in the UK should be updated. However, the specific goal of
120 this phase of the project was purely an exploratory analysis to evaluate the potential
121 usefulness of the AB words.

122

123 **Method**

124 **Study participant inclusion criteria**

125 There were 134 data sets collated by December 1st 2015. Ninety seven of those had scores
126 for both BKB sentences and AB words in quiet; 37 were missing data on one or both
127 measures. This represented all data sets entered and verified at the point of analysis.
128 Participants had all been seen by a participating adult implant programme and were able to
129 undergo speech perception testing. Each centre registered the data collection with their
130 individual Hospital Research and Development boards as part of a multi-site service
131 evaluation for the British Cochlear Implant Group.

132

133 **Speech perception test delivery**

134 The BKB sentences and AB words were delivered in sound-treated rooms at the individual
135 hospitals and presented at 70 dBSPL. The BKB sentences were scored by the number of key
136 words correct and the AB word test was scored by both phoneme and word.

137

138 **Demographic information**

139 Data were also *collated* ?collected? on home language, aetiology, duration of profound
140 deafness and age at implantation.

141

142 **Analysis**

143 The scores on these measures were evaluated in two ways. First, the scores for all participants
144 were reviewed to determine if their AB words or BKB sentences fell within the candidacy
145 range proposed by Doran and Jenkinson (2016); i.e. <50% on BKB sentences and/or <15%
146 on AB words. If there were any anomalies the demographic factors were explored to
147 determine if the findings could be explained and highlighting areas to be included in the final
148 actuarial model.

149

150 Secondly, comparisons of the AB word and BKB sentence scores were made using Receiver
151 Operating Characteristic (ROC) curve analysis. Patients were first classified as either eligible
152 or ineligible based on their BKB sentences score, where eligible patients were those with a
153 score less than 50% correct, in line with NICE guidance. An ROC curve was then fitted to
154 this binary variable and their corresponding scores on the AB words test. Two analyses were
155 performed: one using the word score and another using the phoneme score. These ROC
156 curves were fit and compared using the pROC package for the R statistical programming
157 environment (Robin et al., 2011). The ‘optimal’ word and phoneme cut-off scores for the AB
158 word test were determined as those scores that simultaneously maximised the sensitivity and
159 specificity of the test. In this context, sensitivity refers to the test’s ability to identify eligible
160 patients and specificity to its ability to identify ineligible patients.

161

162 **Results**

163 The mean age at implantation was 57 years (median 61 years; range 16 to 88 years) and the
164 mean duration of deafness was 22.5 years (median 15 years; range 0 to 80 years). Eight
165 individuals had English as an additional language (EAL) and five people used British Sign
166 Language as their primary language. Table 1 shows the number of individuals who would
167 pass or fail the BKB sentences (pass cut off level was $\geq 50\%$) or the AB words (pass cut off
168 level was $>15\%$). If the AB words were used as the only candidacy measure fewer people
169 would fall within the candidacy range (78 people) versus 85 people with the BKB sentence
170 criteria. If both assessments were used together an additional one person would be considered
171 for implantation on top of the current BKB cutoff at 50%.

172 -----

173 Table 1 here

174 -----

175 Table 2 provides the demographic details of the individuals who had different pass/fail
176 criteria based on the two measures.

177 -----

178 Table 2 here

179 -----

180 The single individual whose scores would fall within candidacy with the use of AB words is
181 shown on the first row of table 2. This individual was pre-lingually deafened and had a
182 conductive component to the hearing loss. This individual was able to detect phonemes but
183 could not identify the words. This individual did remarkably well with the BKB sentences
184 maybe because of the vocabulary level of the materials being appropriate for children. The
185 eight other individuals were those who passed the AB words and failed the BKB sentences,
186 these individuals either did not have English as their primary language or were older than the
187 mean age of the group. Both aspects might make the audibility of phonemes and words easier
188 than understanding the BKB sentences, where gap filling might be necessary to correctly
189 perceive the sentences. Within the dataset forty individuals were over the UK retirement age
190 (65 years old), highlighting the importance of accurately defining the appropriate assessment
191 tools for older adults.

192 It should also be noted that for some tasks the scores were borderline (between 50 and 60%
193 on BKB sentences (6 people) and between 15 and 33% on AB words (11 people)).

194

195 Figure 1 plots the ROC curves for the word and phoneme scores derived from the AB words
196 test. It would be straightforward to choose a very high AB word or phoneme score as a cut-
197 off that would successfully identify all patients with BKB scores <50%; i.e. eligible patients.
198 Similarly, it would be easy to choose a very low word/phoneme score as a cut-off that would
199 successfully identify all patients with BKB scores \geq 50%; i.e. ineligible patients. ROC curves

200 are used to identify the cut-off score that tries to do both of these tasks simultaneously; i.e.
201 the score that would classify patients as eligible or ineligible as similarly as possible to the
202 'gold standard' of <50% on the BKB sentence test. The closer the curves are to the top-left
203 corner of the Figure 1, the more similar they are to the BKB 50% cut-off at determining
204 candidacy.

205

206 Both word and phoneme scores showed high sensitivity (accuracy with identifying eligible
207 patients) and specificity (accuracy with identifying ineligible patients). The optimal cut-off
208 word score was found to be 18.5% (sensitivity 90.9%, specificity 96.3%) and the optimal
209 phoneme score was 34.5% (sensitivity 100%, specificity 93.0%). A paired comparison of the
210 two curves using DeLong's test (DeLong et al., 1988) revealed no significant difference ($Z=-$
211 1.02 , $p=0.31$), suggesting that both word and phoneme scores assessed candidacy in a similar
212 way. **I FOUND THIS RATHER DIFFICULT TO UNDERSTAND! DO YOU THINK**
213 **KIT MIGHT BE POSSIBLE TO EXPLAIN IT MORE CLEARLY TO THOSE NOT**
214 **FAMILIAR WITH ROC CURVES AND THEIR USE?!**

215

216 -----

217 Figure 1 here

218 -----

219

220 Discussion

221 This exploratory analysis was conducted to determine if the data suggests that the addition of
222 the AB words to the candidacy test battery or even the replacement of BKB sentences with
223 AB words would be appropriate. The candidacy criteria recommended by Doran and
224 Jenkinson (2016) were used to determine how the numbers of candidates would change if the
225 AB word criteria were added to the BKB criteria or if the AB words were used alone. The

226 analysis suggested that fewer people would fall within the candidacy range if only the AB
227 words (<15%) were used. This might suggest that the relationship between the two measures
228 was not appropriately set. Further exploratory analyses based on ROC curve analysis
229 produced an AB word cutoff (18.5%) which was marginally higher but very similar to the
230 15% suggested. This shows that the BKB and AB tests produce quite consistent data when
231 assessing candidacy across different CI centres and patient catchment areas.

232

233 At the current 50% BKB cut-off, both word and phoneme scores from the AB word test seem
234 to be highly sensitive and specific. This suggests that they may be suitable replacements/or
235 additions to the test battery for those individuals where there might be concerns over the
236 validity of the BKB sentence test; e.g. EAL, pre-lingually deafened, first-language BSL user.

237

238 Current results suggest that it may be productive to use ROC analysis to compare BKB and
239 AB word tests. The future analysis of the BCIG service evaluation data should replicate the
240 current approach with a larger sample of data. Both pre and post-implantation data can be
241 used to ensure that the entire performance range is covered. A larger sample would also allow
242 the analysis to be conducted on only the subset of post-lingually deafened patients with
243 English as a first language to eliminate any concerns about the validity of BKB scores. The
244 cut-off produced by that data could then be applied to the sub-population of patients where
245 BKB scores may not be valid to see whether AB word scores produce a similar proportion of
246 eligible candidates.

247

248 This analysis has demonstrated that the AB words may have potential as a measure in the
249 candidacy test battery. There is no evidence at this stage to suggest that it should replace the
250 BKB sentences, but the exploration did identify cases (EAL, older adults, BSL users and

251 individuals with high levels of cognitive processing) where it could be beneficial as an
252 additional tool for assessment. The absolute values that are appropriate for candidacy and
253 the actuarial equation will re-evaluated in the future analysis.

254

255 **References**

256 Blamey P, Artieres F, Başkent D, Bergeron F, Beynon A, Burke E, Dillier N, Dowell R,
257 Fraysse B, Gallégo S, Govaerts PJ, Green K, Huber AM, Kleine-Punte A, Maat B,
258 Marx M, Mawman D, Mosnier I, O'Connor AF, O'Leary S, Rousset A, Schauwers, K,
259 Skarzynski H, Skarzynski PH, Sterkers O, Terranti A, Truy E, Van de Heyning P,
260 Venail F, Vincent C, Lazard DS (2013). Factors affecting auditory performance of
261 postlinguistically deaf adults using cochlear implants: an update with 2251 patients.
262 *Audiology and Neurootology*, 18 (1), 36-47

263

264 Craddock, Cooper, Riley, Wright (2016) Cochlear Implants for Pre-Lingually Profoundly
265 Deaf Adults. **This supplement**

266

267 Doran M & Jenkinson L (2016) Mono-syllabic word test as a pre-operative assessment
268 criterion for cochlear implant candidature in adults with acquired hearing loss. **This**
269 **supplement**

270

271 National Institute for Health and Care Excellence Technology Appraisal Guidance (2009).
272 Cochlear implants for children and adults with severe to profound deafness. *NICE*
273 *technology appraisal guidance [TAG166]*. <http://www.nice.org.uk/ta166> accessed
274 29/01/2016

275

- 276 Thornton and Raffin (1978) Speech discrimination scores modelled as a binomial variable.
277 Journal of Speech and Hearing Research. 21, 507-518
278
- 279 UK CI study group (2004) Criteria of Candidacy for Unilateral Cochlear Implantation in
280 Postlingually Deafened Adults I: Theory and Measures of Effectiveness. *Ear &*
281 *Hearing* 25, 310–335
282
- 283 Vickers D, Eyles J, Brinton J, Glasberg B, Graham J. (2009) Conversion of scores
284 between Bamford, Kowal and Bench (BKB) sentences and Arthur Boothroyd (AB) words
285 in quiet for cochlear implant patients. *Cochlear Implants International*. 10(3), 142-149
286
- 287 DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L. (1988). Comparing the areas under two or more
288 correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44,
289 837–845.
- 290 Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J-C, Müller, M. (2011). pROC:
291 an open-source package for R and S+ to analyze and compare ROC curves. *BMC*
292 *Bioinformatics* 12, 77.
293

294 Table 1 Number of individuals who would or would not fall within the implant candidacy range based
 295 on BKB and AB word cutoffs. The two shaded boxes indicate the decisions that are the same if
 296 either BKB or AB criteria are used. If only AB words were used an additional one individual would fall
 297 into criteria but eight individuals would fall outside criteria. If both tests were used together one
 298 extra individual would fall within the candidacy range.

	BKB < 50%	BKB ≥ 50%	
AB ≤ 15% (scored by word)	77	1	78 Would receive CI with only AB criteria
AB > 15% (scored by word)	8	11	19 Would NOT receive CI with only AB criteria
	85 Would receive CI with only BKB criteria	12 Would NOT receive CI with only BKB criteria	

299

300

301 Table 2 Demographic details for nine individuals who had different pass/fail criteria for BKBs and AB
 302 words. The shaded row is the one participant who passed BKB sentences but failed the AB words.
 303 They hadn't gone forward for a CI and received a bone conduction device. The other failed the BKB
 304 sentences but passed the AB words. The individual with Polish as primary language didn't receive an
 305 implant due to audiometric thresholds being outside candidacy range.

Primary Language	Aetiology	Age at testing	Duration of Profound Deafness (years)	Onset	% BKBs in quiet	% ABs (phoneme score)	% ABs (word score)
English	Atresia	62	60	Pre-lingual	56	36	0
English	Otosclerosis	60	Unknown	Post-lingual	48	27	16
English	Unknown	70	8	Post-lingual	0	23	17
English	Unknown	78	28	Post-lingual	0	23	17
English	Anaesthesia	78	15	Post-lingual	0	28	17
English	Unknown	69	1	Post-lingual	40	39	17
English	Unknown	76	Unknown	Post-lingual	36	33	20
BSL	Waardenburg type II Syndrome	39	36	Pre-lingual	47	29	20
Polish	Unknown	88	10	Post-lingual	8	45	33

306

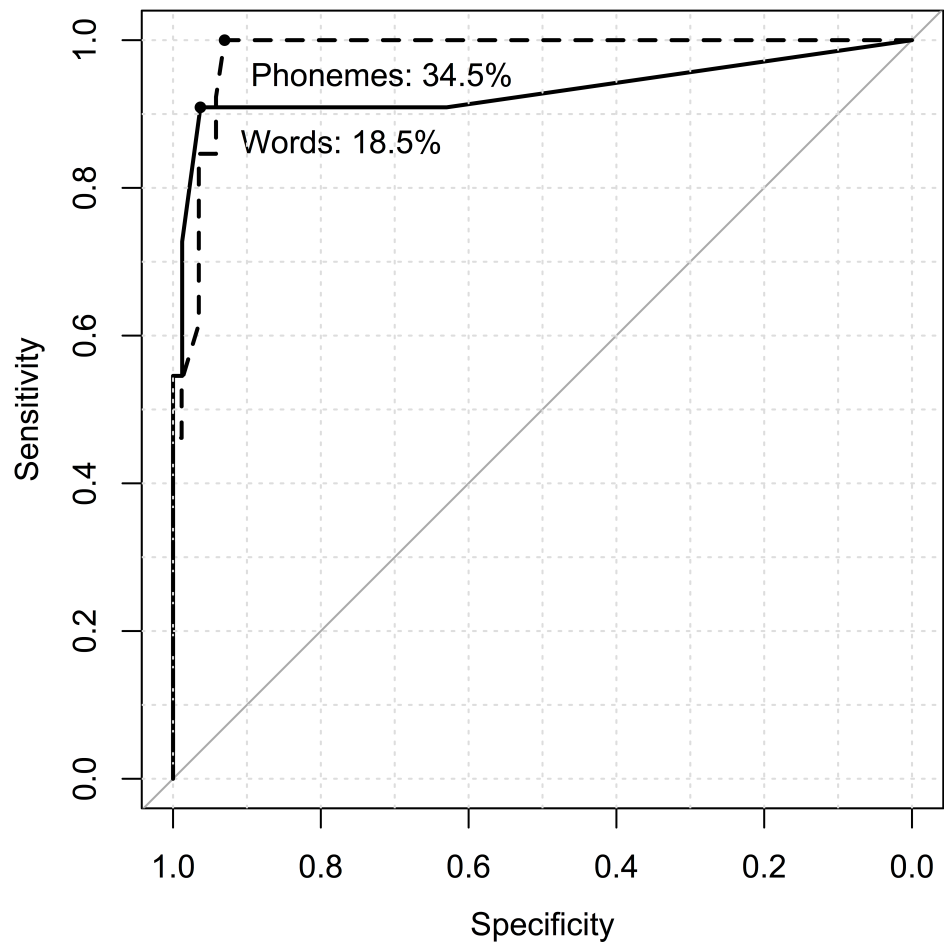
307 Figure legend

308

309 Figure 1 Receiver Operating Characteristic (ROC) curves for the word (solid curve) and
310 phoneme (dashed curve) scores from the AB word test. The symbols mark the score that
311 maximises the sensitivity and specificity of each scoring method; i.e. the cutoff score that
312 identifies eligible and non-eligible candidates as close as possible to the BKB sentence test
313 using the 50% cutoff criterion.

314

315



316
317

Figure 1