1 **Cost-effectiveness of implementing a digital psychosocial intervention for**
2 **patients with psychotic spectrum disorders in low- and middle-income**
3 **countries in Southeast Europe: economic evaluation alongside a cluster**
4 **randomised trial**

5

6 Feng Y[1*], Roukas C[1], Russo M[2], Repišti S[3], Džubur Kulenović A[4], Injac Stević L[3], Konjufca J[5],
7 Markovska-Simoska S[6], Novotni Lj[7], Ristić I[8], Smajić-Mešević E[4], Uka F[5], Zebić M[8], Vončina L[9], Bobinac
8 A[10], Jovanović N[2]

9 1: Wolfson Institute of Population Health, Queen Mary University of London, UK
10 2: Unit for Social and Community Psychiatry, WHO Collaborating Centre for Mental Health Services
11 Development, Queen Mary University of London, UK
12 3: Psychiatric Clinic, Clinical Centre of Montenegro, Podgorica, Montenegro
13 4: Clinical Center University of Sarajevo, Sarajevo, Bosnia and Herzegovina
14 5: Department of Psychology, University of Pristina, Kosovo by United Nations resolution
15 6: Macedonian Academy of Sciences and Arts, Skopje, Republic of North Macedonia
16 7: University Psychiatry Clinic, Skopje, Republic of North Macedonia
17 8: Faculty of Medicine, University of Belgrade, Belgrade, Serbia
18 9: Faculty of health studies, University of Rijeka, Croatia
19 10: Center for Health economics and Pharmacoeconomics, Faculty of Economics and
20 Business University of Rijeka, Croatia
21

22 ***Corresponding author**

23 Dr Yan Feng

24 Wolfson Institute of Population Health
25 Barts and The London School of Medicine and Dentistry
26 Queen Mary University of London
27 Yvonne Carter Building
28 58 Turner Street
29 Whitechapel
30 London UK E1 2AB

31 Email address: yan.feng@qmul.ac.uk
32

33 **A shorten version of the title**: Economic evaluation alongside a cluster randomised trial for a
34 psychosocial intervention to improve treatment of patients with psychosis

35

This peer-reviewed article has been accepted for publication but not yet copyedited or typeset, and so may be subject to change during the production process. The article is considered published and may be cited using its DOI.

36 # Cost-effectiveness of implementing a digital psychosocial intervention for
37 patients with psychotic spectrum disorders in low- and middle-income
38 countries in Southeast Europe: economic evaluation alongside a cluster
39 randomised trial

40

41 **Abstract**

42 **Background:** DIALOG+ is a digital psychosocial intervention aimed at making routine meetings
43 between patients and clinicians therapeutically effective. This study aims to evaluate the cost-
44 effectiveness of implementing DIALOG+ treatment for patients with psychotic disorders in five low-
45 and middle-income countries in Southeast Europe alongside a cluster randomised trial.

46 **Methods:** Resource use and quality of life data were collected alongside the multi-country cluster
47 randomised trial of 468 participants with psychotic disorders. Due to COVID-19 interruptions of the
48 trial's original 12-month intervention period, adjusted costs and quality-adjusted life years (QALYs)
49 were estimated at the participant level using a mixed-effects model over the first 6 months only. We
50 estimated the incremental cost-effectiveness ratio (ICER) with uncertainty presented using a cost-
51 effectiveness plane and a cost-effectiveness acceptability curve. Seven sensitivity analyses were
52 conducted to check the robustness of the findings.

53 **Results:** The average cost of delivering DIALOG+ was €91.11 per participant. DIALOG+ was associated
54 with an incremental health gain of 0.0032 QALYs (95% CI -0.0015, 0.0079), incremental costs of €84.17
55 (95% CI -8.18, 176.52), and an estimated ICER of €26,347.61. The probability of DIALOG+ being cost-
56 effective against three times the weighted gross domestic product (GDP) per capita for the five
57 participating countries was 18.9%.

58 **Conclusion:** Evidence from the cost-effectiveness analyses in this study suggested that DIALOG+
59 involved relatively low costs. However, it is not likely to be cost-effective in the five participating
60 countries compared with standard care against a willingness-to-pay threshold of three times the
61 weighted GDP per capita per QALY gained.

62

63 **Key words:** Cost-effectiveness, Cluster Randomised Trial, Psychotic Disorders, DIALOG+, Low- and
64 Middle-Income Countries in Southeast Europe

65

66 **Trial registration:** ISRCTN11913964

67 **1. Introduction**

68 The international prevalence of psychotic disorders is around 0.75% [1], and life expectancy of people
69 with psychosis is 10 to 15 years shorter than the general population [2]. These illnesses are usually
70 associated with poor quality of life and multi-morbidity [3]. They also often lead to high societal costs,
71 including direct costs for patients' healthcare and costs related to productivity losses [4]. In low- and
72 middle-income countries (LMICs) in Southeast Europe, an estimated 45% patients with psychotic
73 disorders have experienced a treatment gap (i.e., difference between the treatment they require and
74 the treatment they receive) [5-7]. This is the result of shortages in funding and qualified staff, and a
75 high patient load. Reducing the treatment gap in those countries through implementation of effective
76 and low-cost interventions is an urgent need.

77 DIALOG+ is an app-based psychosocial therapeutic intervention. Previously, interactions between
78 psychotic patients and clinicians in routine face-to-face clinical meetings were guided solely by clinical
79 judgement rather than evidence-based methods [8-9]. DIALOG+ was originally developed to make
80 meetings therapeutically effective [8]. To do this, the intervention implements a structured self-
81 assessment for patients during the meetings as well as provides guidance for clinicians on how to
82 respond to patients' ratings. Previous studies have shown that using DIALOG+ is effective in improving
83 the quality of life for patients with psychosis in UK community-based settings [8, 10]. Furthermore,
84 the effectiveness of DIALOG+ has been extensively studied in mental health care across multiple
85 countries and in different healthcare settings [11-14].

86 Since DIALOG+ is used in existing routine patient–clinician meetings, it does not require the formation
87 of new services or hiring of new staff, and only requires that the existing service makes a one-off
88 investment in computer tablets. The intervention can then be widely used by the clinicians with
89 minimal training, making it a good fit for healthcare systems with scare resources [15]. Evidence from
90 high-income settings suggests that DIALOG+ is a cost-saving intervention for people with mental
91 disorders [10]. The intervention also has potential to deliver benefits for psychotic patients in low-
92 resource settings. However, no study has previously evaluated its implementation in LMICs in
93 Southeast Europe. A multi-country cluster randomised trial within the IMPULSE study was conducted
94 to fill this empirical gap. The trial aimed to evaluate the effectiveness and cost-effectiveness of
95 implementing DIALOG+ in five LMICs in Southeast Europe compared to standard care for patients with
96 psychotic disorders [15].

97 The primary aim of this paper is to report the cost-effectiveness analyses of the DIALOG+ intervention
98 versus standard care carried out in five Southeast European countries alongside the cluster
99 randomised trial within the IMPULSE study.

100 **2. Methods**

101 2.1 Trial Design

102 The cluster randomised trial within the IMPULSE study recruited participants from five Southeast
103 European countries: Bosnia and Herzegovina, Kosovo (UN Resolution), Montenegro, North
104 Macedonia, and Serbia. These countries shared similar socioeconomic and political backgrounds
105 before the 1990s, which facilitated the trial setup and mutual learning across sites [15]. Eligible
106 participants were identified through review of medical records. Participants were eligible if they had:
107 a primary diagnosis of psychosis or related disorder in remission with ICD-10 code F20-29 or F31; a
108 lifetime history of being admitted to hospital at least once; a record of attending outpatient psychiatric
109 services; and the capacity to provide written informed consent. Participants with diagnoses of organic
110 brain disorders and/or severe cognitive deficits were excluded from the trial. Clinicians were

3

111 randomised to either the intervention group (DIALOG+) or control group (standard care). Details about
112 the trial methodology and implementation of the intervention can be found in the trial protocol [15].
113 The trial was launched in March 2019 and completed in July 2020.

114 2.2 DIALOG+ intervention and standard care

115 *DIALOG+ intervention*

116 DIALOG+ is a full therapeutic intervention which aims to make existing routine patient–clinician
117 meetings therapeutically effective. The intervention is based on quality of life research, and embeds
118 the concepts of a patient-centred approach and solution-focused therapy in order to provide an
119 evidence-based structure to routine clinical meetings between patients and clinicians. The
120 intervention consists of two parts: (1) a patient self-rating exercise of satisfaction with their life and
121 treatment, followed by (2) a four-step solution-focused discussion which aims to address the patients'
122 concerns and agree on further actions.

123 The trial was designed so that participants in the intervention group would receive 6 sessions of
124 treatment during their routine outpatient consultations over a 12-month period. In accordance with
125 the DIALOG+ manual [16], each session lasted between 30 and 60 minutes. In the first three months,
126 participants received one session per month, followed by one session every three months.

127 Every intervention session started with the patient self-rating their satisfaction with eight life domains
128 (mental health, physical health, job satisfaction, accommodation, leisure activities, partner/family,
129 friendships, personal safety) and three treatment domains (medication, practical help, meetings with
130 clinician) using the DIALOG+ app installed in computer tablets. Next, clinicians were instructed to
131 provide positive feedback to patients for any domain that was scored highly by patients and (from
132 session two onwards) for domains with an improvement in rating from previous sessions. After the
133 self-rating exercise, clinicians and patients identified a maximum of three domains for discussions.
134 These discussions were guided by a four-step approach based on the principles of solution-focused
135 therapy. Finally, the patients and clinicians jointly agreed on actions to improve the patients'
136 satisfaction with the discussed domain(s). At the beginning of the next session, they reviewed those
137 actions together [17]. Each clinician in the intervention group received face-to-face training by a local
138 research team member before the first DIALOG+ session, followed by top-up training after delivering
139 the third session. Clinicians were also able to access individual supervision provided by the study
140 researchers after each session. A computer tablet with DIALOG+ installed was offered to each clinician
141 prior to the first session.

142 *Standard care*

143 Standard care included consultations on medication, psychological support, and discussion with
144 patients on other aspects of care. Participants receiving standard care were offered 6 sessions of
145 treatment over the 12-month trial period following the same delivery schedule as participants in the
146 intervention group.

147 2.3 Impact of the COVID-19 pandemic

148 Although the trial intervention was originally designed to last 12 months, interruption due to the
149 COVID-19 pandemic from March 2020 onward led to significant changes in the intervention, patient
150 assessments, data collection, and retention in the last stage of the trial [14]. Only Serbia completed
151 the 6 sessions and the last assessment (at month 12) as per protocol before the introduction of local
152 restrictions. The other four countries adapted the DIALOG+ manual, and delivered the last 2 sessions
153 (fifth and sixth) and the last assessment remotely. Because of these changes, the effect of the

4

154 complete intervention at 12 months (i.e., 6 sessions) could not be explored. Therefore, the economic
155 evaluation was based on the first 6 months of trial data (first 4 sessions), starting from implementation
156 of the intervention at baseline.

157 2.4 Study measures

158 *Outcome measures*

159 Three instruments were used to assess quality of life for participants, including the 5L version of the
160 EQ-5D (EQ-5D-5L) [18], Manchester Short Assessment of Quality of Life (MANSA) [19], and the 10-item
161 version of Recovering Quality of Life (ReQoL-10) [20]. Due to COVID-19 pandemic (see section 2.3),
162 only data collected at baseline and 6 months after randomisation were used in analysis.

163 The EQ-5D-5L measured the primary economic evaluation outcome. EQ-5D-5L data were converted
164 to index scores by applying the EQ-5D-5L value set. There was no country-specific value set available
165 for any of the five participating countries, so we applied the newly published EQ-5D-5L value set for
166 Poland [21] in Central Europe as the best proxy available. Quality-adjusted life years (QALYs) for
167 participants during the first 6-month period of the trial were calculated using the area-under-the-
168 curve method and EQ-5D-5L index scores [22]. MANSA measured the primary clinical effectiveness
169 outcome in the IMPULSE trial. MANSA scores were calculated as the mean of the instrument's 12
170 individual item scores. ReQoL-10 is a new instrument for measuring quality of life in people
171 with mental health conditions. For ReQoL-10 data, simple sum scores on the instrument's 10
172 questions were calculated.

173 For all three outcome measures, lower score indicates poorer quality of life. EQ-5D-5L index scores
174 have a theoretical range between -0.590 and 1. The range is 1 to 7 for MANSA scores, and 0 to 40 for
175 ReQoL-10 scores.

176 *Costs data*

177 The retrospective costs data 6 months prior to baseline and 6 months after randomisation were
178 collected using an adapted version of the Client Service Receipt Inventory (CSRI) [23]. The CSRI
179 recorded participants' use of inpatient hospital services, community care service, primary care service,
180 and medication. We collected unit cost for each item from the local teams in the five participating
181 countries. Data on participants' socio-demographics, employment status, monthly income, number of
182 days off from work due to mental and/or physical health issues, monetary amount of state benefits
183 claimed, and criminal records were also collected using the CSRI.

184 We developed a health economics inventory form to collect costs data for providing DIALOG+ and
185 standard care treatments. Items included time spent by clinicians on the DIALOG+ training, time spent
186 by clinicians and supporting staff on treatments, quantity of equipment and key materials used for
187 providing treatments. We also collected the unit cost for each item using the inventory form.

188 We converted all unit costs from local currencies to euros at year 2019 level with Purchasing Power
189 Parity ((EU28=1 as the reference base) adjusted [24]. Costs for each item were then calculated as a
190 product of the quantity used and its corresponding unit cost. Finally, we summed all costs together
191 and presented the costs data at participant and assessment time-point levels. There was no discount
192 applied to adjust costs and outcomes data as the time horizon of the study was 6 months [25].

193 Outcome and cost measures used in the economic evaluation are validated scales, including EQ-5D-
194 5L [18], MANSA [19], ReQoL-10 [20] and CSRI [23]. They were translated into the local languages by
195 study researchers from central and local research teams before being administered to participants.

196   2.5 Economic evaluation

197   We compared participant-level costs and outcomes data between the two trial groups at each
198   assessment time point (i.e., baseline and 6 months after randomisation). Independent t-tests were
199   used for all comparisons. The 95% confidence intervals (CI) were constructed using a bootstrap
200   method with 1,000 replications. We also applied a three-level mixed-effects model to recognise the
201   clustered nature of our data where participants nested within clinicians that nested within countries.
202   The model controlled for baseline variable (i.e., costs or outcomes) and covariates (i.e., age of
203   participants, ICD-10 code, and profession of clinicians).

204   We conducted the within-trial analyses from a healthcare perspective under the principle of intention-
205   to-treat. Time horizon for the economic evaluation was 6 months, starting from implementation of
206   the intervention at baseline. This was consistent with the time horizon for the effectiveness evaluation
207   of DIALOG+ in the IMPULSE trial [14].

208   Cost-utility analysis was used to conduct the base case economic evaluation. Costs included
209   intervention costs, health service costs, and medication costs. The primary economic outcome
210   measure used QALYs calculated from the EQ-5D-5L index scores. We estimated the incremental costs
211   (and incremental QALYs) as the difference between the intervention and control groups over the first
212   6 months of the trial period, controlling for baseline values, participants' ages, ICD-10 code, and
213   profession of clinicians. A three-level mixed-effects model was applied. The pattern of missing values
214   with three variables (i.e., costs at baseline, costs, and QALYs over the 6-month period) was assumed
215   as missing at random. Multiple imputation with chained equations was applied to generate 70
216   imputed data sets (the largest fraction of missing information was 0.5258). The point estimate of the
217   incremental cost-effectiveness ratio (ICER) was calculated by dividing the estimated incremental costs
218   by the estimated incremental QALYs. To explore the uncertainty around the point estimate, we used
219   the non-parametric bootstrap approach with 1,000 replications to estimate the 95% CI around the
220   ICER [26]. The result was presented using a cost-effectiveness plane. We also constructed a cost-
221   effectiveness acceptability curve to show the probability that DIALOG+ was cost-effective compared
222   with standard care for a range of willingness-to-pay values for an additional QALY gained.

223   There is no evidence-based cost-effectiveness threshold to apply in multi-country trials for LMICs [27].
224   The World Health Organisation has recommended using one to three times the gross domestic
225   product (GDP) per capita of an LMIC as the cost-effectiveness threshold for the country [28, 29]. An
226   intervention with an estimated ICER less than three times the national annual GDP per capita is
227   considered cost-effective. In our base case evaluation, we compared our point estimate of the ICER
228   against one to three times the weighted GPD per capita. The weights are proportions of participants
229   from each country out of the total trial sample size.

230   To check the robustness of the findings from the base case evaluation, we conducted seven sensitivity
231   analyses. First, we ran the base case analysis with complete cases only (i.e., without missing values).
232   Second, the seemingly unrelated regression model without robust standard error was applied to
233   compare the impact of the model choice [30]. Third, we estimated two ICERs using the minimum (and
234   maximum) unit costs, respectively, for all medications from each country when unit costs for some
235   medications were reported in a range. Fourth, we undertook analyses using a broader analytical
236   perspective, including costs due to productivity lost as a result of mental or physical health problems.
237   In the fifth and sixth sensitivity analyses, we replaced the outcome measure EQ-5D-5L index scores
238   with MANSA scores and ReQoL-10 sum scores, respectively. Finally, we estimated country-specific
239   ICERs by applying the method developed by Willke and colleagues [31].

240 Statistical significance was determined at the 5% level (P<0.05). All analyses were
241 performed with software package STATA/MP 17 [32].

3. Results

243 3.1 Characteristics of the sample

244 We present the characteristics of all participants at baseline in **Table 1**. In total, 468 eligible
245 participants were recruited, with 236 receiving the DIALOG+ treatment and 232 receiving standard
246 care. There were 424 participants at 6 months after randomisation. The trial recruited 81 clinicians
247 from 11 clinics across five countries. The average age of participants in the trial was 42.59 years old
248 (SD=11.30). More than half of the participants were male (54.3%), single (54.3%), unemployed
249 (59.7%), not receiving any state benefits (56.8%), and reported the highest level of education as high
250 school (60.5%). Montenegro contributed the largest trial sample (n=122, 26.1%), followed by Kosovo
251 (UN Resolution; n=103, 22%), North Macedonia (n=82, 17.5%), Bosnia and Herzegovina (n=81, 17.3%),
252 and Serbia (n=80, 17.1%).

253 3.2 Costs for DIALOG+ and standard care interventions

254 The average cost of delivering DIALOG+ for each participant was €91.11 during the 6-month trial
255 period. The majority of this cost was for clinicians' time, with €50.92 spent on delivering DIALOG+ and
256 €14.69 on training. The cost also included key resource use (€17.66; computer tablets, fee for
257 translating DIALOG+ manual to local language, room booking for DIALOG+ training), and other
258 equipment use (€6.59; cell phones, recording devices, stationery). Costs from other staff that
259 supported the delivery of DIALOG+ were minor at €1.24 per participant. Average total cost for
260 delivering standard care sessions during the 6-month trial period was €20.87 per participant.

261 3.3 Resource use and costs

262 **Table 2** presents the quantity of resource use at the participant level over the 6-month trial period,
263 while **Appendix 1** reports the unit costs for each resource use item. **Table 3** shows the average cost
264 per participant for resource use over the 6-month trial period. The single most costly resource was
265 medication. On average, the medication cost for participants in the intervention group was €237.23
266 per participant, while average medication cost in the control was €243.35. The total cost in the
267 intervention group was €565.95 per participant, and €497.78 per participant in the control. The
268 difference in total cost between the groups was €68.17 (95% CI -54.26, 168.60), but this was not
269 statistically significant as suggested by independent t test. While controlling for the differences in total
270 costs and the list of other covariates at baseline, the mixed-effects models produced qualitatively
271 similar results. The difference in total cost was estimated as €98.42 (95% CI -29.49, 208.30), although
272 this was not statistically significant.

273 We found differences between two groups in costs for total resource use over 6 months before
274 randomisation (**Appendix 2**), and these differences were not statistically significant.

275 3.4 Outcome measures

276 **Table 4** shows the participant level EQ-5D-5L index scores (and estimated QALYs), MANSA scores, and
277 ReQoL-10 sum scores at each assessment time point (baseline and 6 months) by trial group
278 (intervention and control). After adjusting for the baseline differences in EQ-5D-5L index scores and
279 the list of covariates, the mixed-effect model resulted in a difference of 0.0035 QALYs (95% CI -0.0021,
280 0.0089) between the intervention and control groups over the 6-month period, a difference of 0.1810
281 points (95% CI 0.0315 to 0.3158) for the MANSA, and a difference of 0.7237 points (95% CI -0.2798 to

7

282   1.9375) for the ReQoL-10. All three outcome measures suggested a health improvement after 6
283   months of treatment with DIALOG+, however, only the difference in MANSA scores was statistically
284   significant.

285   3.5 Cost-effectiveness base case analysis

286   **Table 5** reports results from the base case evaluation. Cost per QALY gained from implementing
287   DIALOG+ was €26,347.61, achieved by dividing incremental costs of €84.17 (95% CI -8.18, 176.52) by
288   incremental QALYs of 0.0032 (95% CI -0.0015, 0.0079). The weighted GDP per capita was €4,587, and
289   three times this value was €13,761. **Figure 1** shows the uncertainty around our point estimate of the
290   ICER using a cost-effectiveness plane, including 1,000 pairs of incremental costs and incremental
291   QALYs from bootstrap replications. **Figure 2** presents the cost-effectiveness acceptability curve
292   showing that the probability of DIALOG+ being cost-effective compared with standard care was 3.8%
293   at a willingness-to-pay of €4,587 per QALY, and 18.9% at a willingness-to-pay of €13,761 per QALY.
294   The base case analysis suggested that DIALOG+ was unlikely to be cost-effective.

295   3.6 Sensitivity analyses

296   **Table 5** reports results from seven sensitivity analyses. The first four sensitivity analyses produced
297   results consistent with the base case analysis: the point estimate of the ICER was above three times
298   the weighted GDP per capita per QALY gained threshold. When ReQoL-10 sum scores were applied as
299   the outcome measure, one score of improvement in ReQoL-10 was associated with additional costs
300   of €119.02 (sensitivity analysis five). Analysis of MANSA scores suggested that an improvement of one
301   score in MANSA was associated with additional costs of €523.53 (sensitivity analysis six). In sensitivity
302   analysis seven, we attempted to estimate country-specific ICERs. DIALOG+ treatment was consistently
303   found not to be cost-effective in four participating countries; Kosovo (UN Resolution) was the only
304   country where the intervention was more effective and less costly than standard care.

305   **4. Discussion**

306   The main cost-effectiveness analysis suggested that DIALOG+ is slightly more costly and slightly more
307   effective than standard care over the first 6 months of the trial period. The point estimate of the ICER
308   was higher than the willingness-to-pay value at three times the weighted GDP per capita of the five
309   participating countries. Regarding the uncertainty of this point estimate, our results suggested that
310   the probability was low (18.9%) that DIALOG+ was cost-effective compared with standard care at the
311   provider's willingness-to-pay threshold. We conducted sensitivity analyses to explore the impact of
312   missing values, estimation methods, key parameters for costs, and evaluation perspectives. None of
313   these analyses challenged the main finding. In country-specific analyses, we found DIALOG+ was more
314   effective and more costly in four of the five participating countries (and the point estimate of the ICER
315   was not cost-effective). Kosovo (UN Resolution) alone showed DIALOG+ as more effective and less
316   costly than standard care. This result should be interpreted with caution as the trial was not powered
317   to detect country-specific treatment effects (in particular, for the EQ-5D-5L measure). Cost analyses
318   shared similar limitations. Additionally, a few unit costs for resource use in Kosovo (UN Resolution)
319   were proxied by the lowest unit price among the other four participating countries due to absence of
320   an official local data source. Country-specific costs for total resource use per participant and outcomes
321   by group is reported in **Appendices 3 and 4**, respectively.

322   In this trial, we observed modest improvements of quality of life measured by three instruments. Only
323   the difference in MANSA scores (i.e., the primary clinical effectiveness outcome in the IMPULSE trial)
324   between the intervention and control groups was statistically significant [14]. The primary economic
325   evaluation relied on QALYs derived from the EQ-5D-5L data as the outcome measure. It should be

8

noted that the EQ-5D-5L has been criticised for its sensitivity regarding people with psychotic disorders and severe and complex nonpsychotic disorders [33]. It has been argued that a condition-specific instrument might be more sensitive in reflecting changes in quality of life in these populations than a generic instrument like the EQ-5D-5L.

DIALOG+ has previously been applied in community care settings in the UK for patients with psychosis [10]. However, the UK study found that the treatment was less costly than standard care, which was not in line with the results from our IMPULSE study. The UK study did not collect EQ-5D-5L data, which was one of the limitations reported by its authors. We, therefore, were unable to make a direct comparison between IMPULSE and the UK study of patients' self-reported EQ-5D-5L and QALYs.

Evidence of cost-effectiveness analyses of treatments for severe mental illness in Southeast Europe is scarce [15]. Treatments are predominantly provided in large psychiatric hospitals with limited community-based alternatives. However, a recently published economic evaluation in the Czech Republic showed that it is cost-effective to discharge patients with chronic psychotic disorders to community care compared with care in psychiatric hospitals [4]. This finding supports one of the aims of introducing DIALOG+ in the LMIC settings, namely, to provide effective and cost-effective mental health treatment for psychotic patients through community-based services.

To our knowledge, this study reports the first cost-effectiveness evaluation of implementing (non-pharmacological) psychosocial treatments for people with psychosis in Southeast Europe. A strength of this study is the trial data that we collected. The challenges around data collection and lack of country-specific unit cost data in multi-country randomised controlled trials are well documented in the literature [29]. It has widely been observed in economic evaluations of multi-country clinical trials that the analyses applied unit costs from one country to all participating countries due to lack of unit cost data from all individual countries [29, 34]. A concern with this approach is around the possibility of generating biased (over/under) estimates for costs. In IMPULSE trial, we collected resource use and outcomes data at the patient level, as well as country-specific unit costs for each resource item used. This strategy for data collection enabled patient-level data analyses with multi-country costing.

This study has several limitations that should be considered. First, there were no country-specific value sets for the three outcome measures (EQ-5D-5L, MANSA, ReQoL-10). Since we observed minimal improvements in QALYs for EQ-5D-5L data, the impact of value set choice on the estimated ICERs could, therefore, be very limited. We reported the results of cost-effectiveness analyses in this paper using ReQoL-10 and MANSA to enable comparisons with future research. Another consideration is around the generalisability of our findings. This issue is well documented for economic evaluations of multi-country randomised controlled trials [29, 35]. We showed different results in cost analyses from the application of the DIALOG+ in the UK [10]. Care should be taken when interpreting our findings to inform decision making in a different context or/and for a different population. A final limitation of the study relates to the COVID-19 pandemic. The trial was designed to last 12 months, but only the first 6 months of data was interpretable due to disruptions in the study's delivery relating to pandemic restrictions [14].

Future research might consider producing value sets or conducting mapping exercises to convert scores from MANSA and ReQoL instruments to health utilities in LMIC settings. Furthermore, we found limited research evidence on country-specific cost-effectiveness thresholds in LMICs [36]. The empirical evidence and methodological research in this area are much needed. Finally, we did not find an agreed approach for estimating country-specific cost-effectiveness of an intervention in multi-country clinical trials. Additional research is required in this area in order to inform policy makers regarding resource allocation decisions at the country-specific level.

371 **5. Conclusion**

372 This paper reports an economic evaluation of the DIALOG+ intervention alongside the IMPULSE trial.
373 Within the trial, DIALOG+ was shown to be more costly and also more effective for patients with
374 psychosis compared with standard care. The probability was low that DIALOG+ was a cost-effective
375 treatment at the willingness-to-pay threshold of three times the weighted GDP per capita of the five
376 participating countries.

377 **Abbreviations**

378 CI: confidence interval

379 CSRI: Client Service Receipt Inventory

380 EQ-5D-5L: The 5-level EQ-5D version

381 GDP: gross domestic product

382 ICD-10: International Classification of Diseases, Tenth Revision

383 ICER: incremental cost-effectiveness ratio

384 LMICs: low- and middle-income countries

385 IMPULSE: Implementation of an effective and cost-effective intervention for patients with psychotic
386 disorders in low and middle-income countries in Southeast Europe

387 MANSA: Manchester Short Assessment of Quality of Life

388 QALY: Quality-Adjusted Life Year

389 ReQoL-10: Recovering Quality of Life, a short 10-item version

390

391 **References**

392 1. Moreno-Küstner B, Martín C, Pastor L. Prevalence of psychotic disorders and its association
393   with methodological issues. A systematic review and meta-analyses. PLoS One. 2018 Apr
394   12;13(4):e0195687. doi: 10.1371/journal.pone.0195687.
395 2. Simon GE, Stewart C, Yarborough BJ, Lynch F, Coleman KJ, Beck A, Operskalski BH, Penfold RB,
396   Hunkeler EM. Mortality Rates After the First Diagnosis of Psychotic Disorder in Adolescents
397   and Young Adults. JAMA Psychiatry. 2018 Mar 1;75(3):254-260. doi:
398   10.1001/jamapsychiatry.2017.4437.
399 3. Rodrigues M, Wiener JC, Stranges S, Ryan BL, Anderson KK. The risk of physical multimorbidity
400   in people with psychotic disorders: A systematic review and meta-analysis. J Psychosom Res.
401   2021 Jan;140:110315. doi: 10.1016/j.jpsychores.2020.110315.
402 4. Winkler P, Koeser L, Kondrátová L, Broulíková HM, Páv M, Kališová L, Barrett B, McCrone P.
403   Cost-effectiveness of care for people with psychosis in the community and psychiatric
404   hospitals in the Czech Republic: an economic analysis. Lancet Psychiatry. 2018
405   Dec;5(12):1023-1031. doi: 10.1016/S2215-0366(18)30388-2.
406 5. mhGAP: Mental Health Gap Action Programme: Scaling Up Care for Mental, Neurological and
407   Substance Use Disorders. Geneva: World Health Organization; 2008.
408 6. World Health Organization. Mental Health Atlas 2014,
409   https://apps.who.int/iris/handle/10665/178879; [accessed 12 January. 2022].
410 7. World Health Organization. Regional Office for Europe, European Observatory on Health
411   Systems and Policies, McDaid, David, Knapp, Martin & Curran, Claire. (2005). Mental health III:
412   funding mental health in Europe. World Health Organization. Regional Office for
413   Europe. https://apps.who.int/iris/handle/10665/107633.
414 8. Priebe S, Kelley L, Omer S, Golden E, Walsh S, Khanom H, Kingdon D, Rutterford C, McCrone
415   P, McCabe R. The Effectiveness of a Patient-Centred Assessment with a Solution-Focused

11

416   Approach (DIALOG+) for Patients with Psychosis: A Pragmatic Cluster-Randomised Controlled
417   Trial in Community Care. Psychother Psychosom. 2015;84(5):304-13. doi: 10.1159/000430991.
418 9. Priebe S, McCabe R. The therapeutic relationship in psychiatric settings. Acta Psychiatr Scand
419   Suppl. 2006;(429):69-72. doi: 10.1111/j.1600-0447.2005.00721.x.
420 10. Priebe S, Golden E, Kingdon D, Omer S, Walsh S, Katevas K, McCrone P, Eldridge S, McCabe R.
421   Effective patient–clinician interaction to improve treatment outcomes for patients with
422   psychosis: a mixed-methods design. Southampton (UK): NIHR Journals Library; 2017 Feb.
423 11. Matanov A, McNamee P, Akther S, Barber N, Bird V. Acceptability of a technology-supported
424   and solution-focused intervention (DIALOG+) for chronic depression: views of service users
425   and clinicians. BMC Psychiatry. 2021 May 20;21(1):263. doi: 10.1186/s12888-021-03256-5.
426 12. van Loggerenberg F, McGrath M, Akena D, Birabwa-Oketcho H, Méndez CAC, Gómez-Restrepo
427   C, Džubur Kulenoviĉ A, Muhić M, Sewankambo NK, Sikira H, Priebe S. Feasibility, experiences
428   and outcomes of using DIALOG+ in primary care to improve quality of life and mental distress
429   of patients with chronic conditions: an exploratory non-controlled trial in Bosnia and
430   Herzegovina, Colombia and Uganda. Pilot Feasibility Stud. 2021 Sep 30;7(1):180. doi:
431   10.1186/s40814-021-00914-z.
432 13. van den Brink R, Wiersma D, Wolters K, Bullenkamp J, Hansson L, Lauber C, Martinez-Leal R,
433   McCabe R, Rössler W, Salize H, Svensson B, Torres-Gonzales F, Priebe S. Non-uniform
434   effectiveness of structured patient-clinician communication in community mental healthcare:
435   an international comparison. Soc Psychiatry Psychiatr Epidemiol. 2011 Aug;46(8):685-93. doi:
436   10.1007/s00127-010-0235-x.
437 14. Jovanović N, Russo M, Pemovska T, Francis JJ, Arenliu A, Bajraktarov S et al. Improving
438   treatment   of   patients   with   psychosis   in   low-and-middle-income   countries
439   in Southeast Europe: results from a hybrid effectiveness-implementation, pragmatic, cluster-
440   randomised clinical trial (IMPULSE). Accepted for publication in Eur Psychiatry, June 2022.
441 15. Jovanovic N, Francis J, Maric NP, Arenliu A, Barjaktarov S, Kulenovic AD, Injac L, Feng Y,
442   Novotni A. Implementing a psychosocial intervention DIALOG+ for patients with psychotic
443   disorders in low and middle income countries in South Eastern Europe: protocol for a hybrid
444   effectiveness-implementation cluster randomized clinical trial (IMPULSE). Global Psychiatry.
445   2020 May;3(1): 83-96. https://doi.org/10.2478/gp-2019-0020.
446 16. Priebe S, Golden E, Katevas H, Healey P, McCabe R. DIALOG+ Manual,
447   https://www.elft.nhs.uk/sites/default/files/DIALOG%20Manual%20%282021%29.pdf;
448   [accessed 9 July 2022].
449 17. Omer S, Golden E, Priebe S. Exploring the Mechanisms of a Patient-Centred Assessment with
450   a Solution Focused Approach (DIALOG+) in the Community Treatment of Patients with
451   Psychosis: A Process Evaluation within a Cluster-Randomised Controlled Trial. PLoS One. 2016
452   Feb 9;11(2):e0148415. doi: 10.1371/journal.pone.0148415.
453 18. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, Bonsel G, Badia X. Development
454   and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011
455   Dec;20(10):1727-36. doi: 10.1007/s11136-011-9903-x.
456 19. Priebe S, Huxley P, Knight S, Evans S. Application and results of the Manchester Short
457   Assessment of Quality of Life (MANSA). Int J Soc Psychiatry. 1999 Spring;45(1):7-12. doi:
458   10.1177/002076409904500102.
459 20. Keetharuth AD, Brazier J, Connell J, Bjorner JB, Carlton J, Taylor Buck E, Ricketts T, McKendrick
460   K, Browne J, Croudace T, Barkham M. Recovering Quality of Life (ReQoL): a new generic self-
461   reported outcome measure for use with people experiencing mental health difficulties. Br J
462   Psychiatry. 2018 Jan;212(1):42-49. doi: 10.1192/bjp.2017.10.
463 21. Golicki D, Jakubczyk M, Graczyk K, Niewada M. Valuation of EQ-5D-5L Health States in Poland:
464   the First EQ-VT-Based Study in Central and Eastern Europe. Pharmacoeconomics. 2019
465   Sep;37(9):1165-1176. doi: 10.1007/s40273-019-00811-7.

12

22. Matthews JN, Altman DG, Campbell MJ, Royston P. Analysis of serial measurements in medical research. BMJ. 1990 Jan 27;300(6719):230-5. doi: 10.1136/bmj.300.6719.230.

23. Beecham J, Knapp M. *Costing psychiatric interventions.* In: Thornicroft Graham, editor. Measuring Mental Health Needs (Second Edition), London: Royal College of Psychiatrists; 2001, p. 200-224.

24. Eurostat. Purchasing power parities (PPPs), price level indices and real expenditures and real expenditures for ESA 2010 aggregates, https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=prc_ppp_ind&lang=en; [accessed 6 April 2021].

25. National Institute for Health and Clinical Excellence. Guide to the Methods of Technology Appraisal 2013, https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781; [accessed 15 February 2021].

26. Briggs AH, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. Health Econ. 1997 Jul-Aug;6(4):327-40. doi: 10.1002/(sici)1099-1050(199707)6:4<327::aid-hec282>3.0.co;2-w.

27. Woods B, Revill P, Sculpher M, Claxton K. Country-Level Cost-Effectiveness Thresholds: Initial Estimates and the Need for Further Research. Value in Health. 2016 Dec;19(8):929-935. doi: 10.1016/j.jval.2016.02.017.

28. Bertram MY, Lauer JA, De Joncheere K, Edejer T, Hutubessy R, Kieny MP, Hill SR. Cost-effectiveness thresholds: pros and cons. Bull World Health Organ. 2016 Dec 1;94(12):925-930. doi: 10.2471/BLT.15.164418.

29. Oppong R, Jowett S, Roberts TE. Economic Evaluation alongside Multinational Studies: A Systematic Review of Empirical Studies. PLoS One. 2015 Jun 29;10(6):e0131949. doi: 10.1371/journal.pone.0131949.

30. Gomes M, Ng ES, Grieve R, Nixon R, Carpenter J, Thompson SG. Developing appropriate methods for cost-effectiveness analysis of cluster randomized trials. Med Decis Making. 2012 Mar-Apr;32(2):350-61. doi: 10.1177/0272989X11418372.

31. Willke RJ, Glick HA, Polsky D, Schulman K. Estimating country-specific cost-effectiveness from multinational clinical trials. Health Econ. 1998 Sep;7(6):481-93. doi: 10.1002/(sici)1099-1050(199809)7:6<481::aid-hec353>3.0.co;2-k.

32. StataCorp, L.L.C., Stata Statistical Software. Release 16. College Station, TX. 2019, Stata Press.

33. Brazier J. Is the EQ-5D fit for purpose in mental health? Br J Psychiatry. 2010 Nov;197(5):348-9. doi: 10.1192/bjp.bp.110.082453.

34. Reed SD, Anstrom KJ, Bakhai A, Briggs AH, Califf RM, Cohen DJ, Drummond MF, Glick HA, Gnanasakthy A, Hlatky MA, O'Brien BJ, Torti FM Jr, Tsiatis AA, Willan AR, Mark DB, Schulman KA. Conducting economic evaluations alongside multinational clinical trials: toward a research consensus. Am Heart J. 2005 Mar;149(3):434-43. doi: 10.1016/j.ahj.2004.11.001.

35. Sculpher MJ, Pang FS, Manca A, Drummond MF, Golder S, Urdahl H, Davies LM, Eastwood A. Generalisability in economic evaluation studies in healthcare: a review and case studies. Health Technol Assess. 2004 Dec;8(49):iii-iv, 1-192. doi: 10.3310/hta8490.

36. Vončina L, Strbad T, Fürst J, Dimitrova M, Kamusheva M, Vila M, Mardare I, Hristova K, Harsanyi A, Atanasijević D, Banović I, Bobinac A. Pricing and Reimbursement of Patent-Protected Medicines: Challenges and Lessons from South-Eastern Europe. Appl Health Econ Health Policy. 2021 Nov;19(6):915-927. doi: 10.1007/s40258-021-00678-w.

13

510 **Acknowledgements**

514

515 **Financial support**

519

520 **Conflicts of Interest**

521 The authors declare no conflicts of interest.

522

523 **Supplementary**

524 For supplementary material accompanying this paper, visit cambridge.org/EPA.

525

526 **Data Availability Statement**

527 The data that support the findings of this study are available from the corresponding author, YF, upon
528 reasonable request.

529

530 **Trial registration:** ISRCTN11913964

531

532 **Ethical approval**

533 All procedures in the trial were approved by the following six ethics committees including Bosnia and
534 Herzegovina (Klinicki Centar Univerziteta u Sarajevu – Eticki Komitet 03-02-4216, Eticki komitet JU
535 Psihijatriska bolnica Kantona Sarajevo & JU Zavod za bolesti ovisnosti Kantona Sarajevo 02.8 – 408/19);
536 Kosovo (UN Resolution) (Hospital and University Clinical Service of Kosovo – Ethics Committee 2019-
537 85); Montenegro (Javna Zdravstvena Ustanova Klinicki Centar Crne Gore – Eticki komitet 03/01 –
538 29304/1, ZU Specijalna Bolnica za Psihijatriju 'Dobrota' Kotor – Eticki komitet, Eticki Komitet JZU Dom
539 Zdravlja 'DR Nika Labovic' Berane 01-47); Republic of North Macedonia (Eticka Komisija za istrazuvanje
540 na luge, Medicinski Fakultet pri UKIM vo Skopje 03-24219); Serbia (Eticka komisija Medicinskog
541 fakulteta u Beogradu 2650/XII-20 and Eticka komisija Specijalne bolnice 'Dr Slavoljub Bakalovic' Vrsac
542 01-36/1); and the United Kingdom (Queen Mary University of London QMREC2204a, 16 October
543 2018).

Appendix 1: Summary of main resources and unit costs (euros) by country, adjusted for Purchasing Power Parity

Resource item	Unit	Bosnia and Herzegovina		Kosovo (UN Resolution)		Montenegro		North Macedonia		Serbia	
		Unit cost[1]	Data source	Unit cost[1]	Data source	Unit cost[1]	Data source	Unit cost[1]	Data source	Unit cost[1]	Data source
Psychiatric hospital (voluntary)	Day	119.82	HIRI[2,4]	21.86	Proxy[3]	21.86	HIFM[2,5]	35.77	HIFNM[2,6]	26.69	LIS[2,7]
Psychiatric hospital (involuntary)	Day	119.82	HIRI	21.86	Proxy	21.86	HIFM	35.77	HIFNM	26.69	LIS
Physical hospital	Day	116.69	HIRI	21.86	Proxy	21.86	HIFM	44.71	HIFNM	26.69	LIS
General Practitioner	Visit	10.94	HIRI	4.24	Proxy	12.04	HIFM	4.24	HIFNM	4.90	LIS
Psychiatrist	Visit	16.67	HIRI	9.16	Proxy	27.85	HIFM	74.51	HIFNM	9.16	LIS
Psychologist	Visit	22.40	HIRI	21.98	Proxy	21.98	HIFM	33.53	HIFNM	35.40	LIS
Dentist	Visit	4.69	HIRI	4.69	Proxy	18.87	HIFM	63.33	HIFNM	5.57	LIS
Emergency service	Visit	15.63	HIRI	15.63	Proxy	17.41	HIFM	29.06	HIFNM	41.31	WHO Choice
Other mental health professional	Visit	17.30	HIRI	12.46	Proxy	12.46	HIFM	50.51	WHO Choice[2,8]	43.05	WHO Choice
Other specialist doctor	Visit	8.17	HIRI	8.17	Proxy	11.61	HIFM	48.49	WHO Choice	41.31	WHO Choice
Lost work by patients at baseline	Daily income	35.81	World Bank[9]	35.81	Proxy	57.89	World Bank	39.58	World Bank	44.22	World Bank
Medicine at baseline	Patient	407.58	CRF[2]	130.86	CRF	286.29	CRF	308.31	CRF	1393.29	CRF
DIALOG+ clinician	Hourly rate	12.28	HEI[2]	3.34	HEI	16.74	HEI	24.44	HEI	12.43	HEI

15

Standard care clinician	Hourly rate	13.54	HEI	3.71	HEI	17.37	HEI	32.06	HEI	9.96	HEI

1: Unit costs were Purchasing Power Parity (EU28=1) adjusted.

2: CRF: Case Report Form; HEI: Health Economics Inventory form; HIRI: Health Insurance and Reinsurance Institute of the Federation of Bosnia and Herzegovina; HIFM: Health Insurance Fund of Montenegro; HIFNM: Health Insurance Fund of Republic of North Macedonia; WHO: World Health Organization; LIS: Legal Information System of the Republic of Serbia.

3: In absence of official data source, unit costs for Kosovo (UN Resolution) were derived using the lowest unit price among the other four participating countries in the trial.

4: Data source: http://www.zzofbih.ba/bs/dokument/tarifnik/68. Last accessed 12 December 2020.

5: Data source: https://fzocg.me/davaoci_zdravstvenih_usluga.php?type=prices2. Last accessed 12 December 2020.

6: Data source: http://www.fzo.org.mk/default-en.asp. Last accessed 12 December 2020.

7: Data source: http://www.pravno-informacioni-sistem.rs/SlGlasnikPortal/eli/rep/sgrs/drugidrzavniorganiorganizacije/pravilnik/2019/55/2/reg. Last accessed 12 December 2020.

8: Data source: https://www.who.int/choice/cost-effectiveness/inputs/health_service/en/. Last accessed 12 December 2020.

9. Data source: https://data.worldbank.org/indicator/NY.ADJ.NNTY.PC.CD. Last accessed 17 February 2021.

Appendix 2: Mean costs (euros) for resource use over 6 months before baseline by group, adjusted for Purchasing Power Parity

	DIALOG+ intervention (N=236)[1]		Standard care (N=232)[1]		Difference (no adjustment)[2]
	N	Mean (SD)	N	Mean (SD)	Difference (P value) (95% CI)
Inpatient service					
Voluntary admission to psychiatric hospital (days)	236	676.76 (1596.50)	232	392.49 (1511.86)	284.27 (-5.86, 562.59)
Involuntary admission to psychiatric hospital (days)	236	206.60 (1057.21)	232	96.07 (611.65)	110.53 (-45.04, 271.58)
Admission to hospital for physical health (days)	236	6.06 (60.17)	232	29.81 (271.44)	-23.75 (-71.56, 3.18)
Sub total	**236**	**889.42 (2387.47)**	**232**	**518.36 (1960.70)**	**371.06 (-40.63, 762.05)**
Primary/community service[3]					
General Practitioner	234	25.52 (40.92)	231	30.77 (38.59)	-5.26 (-12.17, 1.89)
Psychiatrist	232	139.45 (231.43)	231	118.09 (138.96)	21.36 (-11.06, 60.32)
Psychologist	235	20.00 (73.99)	231	53.73 (285.27)	-33.73 (-76.16, 1.13)
Dentist	236	16.84 (60.90)	230	20.38 (92.52)	-3.54 (-21.63, 9.40)
Emergency services	211	2.67 (9.06)	198	2.68 (9.64)	-0.01 (-2.09, 1.74)
Other mental health professional	236	42.72 (133.80)	230	52.78 (180.20)	-10.05 (-40.27, 18.87)
Other specialist doctor	236	16.86 (49.36)	232	11.30 (29.99)	5.56 (-1.45, 12.78)
Sub total	**206**	**211.90 (225.61)**	**195**	**264.42 (430.25)**	**-52.52 (-119.83, 10.15)**

Patients' other costs					
Lost work by patients	232	141.19 (813.86)	230	248.61 (1310.80)	-107.42 (-332.78, 76.63)
Medicine	236	332.16 (577.65)	232	482.09 (2411.37)	-149.93 (-579.02, 65.93)
Total costs with productivity lost	**203**	**1640.46 (2789.70)**	**193**	**1633.08 (3691.84)**	**7.38 (-653.80, 661.01)**
Total costs without productivity lost	**206**	**1478.06 (2628.71)**	**195**	**1357.47 (3332.29)**	**120.59 (-511.82, 678.01)**

1: N refers to the number of participants who responded to each question.

2: Independent t-tests are reported; CI was produced using bootstrapping method with 1,000 replications; * P value is <0.05.

3: Those contacts do not include care that participants received in the IMPULSE trial.

Appendix 3: Mean costs (euros) for resource use over 6 months before and 6 months after randomisation by country and group, adjusted for Purchasing Power Parity

	6 months after randomisation		6 months before randomisation	
	DIALOG+ intervention	Standard care	DIALOG+ intervention	Standard care
Bosnia and Herzegovina	627.10	466.26	2327.47	2162.24
Kosovo (UN Resolution)	331.51	548.52	604.68	634.92
Montenegro	552.87	445.20	1761.03	653.23
North Macedonia	643.02	538.20	644.53	1462.26
Serbia	728.87	510.20	1632.55	2444.61

Appendix 4: Comparisons of EQ-5D-5L index scores, MANSA scores, and ReQoL-10 sum scores by country and group

	EQ-5D-5L index scores		MANSA scores		ReQoL-10 sum scores	
	DIALOG+ intervention	Standard care	DIALOG+ intervention	Standard care	DIALOG+ intervention	Standard care
Bosnia and Herzegovina						
At baseline	0.926	0.970	4.898	5.036	30.325	30.951
At 6 months	0.964	0.961	4.901	4.912	27.757	29.769
Kosovo (UN Resolution)						
At baseline	0.829	0.880	4.190	4.128	23.788	23.294
At 6 months	0.922	0.927	4.775	4.519	27.213	25.044
Montenegro						
At baseline	0.891	0.933	4.332	4.604	23.677	25.233
At 6 months	0.932	0.942	4.650	4.654	25.378	25.632
North Macedonia						
At baseline	0.921	0.943	4.715	4.774	27.341	29.024
At 6 months	0.950	0.948	4.979	4.732	28.600	29.171
Serbia						
At baseline	0.906	0.915	4.427	4.197	24.805	20.385
At 6 months	0.907	0.890	4.935	4.424	27.162	21.056

Figure 1: Cost-effectiveness plane (1,000 iterations)
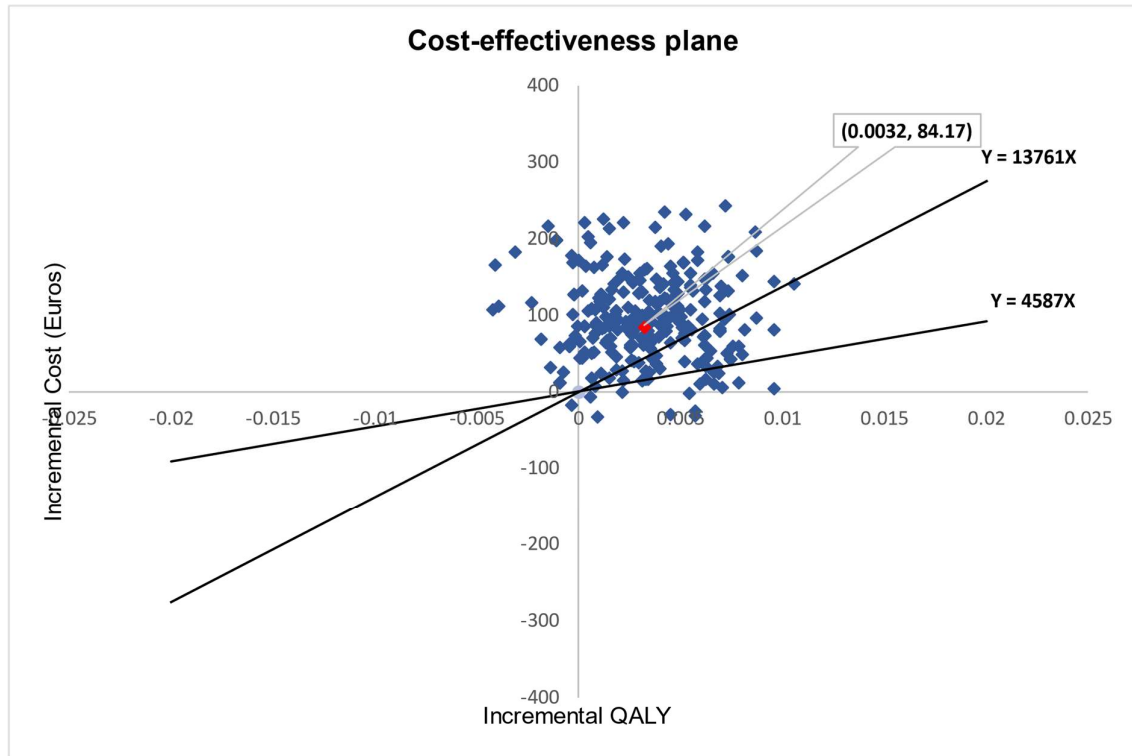
Figure 2: Cost-effectiveness acceptability curve

Cost-effectiveness acceptability curve

Y-axis: Proportion at incremental net benefits>0

X-axis: Willingness-to-pay per QALY gained (€ 5,000 to € 1,95,000)
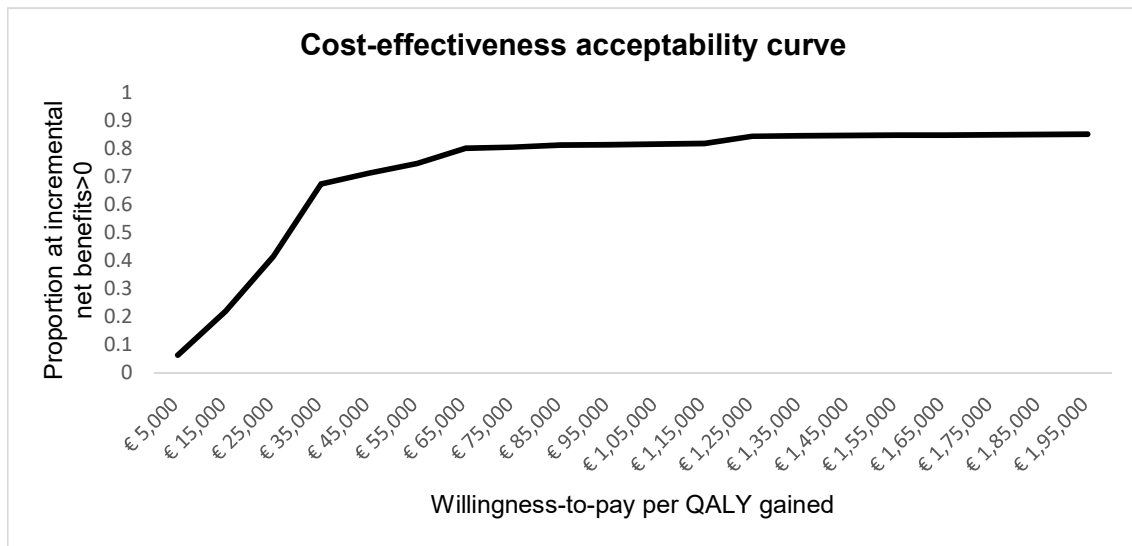
1

Table 1: Baseline characteristics of trial participants by group for five participating countries

	DIALOG+ intervention (N=236)	Standard care (N=232)	Overall sample (N=468)
Age in years (Mean, SD)	44.34 (11.09)	40.81 (11.26)	42.59 (11.30)
Sex (% female)	103 (43.64%)	111 (47.84%)	214 (45.73%)
Countries (N, %)			
Bosnia and Herzegovina	40 (16.95%)	41 (17.67%)	81 (17.31%)
Kosovo (UN Resolution)	52 (22.03%)	51 (21.98%)	103 (22.01%)
Montenegro	62 (26.27%)	60 (25.86%)	122 (26.07%)
North Macedonia	41 (17.37%)	41 (17.67%)	82 (17.52%)
Serbia	41 (17.37%)	39 (16.81%)	80 (17.09%)
Marital status (N, %)			
Single	121 (51.27%)	133 (57.33%)	254 (54.27%)
Married/Co-living/Any partnership	66 (27.97%)	59 (25.43%)	125 (26.71%)
Separated/Divorced	38 (16.10%)	37 (15.95%)	75 (16.03%)
Widow/Widower	11 (4.66%)	3 (1.29%)	14 (2.99%)
Educational level (N, %)			
Less than elementary school	2 (0.85%)	7 (3.02%)	9 (1.92%)
Elementary school graduate	49 (20.76%)	30 (12.93%)	79 (16.88%)
High school graduate	139 (58.90%)	144 (62.07%)	283 (60.47%)
University/College graduate	40 (16.95%)	45 (19.40%)	85 (18.16%)
Postgraduate/professional qualification	4 (1.69%)	4 (1.72%)	8 (1.71%)
Other qualification	2 (0.85%)	2 (0.86%)	4 (0.85%)
Employment status (N, %)[1]			
Paid employment	29 (12.29%)	39 (16.81%)	68 (14.56%)
Sheltered employment	1 (0.42%)	1 (0.43%)	2 (0.43%)
Training/Education	7 (2.97%)	13 (5.60%)	20 (4.28%)
Unemployed	140 (59.32%)	139 (59.91%)	279 (59.74%)
Retired	54 (22.88%)	39 (16.81%)	93 (19.91%)
Other	4 (1.69%)	1 (0.43%)	5 (1.07%)
State benefits (N, %)[2]			
No	128 (54.24%)	138 (59.48%)	266 (56.84%)
Yes	106 (44.92%)	90 (38.79%)	196 (41.88%)

1: There is one observation missing in the DIALOG+ group with N = 235.
2: There are two observations missing in the DIALOG+ group with N = 234 and four observations missing in the standard care group with N = 228.

Table 2: Mean resource use in quantities over the first 6 months of the trial by group

	DIALOG+ intervention (N=236)		Standard care (N=232)	
	N[1]	Mean (SD) [min, max]	N[1]	Mean (SD) [min, max]
Inpatient service				
Voluntary admission to psychiatric hospital (days)	206	2.00 (14.15) [0, 180]	218	1.20 (7.03) [0, 60]
Involuntary admission to psychiatric hospital (days)	206	0.54 (4.41) [0, 54]	218	0.06 (0.62) [0, 7]
Admission to hospital for physical health (days)	206	0.19 (1.50) [0, 15]	218	0.05 (0.50) [0, 7]
Primary/community service[2]				
General Practitioner (visits)	206	3.31 (4.81) [0, 48]	218	3.41 (3.80) [0, 22]
Psychiatrist (visits)	206	2.53 (3.25) [0, 19]	218	1.73 (2.77) [0, 24]
Psychologist (visits)	206	0.80 (3.39) [0, 24]	218	1.42 (8.15) [0, 96]
Dentist (visits)	205	0.55 (1.71) [0, 20]	218	0.70 (1.62) [0 ,10]
Emergency service (visits)	205	0.08 (0.38) [0, 3]	214	0.09 (0.51) [0, 5]
Other mental health professional (visits)	206	1.77 (5.51) [0, 48]	218	5.41 (17.46) [0, 120]
Other specialist doctor (visits)	205	0.65 (2.56) [0, 24]	218	0.53 (1.44) [0, 12]
Patients' other costs				
Lost work as physical health (days)	196	1.16 (8.40) [0, 90]	198	0.35 (2.45) [0, 30]
Lost work as mental health (days)	197	2.54 (18.66) [0, 180]	198	1.21 (13.00) [0, 180]
Medicine (euros)	206	237.23 (234.34) [0, 1598.10]	218	243.35 (509.97) [0, 6169.04]

1: N refers to the number of participants who responded to each question.

2: Those contacts do not include care that participants received in the IMPULSE trial.

Table 3: Mean costs (euros) for resource use over the first 6 months of the trial by group with Purchasing Power Parity adjusted

	DIALOG+ intervention (N=236)		Standard care (N=232)		Difference (no adjustment)[1]	Difference (with adjustment)[2]
	N[3]	Mean (SD)	N[3]	Mean (SD)	Difference (95% CI)	Difference (95% CI)
Inpatient service						
Voluntary admission to psychiatric hospital (days)	206	52.40 (377.41)	218	32.78 (196.95)	19.62 (-30.32,91.52)	4.58 (-42.70, 76.63)
Involuntary admission to psychiatric hospital (days)	206	11.89 (96.50)	218	1.30 (13.62)	10.58 (0.82, 28.24)	11.35 (-0.20, 29.94)
Admission to hospital for physical health (days)	206	7.70 (62.84)	218	1.00 (10.86)	6.69 (0.39, 17.30)	9.12 (-0.98, 22.68)
Sub total	**206**	**71.98 (392.17)**	**218**	**35.09 (197.33)**	**36.89 (-13.51, 103.08)**	**29.45 (-21.42, 106.93)**
Primary/community service[4]						
General Practitioner	206	27.01 (42.49)	218	28.82 (42.60)	-1.81 (-9.49, 6.41)	0.29 (-6.58, 7.15)
Psychiatrist	206	64.29 (128.22)	218	36.27 (68.16)	28.02* (10.02, 48.20)	23.92* (9.71, 40.64)
Psychologist	206	19.83 (80.18)	218	33.84 (185.03)	-14.01 (-44.21, 7.45)	-19.69 (-48.72, 5.12)
Dentist	205	8.71 (27.02)	218	15.60 (56.77)	-6.89 (-18.22, -0.03)	-3.75 (-9.41, 1.56)
Emergency services	205	1.80 (8.48)	214	1.62 (8.92)	0.19 (-1.52, 1.83)	0.31 (-1.52, 1.84)
Other mental health professional	206	22.02 (68.62)	218	72.11 (231.14)	-50.09* (-86.04, -20.00)	-52.33* (-83.94, -25.13)
Other specialist doctor	205	15.07 (65.81)	218	11.99 (29.94)	3.09 (-5.08, 14.01)	2.70 (-6.70, 14.87)
Sub total	**205**	**158.63**	**214**	**202.94**	**-44.31**	**-50.08**

25

		(202.81)		(362.18)	(-106.03, 5.60)	(-105.06, 3.90)
Patients' other costs						
Lost work by patients	196	169.28 (1125.73)	198	81.35 (753.46)	87.93 (-107.38, 289.20)	106.81 (-84.55, 307.61)
Medication	206	237.23 (234.34)	218	243.35 (509.97)	-6.12 (-92.56, 55.65)	37.03 (-40.88, 78.90)
DIALOG+/Standard care treatments						
DIALOG+ training	236	14.69 (9.13)	-	-	-	-
Other staff support for DIALOG+	236	1.24 (2.42)	-	-	-	-
Provision of DIALOG+/standard care	236	50.92 (62.63)	232	20.22 (21.14)	-	-
Other equipment	236	6.59 (9.02)	232	0.04 (0.08)	-	-
Other key resources	236	17.66 (10.94)	232	0.61 (1.14)	-	-
Sub total	**236**	**91.11 (62.86)**	**232**	**20.87 (20.71)**	-	-
Total costs with productivity lost	**195**	**714.49 (1247.26)**	**194**	**584.44 (986.27)**	**130.05 (-81.79, 352.24)**	**154.65 (-110.94, 422.73)**
Total costs without productivity lost	**205**	**565.95 (516.45)**	**214**	**497.78 (642.55)**	**68.17 (-54.26, 168.60)**	**98.42 (-29.49, 208.30)**

1: Independent t-tests are reported; 95% CI was produced using bootstrapping method with 1,000 replications; * P value is <0.05.

2: Mixed-effect model with baseline cost and covariates (patients' age, ICD code, and clinicians' profession) controlled. 95% CI was produced using bootstrapping replication for 1,000 times with bias corrected. * P value is <0.05.

3: N refers to the number of participants who responded to each question.

4: Those contacts do not include care that participants received in the IMPULSE trial.

Table 4: Comparisons of EQ-5D-5L index scores, MANSA scores, and ReQoL-10 sum scores by group

	DIALOG+ intervention (N=236)		Standard care (N=232)		Difference (no adjustment)[1]	Difference (with adjustment)[2]
	N[3]	Mean (SD) [min, max]	N[3]	Mean (SD) [min, max]	Difference (95% CI)	Difference (95% CI)
EQ-5D-5L						
Index at baseline	235	0.891 (0.16) [0.173, 1]	232	0.927 (0.13) [0.008, 1]	-0.0351* (-0.0609, -0.0088)	-
Index at 6 months	206	0.934 (0.13) [-0.141, 1]	218	0.935 (0.12) [0.075, 1]	-0.0005 (-0.0290, 0.0190)	0.0140 (-0.0083, 0.0355)
QALYs over 6 months[4]	**206**	**0.458 (0.06) [0.095, 0.5]**	**218**	**0.465 (0.050) [0.195, 0.5]**	**-0.0074 (-0.0190, 0.0027)**	**0.0035 (-0.0021, 0.0089)**
MANSA						
At baseline	236	4.480 (0.95) [1.917, 7]	232	4.537 (0.96) [1.083, 6.833]	-0.0576 (-0.2304, 0.1242)	-
At 6 months	206	4.839 (0.98) [2, 6.917]	218	4.649 (0.97) [1, 7]	0.1896* (0.0061, 0.3645)	0.1810* (0.0315, 0.3158)
ReQoL-10						
At baseline	236	25.661 (8.13) [1, 40]	232	25.672 (8.51) [2, 40]	-0.0114 (-1.6213, 1.3952)	-
At 6 months	206	27.170 (7.88) [2, 40]	218	26.161 (8.31) [3, 40]	1.0094 (-0.6621, 2.4348)	0.7237 (-0.2798, 1.9375)

1: Independent t-tests are reported; 95% CI was produced using bootstrapping method with 1,000 replications; * P value is <0.05.

2: Mixed-effect model with baseline outcome measure and covariates (patients' age, ICD code, and clinicians' profession) controlled. 95% CI was produced using bootstrapping replication for 1,000 times with bias corrected. * P value is <0.05.

3: N refers to the number of participants who responded to each question.

4: Formula used to calculate QALYs over 6 months: QALY = 0.25 X (index at baseline + index at 6 months).

Table 5: Cost-effectiveness analysis for point estimate of the ICER and sensitivity analyses

	Differences (95% CI)	ICER [1]	One to three times GDP per capita in euros [2,3]
Base case analysis (EQ-5D-5L at 6 months)			
Costs	84.17 (-8.18, 176.52)	€26,347.61	4,587 – 13,761
Outcomes	0.0032 (-0.0015, 0.0079)		
Sensitivity analysis 1 (Complete case analysis)			
Costs	98.42 (-48.08, 244.91)	€28,062.05	4,587 – 13,761
Outcomes	0.0035 (-0.0031, 0.0101)		
Sensitivity analysis 2 (Seemingly Unrelated Regression)			
Costs	66.09 (-44.86, 177.05)	€19,667.97	4,587 – 13,761
Outcomes	0.0034 (-0.0024, 0.0091)		
Sensitivity analysis 3.1 (minimum drug price)			
Costs	63.18 (-68.37, 194.73)	€18,649.54	4,587 – 13,761
Outcomes	0.0034 (-0.0031, 0.0099)		
Sensitivity analysis 3.2 (maximum drug price)			
Costs	78.86 (-71.41, 229.14)	€22,767.93	4,587 – 13,761
Outcomes	0.0035 (-0.0030, 0.0099)		
Sensitivity analysis 4 (societal perspective)			
Costs	105.48 (-136.19, 347.15)	€31,303.61	4,587 – 13,761
Outcomes	0.0034 (-0.0031, 0.0099)		
Sensitivity analysis 5 (ReQoL-10 as outcome measure)			
Costs	85.30 (-45.63, 216.22)	€119.02	
Outcomes	0.72 (-0.4880, 1.9212)		
Sensitivity analysis 6 (MANSA as outcome measure)			
Costs	89.06 (-41.91, 220.03)	€523.53	
Outcomes	0.17 (0.01, 0.33)		
Sensitivity analysis 7.1 [4]			
Bosnia perspective		€22,464.30	4,199 – 12,597
Sensitivity analysis 7.2 [4]			
Kosovo (UN Resolution) perspective		Dominant	3,036 – 9,108
Sensitivity analysis 7.3 [4]			

Montenegro perspective	€30,514.02	6,124 – 18,372
Sensitivity analysis 7.4		
North Macedonia perspective	€61,293.59	4,139 – 12,417
Sensitivity analysis 7.5		
Serbia perspective	€47,205.13	5,095 – 15,285

1: Measure for outcomes was ReQol-10 sum scores in sensitivity analysis 5 and MANSA scores in sensitivity analysis 6. Outcome measure for all other analyses in Table 5 used QALYs.

2: For base case analysis and sensitivity analyses 1 to 4, GDP per capita was calculated as the weighted GDP per capita of the five participating countries. The weights were proportions of participants from each country out of the total trial sample size. The formula used was: (€4198.69 x 17.31 + €3036.39 x 22.01 + €4139.38 x 17.52 + €6123.57 x 26.07 + €5094.54 x 17.09)/100 = €4,587. Three times of the GDP per capita was therefore calculated using €4,587 x 3 = €13,761.

3: For sensitivity analyses 7.1 to 7.5, GDP per capita was country-specific.

4: For sensitivity analyses 7.1 to 7.5, we ran two regressions for each analysis including a structural cost regression and a QALY outcome regression. Country-perspective ICER was calculated using coefficients from three interactions terms of the two regressions. We followed the method proposed by Willke et al (1998).