Reliability Analysis of a Newly Developed Questionnaire for Quality Control of Follow-up Visits in PolyIran Study

Gholamreza Roshandel MD PhD¹, Mohammad Reza Ostovaneh MD², Hossein Poustchi MD PhD³, Fatemeh Malekzadeh MD¹, Sadjad Ghajarieh Sepanlou MD PhD², Mohammad Reza Honarvar MD PhD³, Shahryar Semnani MD¹, Shahin Merat MD², Reza Malekzadeh MD²

Abstract

Background: The PolyIran study is a large-scale pragmatic cluster randomized controlled trial of fixed-dose combination therapy (Polypill) for prevention of cardiovascular diseases (CVD) in Iran. The PolyIran Quality Control Program (PIQCP) including a new questionnaire was developed to assess the quality of data collection during follow-up visits. The aim of this study was to assess the inter-rater reliability of PIQCP questionnaire.

Methods: The study was conducted in 26 (11%) randomly selected clusters (from a total of 236 PolyIran clusters). All participants within these 26 clusters were enrolled. The quality scores were measured according to the PIQCP guidelines by two independent raters. The intraclass correlation coefficients (ICC) were measured. In addition, the quality scores were categorized into good (>70%) and poor (<70%). The kappa coefficient was used to assess inter-rater agreement for this categorical quality scores.

Results: A total number of 945 PolyIran participants were enrolled of which, 501 (53%) were from intervention arm. In 934 participants (98.8%), the quality score could be successfully identified by both raters. The ICC (95%CI) of the overall quality scores was 0.985 (0.983–0.987). It was 0.976 (0.972–0.980) and 0.988 (0.986–0.990) in intervention and control arms, respectively. We found excellent agreement between the two raters in identifying participants with good and poor quality scores (kappa = 0.988, P < 0.001). The kappa values were 0.972 (P < 0.001) and 1.000 (P < 0.001) in intervention and control arms, respectively.

Discussion: Our results suggested that the PIQCP questionnaire is a reliable tool for assessing quality of data collection in PolyIran follow-up visits. Using this measure will help us in efficient monitoring of the PolyIran follow-ups and may ensure high quality data.

Keywords: Follow-up visit, PolyIran study, quality control, reliability

Introduction

Cardiovascular diseases (CVD) are leading causes of death worldwide.¹-³ A fixed-dose combination of established drugs (polypill) was suggested as an effective strategy for controlling CVD.⁴ This approach may reduce costs of treatment and increase participants’ adherence and consequently will reduce CVD incidence and mortality. Regarding the high rates of CVD mortality in Golestan Cohort Study (GCS),⁵ it was suggested to consider the polypill strategy for primary and secondary prevention of CVD in this population. The GCS is a prospective population-based long-term study with 50,000 participants which started 10 years ago, primarily to study esophageal cancer which was exceptionally common in Northeastern Iran (eastern part of the Golestan province) with 99% follow-up success rate. CVD was found to be the etiology of death in 50% of all mortalities among GCS participants. The results of the pilot phase of polypill study suggested that it is feasible and even necessary to conduct a large-scale trial on GCS population.⁶ Therefore, the main phase of polypill study, called “PolyIran study”, was designed. The PolyIran study is a pragmatic cohort multiple cluster randomized trial (cmRCT) to assess the safety and efficacy of a single fixed-dose combination of cardiovascular medications intervention for primary and secondary prevention of major cardiovascular events in middle-aged and elderly Iranians. The details of the PolyIran study have been described elsewhere.¹ The enrolment phase started in February 2011 and was completed by April 2013, with each village considered a cluster. The follow-up phase of the PolyIran project began in May 2011. The intervention arm clusters were followed up at months 1, 2, 3, 6, and then at 6-month intervals whereas the control arm clusters were followed up at months 3, 6, and subsequently at 6-month intervals.

During follow-up visits, different types of information were collected including participants’ demographic data, pill counts, reasons for non-adherence, blood pressure measurement, new medications, symptoms, referrals and cardiovascular events. The full list of information collected in PolyIran follow-up visits is shown in Table 1.

Quality assurance and quality control constitute an important part of any clinical trial.⁷,⁸ Because of the importance and pivotal role of follow-up visits in the PolyIran study, the PolyIran Quality Control Program (PIQCP) was developed to assess the quality of data collection during PolyIran follow-up visits. It contained two questionnaires, one for the intervention group, and the other for

Authors’ affiliations: ¹Golestan Research Center of Gastroenterology and Hepatology, Golestan University of Medical Sciences, Gorgan, Iran. ²Digestive Disease Research Center, Digestive Disease Research Institute, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran. ³Department of Health, Golestan University of Medical Sciences, Gorgan, Iran.

Corresponding author and reprints: Hossein Poustchi MD, Digestive Disease Research Center, Digestive Disease Research Institute, Shariati Hospital, Tehran University of Medical Sciences, Shariati Hospital, 1411713135, Tehran, Iran. Telefax: +982182415104, E-mail: h.poustchi@gmail.com

Accepted for publication: 15 May 2016
the control group asking about the quality of data collection on all of the follow-up activities (Table 1). The questionnaires were completed by an experienced staff after reviewing the PolyIran follow-up data collection forms and by telephone interviews with followed up participants. A PIQCP guideline was also developed, in which the methods of scoring the quality of data were described in detail. Because the PIQCP questionnaire was a new questionnaire developed for the PolyIran, we decided to validate this quality control measure. Validation studies mainly assess the validity and reliability of questionnaires. The different aspects of validity of the PIQCP questionnaire (content validity, face validity and construct validity) were assessed and approved by experts and experienced individuals in this field. In the present study, we aimed to determine the inter-rater reliability (agreement) for the PIQCP questionnaire in assessing quality of data collection in PolyIran follow-up visits.

### Materials and Methods

#### Study population

This study was done on the PolyIran study participants during May and September 2015. A total of 8410 participants were enrolled in PolyIran study, including 4233 participants (120 clusters) and 4177 (116 clusters) in the intervention and control arms, respectively.

#### Sample size

We aimed to enroll about 10% of the PolyIran participants (800 participants) in this validation study with the intervention/control ratio of 1/1.

#### Sampling

Considering a median cluster size of about 30 participant, we aimed to enroll 26 PolyIran clusters to achieve a total sample size of about 800 for the present study. According to the timetable of the PolyIran follow-up, 180 clusters (90 clusters from the control and 90 clusters from the intervention arms) were followed between May and September 2015. Of these followed clusters, 13 intervention and 13 control clusters were randomly selected for this validation study. We aimed to enroll all PolyIran participants within these clusters.

#### Data collection

For all participants within selected clusters, the quality scores were assessed (using the PIQCP questionnaire) by two independent raters within one week after the end of follow-up. The quality scores were measured by the PIQCP questionnaire according to the PIQCP guidelines (Table 1). Thus, the following process was considered to measure the quality score for each participant.

At first, the quality score of each activity (activity-specific quality score) was calculated (Table 1). For each of the PolyIran follow-up activities, a maximum score was earned if the data were correctly and completely collected according to the protocol of the PolyIran follow-up and there was no error. Otherwise, part of the score was earned (based on the level of incompleteness, errors, etc). For example, according to the protocol of the PolyIran follow-up, the section of “demographic data” included 5 items (first name, last name, participants ID, village code, date). So, if there was an error in one of these items, 80% of the maximum score would be earned for this activity, and so on. Then, the sum

### Table 1. Calculating activity-specific and overall quality scores for each participant in the PolyIran Quality Control Program (PIQCP).

<table>
<thead>
<tr>
<th>Activity code</th>
<th>Activity details</th>
<th>Expected (maximum) score (S_e)</th>
<th>Observed (calculated) score (S_o)</th>
<th>Quality score</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>To record participants’ demographic data</td>
<td>10</td>
<td>10</td>
<td>(S_o *100) / S_e</td>
</tr>
<tr>
<td>A2</td>
<td>To calculate and record number of Polypill tablets have been taken/ had to be taken by participants</td>
<td>35</td>
<td>—</td>
<td>(S_o *100) / S_e</td>
</tr>
<tr>
<td>A3</td>
<td>To record the reasons for non-adherence</td>
<td>15</td>
<td>—</td>
<td>(S_o *100) / S_e</td>
</tr>
<tr>
<td>A4</td>
<td>To assess and record blood pressure</td>
<td>35</td>
<td>50</td>
<td>(S_o *100) / S_e</td>
</tr>
<tr>
<td>A5</td>
<td>To refer hypertensive participants to health system physician</td>
<td>15</td>
<td>25</td>
<td>(S_o *100) / S_e</td>
</tr>
<tr>
<td>A6</td>
<td>To obtain and record data on newly occurred cardiovascular events</td>
<td>15</td>
<td>25</td>
<td>(S_o *100) / S_e</td>
</tr>
<tr>
<td>A7</td>
<td>To record drug history (Aspirin, Anti-lipid, Anti-hypertensive)</td>
<td>—</td>
<td>25</td>
<td>(S_o *100) / S_e</td>
</tr>
<tr>
<td>A8</td>
<td>To record new symptoms</td>
<td>30</td>
<td>40</td>
<td>(S_o *100) / S_e</td>
</tr>
<tr>
<td>A9</td>
<td>To record the participant’s follow-up status and reasons for loss to follow up</td>
<td>10</td>
<td>25</td>
<td>(S_o *100) / S_e</td>
</tr>
<tr>
<td>A10</td>
<td>To make and record final decision on drug continuation for the next 6 months</td>
<td>20</td>
<td>—</td>
<td>(S_o *100) / S_e</td>
</tr>
<tr>
<td>A11</td>
<td>To refer participants with adverse events (or any other problems) for physician visit</td>
<td>35</td>
<td>35</td>
<td>(S_o *100) / S_e</td>
</tr>
<tr>
<td>A12</td>
<td>To give the Polypill tablets to participants</td>
<td>35</td>
<td>—</td>
<td>(S_o *100) / S_e</td>
</tr>
<tr>
<td>A13</td>
<td>To prepare the list of participants needed to be visited by the study physician</td>
<td>15</td>
<td>15</td>
<td>(S_o *100) / S_e</td>
</tr>
<tr>
<td>A14</td>
<td>To give the participants information on lifestyle modification</td>
<td>30</td>
<td>50</td>
<td>(S_o *100) / S_e</td>
</tr>
</tbody>
</table>

Overall quality score \( \sum S_o = 300 \)
of scores of all activities (overall observed score) was calculated. Finally, the “overall quality score” was calculated by dividing the “overall observed score” by sum of maximum scores of all activities (overall expected score) (Table 1). Each rater was blinded to the results reported by the other.

Analysis of data
The activity-specific quality scores as well as the overall quality score were calculated for each participant using the above-mentioned formula (Table 1). The intraclass correlation coefficient (ICC) and its 95% confidence intervals (CI) were measured to assess the level of agreement between the raters. Bland-Altman graph10 was also used to assess the concordance of quality scores reported by rater 1 and rater 2.

In addition, the quality of data collection was categorized as good (quality score ³ 70%) or poor (quality score < 70%). The kappa coefficient was used to assess inter-rater agreement for this categorical quality scores. P values of less than 0.05 were considered significant.

Results
A total of 945 PolyIran participants were enrolled, of which, 472 (49.9%) were male and the mean (SD) age of participants was 60.01 (7.14) years. A total of 501 participants (53%) were recruited from the intervention arm. The number of male participants in the intervention and control arms was 246 (49.1%) and 226 (50.9%), respectively (P value = 0.58). The mean (SD) of participants’ age was 59.88 (7.21) and 60.27 (7.07) years in the intervention and control arms, respectively (P value = 0.40).

In 934 participants (98.8%), the quality score could be successfully identified by both raters. The mean (SD) of overall quality scores reported by raters 1 and 2 was 94.8% (13.9) and 95.1% (14.3), respectively. Table 2 shows the means of the overall quality scores in the intervention and control arms.

The quality of data was rated as good in 95.1% and 95.2% of participants by raters 1 and 2, respectively. The number (proportion) of participants with good and poor quality scores reported by raters 1 and 2 in the intervention and control arms is shown in Table 2.

The ICC (95% CI) of the overall quality scores reported by raters 1 and 2 was 0.985 (0.983–0.987). The ICCs (95% CI) of the overall quality scores in the intervention and control arms were 0.976 (0.972–0.980) and 0.988 (0.986–0.990), respectively. Table 3 shows the ICCs (95%CI)s of activity-specific quality scores reported by the two raters in the intervention and control arms.

Figure 1 shows the concordance of two raters in assessing quality scores in PolyIran participants.

Our results showed an excellent agreement between the two raters in identifying participants with good and poor quality scores (kappa = 0.988, P < 0.001). The kappa values were 0.972 (P < 0.001) and 1.000 (P < 0.001) in the intervention and control arms, respectively.

Discussion
The PolyIran study is a large-scale pragmatic randomized controlled trial.7 It was conducted at the community level and large number of healthcare workers (behvarzes) were involved in the process of data collection and follow-up. It was believed that follow-up visits were an important phase of the PolyIran study and quality of data collection in this phase could strongly affect the study results. To ensure high-quality data acquisition and reporting during PolyIran follow-up visits, a quality assurance and monitoring program (PIQCP) was designed for the PolyIran study. A new PIQCP-specific questionnaire was developed and the quality of data collection was regularly assessed according to the PIQCP guideline during each round of follow-up. The PIQCP questionnaire was a newly developed questionnaire and needed to be validated in the study population. The aim of this study was to assess the inter-rater reliability of PIQCP questionnaire in the PolyIran study.

Our findings suggested high inter-rater reliability values for the PIQCP questionnaire. Both quantitative and qualitative analyses showed high concordance for assessing the overall quality scores between the two raters (Table 2 and Figure 1). In addition, table 3 shows high ICCs for all activity-specific quality scores. Therefore, the PIQCP questionnaire was a reliable tool for assessing quality of data collection in PolyIran follow-up visits.

The effects of polypill strategy have been studied on CVD outcomes in different populations.11–15 But, similar to many other researches,16 little information was reported on measures of quality control and quality assurance for follow-up data collection in these studies. Quality control is one of the most important issues of clinical trials, especially those with large number of study sites and many different interviewers.17 Data of poor quality may have strong negative effects on the power of study.18,19 Many different factors may affect the quality of data in clinical researches, including study design, sample size, complexity of data collection, etc. Therefore, different methods may be considered to ensure high quality data in various studies.20,21 In addition, an individual study may need different quality control measures during each phase of the study. Marinez et al. considered an integrated data management system to ensure high quality data.22 In a study

Table 2. The mean (SD) of the overall quality scores as well as the number (proportion) of participants with good (quality score ≥ 70%) and poor (quality score<70%) quality scores reported by raters 1 and 2 in intervention and control arms of the PolyIran study.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Raters</th>
<th>Intervention arm (n = 492)</th>
<th>Control arm (n = 442)</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Rater 1</td>
<td>Rater 2</td>
<td></td>
</tr>
<tr>
<td>Mean (SD) of quality score</td>
<td></td>
<td>95.7 (10.3)</td>
<td>93.8 (15.1)</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>95.9 (10.2)</td>
<td>94.3 (15.7)</td>
<td>0.09</td>
</tr>
<tr>
<td>Number (%) of participants with good</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rater 1</td>
<td>473 (96.1)</td>
<td>415 (93.9)</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>Good quality</td>
<td>473 (96.1)</td>
<td>415 (93.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Poor quality</td>
<td>19 (3.9)</td>
<td>27 (6.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rater 2</td>
<td>474 (96.3)</td>
<td>415 (93.9)</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Good quality</td>
<td>474 (96.3)</td>
<td>415 (93.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Poor quality</td>
<td>18 (3.7)</td>
<td>27 (6.1)</td>
<td></td>
</tr>
</tbody>
</table>
Reliability of PolyIran Quality Control Program from Malawi, using a local quality assessment tool could result in improvements in data quality. In the Diabetes Control and Complications Trial and the follow-up Epidemiology of Diabetes Interventions and Complications study a quality assurance program was implemented to identify trends, data inconsistencies and process variability of results over time. Therefore, it is important to design and implement an appropriate quality control program for different phases of clinical researches, especially large-scale clinical trial.

In conclusion, we found high inter-rater accordance for data collection in PolyIran follow-ups and will consequently ensure high quality data in this large-scale pragmatic randomized controlled trial.

**Acknowledgment**

The authors would like to thank the PolyIran staff for their cooperation in data collection. The efforts of local healthcare workers (behvarzes) in the PolyIran study are also appreciated.
References


