PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The effect of Guided Self-Determination on self-management in persons with Type 1 diabetes mellitus and HbA1c ≥ 64 mmol/mol – a group-based randomized controlled trial
AUTHORS	Mohn, Jannike; Graue, Marit; Assmus, Jõrg; Zoffmann, Vibeke; Thordarson, Hrafnkell; Peyrot, Mark; Rokne, Berit

VERSION 1 - REVIEW

REVIEWER	Dominic Ehrmann Research Institute Diabetes Academy Bad Mergentheim, Germany
REVIEW RETURNED	17-Aug-2016

GENERAL COMMENTS	The authors evaluated an intervention program on guided self-determination (GSD) in a randomized controlled study. The main outcome, improvement of HbA1c, was not achieved but diabetes distress, autonomy-motivated behavior and self-esteem significantly improved in the intervention group. Main struggles of the study are problems with recruitment and high attrition in the intervention group resulting in a high self-selection of willing participants. The paper is clearly written and openly discusses the limitations and struggles of this study. However, there are some additional issues which should be addressed by the authors. • At the end of the introduction, several studies are cited that showed the efficacy of GSD with regard to glycemic control or diabetes distress and lack of motivation. The authors should make a stronger case for why this particular RCT is needed and what separates this study from previous studies. • Do the authors have any information on screening HbA1c that could be relevant in order to determine if patients' glycemic control was already improving or deteriorating at baseline? • The time of the follow-up assessment differed between groups. As stated in "assessments" the intervention group was assessed 9 months after the last session (7 sessions over 14 weeks) whereas the control group was assessed 9 months after inclusion; hence, the control group was assessed 14 weeks earlier than the intervention group. Considering that HbA1c is usually assessed every 3 months (at least in patients with type 1 diabetes), 14 weeks is a substantial difference. Can the authors comment on that? • In Figure 1 the authors state that 13 drop outs were due to non-attendances. Did those 13 participants not attend any session of the intervention or did they not attend the follow-up measurement? The authors should clarify this and the difference between "did not wish

	to participate". • Was the sum-score of the DDS transformed to a 0-100 scale like the PAID in order to allow for comparison between the two scores? • Using a linear regression to analyze group differences is quite unusual (albeit the same results as from an ANCOVA can be expected). Maybe the authors can shortly clarify why the decided to use the regression model. • Performing an intention-to-treat analysis only on HbA1c and "comparing" this result with the remaining per-protocol effects can be questioned. By doing so, the efficacy of the intervention with regard to HbA1c is based on a conservative approach while the efficacy with regard to the other outcomes is based on a liberal approach (due to self-selection). On the one hand, it would be interesting to see whether the per-protocol analysis of HbA1c leads to different results. On the other hand, intention-to-treat analyses of the remaining outcomes would be possible using the last-observation-carried-forward imputation method. In order to achieve a more complete picture of the efficacy, intention-to-treat and per-protocol analyses of all variables should be performed. • Why was the number of SMBG categorized into 3 groups? I think that a lot of information is lost due to the rather broad categorization (< daily vs. 1-6 vs. >7). From previous education studies, it can be estimated that the average number of SMBG in patients with type 1 diabetes is around 4 -5 measurements per day. Thus, patients below and above this estimated average fall into the same category. Can the authors comment on this? • Was there any effect of group size on the outcomes? • In the discussion, the authors wonder that despite "many life resources" people still had high distress. I would delete the aspect of "life resources" from the sentence and only describe the high level of distress. • The aspects of self-selection (high attrition rate in the intervention group) should be discussed a little bit more.

REVIEWER	Thomas Kubiak Johannes Gutenberg University Mainz, Germany
REVIEW RETURNED	18-Aug-2016

GENERAL COMMENTS	This is a report on a RCT testing an innovative group intervention for people with Type 1 Diabetes mellitus with the primary outcome of lowering HbA1C level. The manuscript is very well written and the, generally, the RCT has been carefully designed and conducted. The findings add to what has been known in the field, although conclusions and implications that can be drawn from the study are limited, because of the unfortunate high attrition, particularly in the intervention group. The following specific comments should be addressed: 1. One main issue with the study is - as the authors acknowledge themselves - the high attrition in the intervention group. While this

	cannot be remedied, I suggest that the authors go a bit further to shed some light on underlying reasons for attrition that go beyond comparisons of completers vs. non-completers. For instance, a (stepwise) logit model to prediction completion could offer some insights into the most prominent key baseline correlates of attrition. 2. I agree with the authors that patient education interventions in typ 1 diabetes are largely understudied compared to type 2 diabetes. However, mentioning only DAFNE and the Jena program may suggest that these two are the only ones that have been evaluated and published. This is clearly not the case, as there are other programs and research out there (e.g., PRISMA). Please give a more complete picture. 3. I am not overly fond of the labels "poorly controlled" or "poorly regulated" diabetes / glycemic control. If taken verbatim, this label may relate to anything from recurrent hypoglycemia to chronically elevated HbA1c levels. The focus of the present research was the latter - so please be specific. Minor point - Exclusion criteria: How were these checked? Which conditions did qualify as severe co-morbidities? - p. 6, paragraph 2: Definition of GSD needs a reference. - Regression analyses: are these b-coefficents or betas? This should be clear at first glance by using the appropriate symbol in the regression tables without having to judge the range of values that are reported. - Baseline characteristics: The sample appears to be quite heterogenous in terms of duavetes duration and long-term complications. Please comment and elaborate briefly in the discussions section.

REVIEWER	Dr Mayukh Samanta QIMR Berghofer Medical Research Institute Herston, Brisbane Australia
REVIEW RETURNED	21-Nov-2016

GENERAL COMMENTS	This research study conducted by Mohn and colleague is very well written. It satisfies all the major points to be accepted as a research article in BMJ Open. They carefully and clearly indicated all the CONSORT 2010 reporting checklist for reporting a RCT. The aim of the study and the appropriate statistical analyses were correctly stated and strength and limitations were also indicated clearly. My only concern is why the authors did not have any recent reference? The latest reference I find is 2009. Authors should address that in their revised version. I would like to accept this paper for publication otherwise.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

1. At the end of the introduction, several studies are cited that showed the efficacy of GSD with regard to glycemic control or diabetes distress and lack of motivation. The authors should make a stronger case for why this particular RCT is needed and what separates this study from previous studies.
RE: As also commented from reviewer 2 we have expanded the text to give a more distinct picture of previous research, please see pages 5-6. Still there is a lack of reports on empowerment-based interventions among people with Type 1 diabetes performed in collaborative multidisciplinary teams. The main strength of the present study is that it is performed in a busy clinical setting and claiming comprehensive training of the nurses. Another strength of this particular RCT is that it is theory based.

2. Do the authors have any information on screening HbA1c that could be relevant in order to determine if patients' glycemic control was already improving or deteriorating at baseline?
RE: We agree that this point is interesting, but unfortunately, that information was not available. However, random assignment to the study groups should eliminate this as a concern regarding group differences in outcomes.

3. The time of the follow-up assessment differed between groups. As stated in "assessments" the intervention group was assessed 9 months after the last session (7 sessions over 14 weeks) whereas the control group was assessed 9 months after inclusion; hence, the control group was assessed 14 weeks earlier than the intervention group. Considering that HbA1c is usually assessed every 3 months (at least in patients with type 1 diabetes), 14 weeks is a substantial difference. Can the authors comment on that?
RE: We agree, this is an interesting and important point. However, due to individual differences in days/weeks from point of randomization to start of intervention group, a clear post intervention standardization on the time of the follow-up assessment was difficult to obtain. We have commented on that at page 10.

4. In Figure 1 the authors state that 13 drop outs were due to non-attendances. Did those 13 participants not attend any session of the intervention or did they not attend the follow-up measurement? The authors should clarify this and the difference between "did not wish to participate".
RE: We appreciate this comment. Those 13 drop outs labelled 'non-attendance' did not attend any session of the intervention. The 20 participants labelled 'did not wish to participate' were persons who actively declined participation after randomization. In the intervention group all participants attended at the follow-up. We have clarified this by adding a more detailed description of the drop outs, please see Figure 1 page 8.

5. Was the sum-score of the DDS transformed to a 0-100 scale like the PAID in order to allow for comparison between the two scores?
RE: This is an interesting idea, but we did not transform the DDS scale to a 0-100 scale in the present study. We consider comparison between the two scores as out of scope of this article due to restricted word limits. Both instruments map individual levels of
diabetes-related emotional distress among people with
diabetes. Because of the one-solution factor structure of PAID this scale was better to use in this study with relatively high attrition challenges in the intervention group (compared to the four-factor structure of the DDS).

6. Using a linear regression to analyze group differences is quite unusual (albeit the same results as from an ANCOVA can be expected). Maybe the authors can shortly clarify why they decided to use the regression model.
RE: The regression analysis adjusted for baseline and treatment is in fact an ANCOVA.

7. Performing an intention-to-treat analysis only on HbA1c and “comparing” this result with the remaining per-protocol effects can be questioned. By doing so, the efficacy of the intervention with regard to HbA1c is based on a conservative approach while the efficacy with regard to the other outcomes is based on a liberal approach (due to self-selection). On the one hand, it would be interesting to see whether the per-protocol analysis of HbA1c leads to different results. On the other hand, intention-to-treat analyses of the remaining outcomes would be possible using the last-observation-carried-forward imputation method. In order to achieve a more complete picture of the efficacy, intention-to-treat and per-protocol analyses of all variables should be performed.
RE: We did not have post-intervention HbA1c data for participants who did not follow the protocol, therefore ITT (intention-to-treat) and PPA (per-protocol-analyses) are identical; we have used the term per-protocol-analysis to emphasize this fact. To clarify, the text has been modified, please see pages 12-13.

8. Why was the number of SMBG categorized into 3 groups? I think that a lot of information is lost due to the rather broad categorization (< daily vs. 1-6 vs. >7). From previous education studies, it can be estimated that the average number of SMBG in patients with type 1 diabetes is around 4 -5 measurements per day. Thus, patients below and above this estimated average fall into the same category. Can the authors comment on this?
RE: We are aware that the ADA’s Standards for diabetic care (2013) recommends frequent SMBGs (at least 3-4 times/day). Ideally, we did not want to categorize. However, due to small sizes of some categories we chose to do so. From a clinical perspective the broad categorization was chosen because we experience a rather substantial number of patients with chronically elevated HbA1c measuring only 1-3 times/day. In the current study group we found that 30% measured 1-3 times/day, 30% 4-6 times/day and 12% >= 7 times/day, whereas 16% measured less than every day, 7% less than every week and 5% had no monitoring last 14 days. We considered that dichotomizing would have been too unsubtle, therefore we decided to categorize into three subgroups.

9. Was there any effect of group size on the outcomes?
RE: This is an interesting perspective. Perform such analyses might provide greater insight with regard to the effect of intervention components (i.e., group) size on outcomes. However, with only 48 subjects in the intervention arm we did not have enough power to perform such analyses.

10. In the discussion, the authors wonder that despite “many life resources” people still had high. I would delete the aspect of “life resources” from the sentence and only describe the high level of distress.
RE: We have made the proposed changes, please see the two last lines at page 22.

11. The aspects of self-selection (high attrition rate in the intervention group) should be discussed a little bit more.
RE: The problem of attrition is addressed on page 24 (please see highlighted text, green colour).

Reviewer 2
1. One main issue with the study is - as the authors acknowledge themselves - the high attrition in the intervention group. While this cannot be remedied, I suggest that the authors go a bit further to shed some light on underlying reasons for attrition that go beyond comparisons of completers vs. non-completers. For instance, a (stepwise) logit model to prediction completion could offer some insights into the most prominent key baseline correlates of attrition.
RE: Prediction of completion is beyond the scope of the paper. Our goal in reporting comparisons of the completers vs. non-completers was to indicate the degree to which the completer population was representative of the included population.

2. I agree with the authors that patient education interventions in type 1 diabetes are largely

understudied compared to type 2 diabetes. However, mentioning only DAFNE and the Jena program may suggest that these two are the only ones that have been evaluated and published. This is clearly not the case, as there are other programs and research out there (e.g., PRISMA). Please give a more complete picture.
RE: We have carefully revised references and tried to add a more distinct picture, please see pages 5-6.

3. I am not overly fond of the labels "poorly controlled" or "poorly regulated" diabetes / glycemic control. If taken verbatim, this label may relate to anything from recurrent hypoglycemia to chronically elevated HbA1c levels. The focus of the present research was the latter - so please be specific.
RE: We have made the proposed changes throughout the article.

4. Minor point
- Exclusion criteria: How were these checked? Which conditions did qualify as severe co-morbidities?
- p. 6, paragraph 2: Definition of GSD needs a reference.
- Regression analyses: are these b-coefficents or betas? This should be clear at first glance by using the appropriate symbol in the regression tables without having to judge the range of values that are reported.
- Baseline characteristics: The sample appears to be quite heterogenous in terms of diabetes duration and long-term complications. Please comment and elaborate briefly in the discussions section.
RE: -Exclusion criteria were assessed through medical records, and the WHO's international classification of diseases (ICD-10) system was applied. Severe co-morbidity was defined: end-stage renal disease, severe heart failure, severe cancer and/or a medical history that included a major psychiatric diagnosis, (e.g. schizophrenia, severe depression, bipolar disorder). To clarify for readers an additional sentence and reference are stated at p.7, Recruitment.

- Reference p.6: we have clarified this

-Definition Betas or B-coefficients: We have clarified this in Table 2 by making a footnote explaining B = unstandardized coefficient

-If we understand the comments correctly concerning Baseline characteristics/heterogeneity:
A) duration: both range, median and means are reported; due to the wide range of age (18-55) there is a wide range of duration.
B) Long-term complications: range from 23-36%.
For both A and B: there is no significant difference between groups. We have noted the heterogeneity of the sample on page 22.

Reviewer 3
1. My only concern is why the authors did not have any recent reference? The latest reference I find is 2009. Authors should address that in their revised version.
RE: The following reference numbers are more recent than 2009: #2, 7, 10, 11, 13, 14, 15, 17, 18, 19, 24, 25, 28, 32, 34, 35, 36, 37, 38 and 39

VERSION 2 – REVIEW

REVIEWER	Dominic Ehrmann Research Institute of the Diabetes Academy Mergentheim (FIDAM), Germany
REVIEW RETURNED	12-Feb-2017

GENERAL COMMENTS	While the authors addressed each point in the letter, little changes have been made to the manuscript itself.

	- At the end of the introduction, it is still unclear why this particular study is necessary and what this study adds to the existing literature. By providing a more distinct picture of previous research did not establish the rationale of the study. - I think that the different time frame of follow-up measurement is a major point and has substantial impact on the primary outcome and the interpretation of the study. The authors should consider adding "time till follow-up" as a covariate in their analyses. - Usually, the mean item-score of the DDS and its subscales is reported. The authors even mentioned the range of 1-6 on page 11 while the PAID's range of 0-100 is mentioned. However, when reporting the sum-score it would be helpful to provide the scale range of the PAID and DDS in Table 1. The suggestion to transform the DDS to a scale of 0-100 was not meant to expand the scope of the article but rather to allow the reader to compare the two measures. - Interestingly, the authors claim that the PAID (due to its one factor solution) was better suited for this study than the DDS (due to its four-factor solution). While I cannot fully fathom their argument, what would this mean for the effect on DDS and DDS-emotional burden that is one main results of this study? - While ANCOVA and regression have a common statistical basis, presenting the between-group-effects of an RCT with B-coefficients is unusual. This way, the effects of the intervention (compared to the control group) are not easily accessible. Presenting the mean difference between the two groups adjusted for the covariates would simplify Table 2. - Not having post-intervention data for participants that dropped out after randomisation is one defining element of an intention-to-treat analysis. The conservative assumption would be that those who initially intended to participate did not change and therefore the respective baseline value is carried forward. In their letter, the authors state that they "did not have post-intervention HbA1c data for participants who did not follow protocol" but on page 12 of their manuscript they state that "only HbA1c was assessed for those who did not complete the study". I have a hard time understanding what the authors mean by "therefore, per-protocol and intention-to-treat analyses were identical for all outcomes except HbA1c" (page 12/13). By definition, per-protocol and intention-to-treat analyses comprise different samples and therefore cannot be identical (except no one dropped out and all participants perfectly followed protocol in which case an ITT is obsolete). - The categorisation of SMBG can be questioned altogether and I don't fully understand the reason for this. My initial comment aimed at using SMBG as a continuous variable as that would offer a between-group comparison of the absolute number of measurements per day. On a different note, 5% type 1 diabetes patients with no measurement on the last 14 days is hardly imaginable – did these patients use CGM? - The highlighted text discusses general aspects of targeting distressed persons with diabetes. However, in concordance with reviewer 2, I suggested to discuss possible underlying reasons for

	attrition and, thus, self-selection in this particular study. With an attrition rate that high, it should be in the scope of the article to address this problem rather that merely mentioning it as a limitation. - To highlight the random assignment as a strength of an RCT while the cited studies 17-20 were all RCTs does not demonstrate the strengths of the study in relation to other studies.

VERSION 2 – AUTHOR RESPONSE

-We have clarified this and developed the text further, please see pages 6-7, 9-10, 11, 12, 15 and 17-18

- We have revised to include the following as the rationale for this study. Please see pages 6-7:
1)Patient involvement and person-centered care are highly appreciated and recommended, but difficult to implement as part of clinical care.
2)Guided Self-Determination (GSD) is one of few interventions which clinicians are able to facilitate in routine clinical care after rather short training.
3)Patients have a primary role in GSD, spending their time at home clarifying what is important for them to change and becoming able to express their thoughts in communication with HCPs. Consequently, efficiency of patient-provider communication increases without extra use of HCP resources.
4)GSD has the potential for improvement of HbA1c, as well as increased patient self-determination and decrease of their diabetes-related burden.

-Thank you for your suggestion to control for variation in time to follow-up. In fact, the groups had different lengths of time from randomization to start of intervention in the IG (mean=4.9 months, SD=3.6), see pages 9-10. There was also some variation in time from baseline to follow-up in the control group (mean 10.9, SD=2.4). Unfortunately it is not possible to include follow-up time as an extra covariate in the models in Table 2 because of the strong correlation between group assignment and length of follow up. This will cause collinearity in the regression models. We therefore did linear regression analyses within each treatment group to test if there was an effect of time on the change in outcomes, i.e. if patients with longer follow-up experienced greater change. The results for the intervention group are given in a supplementary table (Table 4). We did not find any significant associations between length of follow up and change in outcome neither in the intervention group nor the control group. It does thus not look like the change was larger for patients with longer follow-up .
We have also made some adjustments in column headings, i.e. replaced ‘9 months’ with ‘follow-up’ in Table 2 and Table 3.
We’ve added information about differences in length of follow-up in the methods-section on page 12, in the results-section on page 15 and under weaknesses of the study on page 18.

- Thank you for pointing out the need to clarify this. We have revised the text of the manuscript to explain how the score was obtained (page 10-11). We agree that this will allow a direct comparison of PAID and DDS scores. We have also proposed additional text for other instruments to clarify, and we added the range of all instruments in the footnote of Table 1.

- We see that our last attempt to clarify this point was not very successful. The effect for PAID was much stronger than for overall DDS, presumably because the DDS measures components of distress were not as sensitive to the intervention. We have revised to indicate that the DDS allows us to identify the component of diabetes distress that was most sensitive to the intervention, i.e., emotional distress.

- The beta coefficient for group membership can be interpreted as the adjusted mean group

difference. We have revised Table 2 to make this clearer.

- We apologize for the confusion regarding this point, and have clarified on page 12 of the manuscript. For HbA1c we performed intention-to-treat (ITT) analysis because we had complete data. We performed per-protocol (PP) analysis for all the questionnaire-based outcomes, excluding patients with missing follow-up questionnaires. This has been clarified on page 12 in the manuscript.

- Certainly, using SMBG as a continuous variable would have been optimal as that would have offered a between-group comparison of the absolute number of measurements per day. However, when designing this study we considered the variation in demands of the disease to differ too much to be able to record SMBG that precisely. The need for measurement might vary from day to day and it can be difficult for the patient to give a valid estimate of the number of measurement per day. Some patients would then report a range instead of a number. Nevertheless, the categorization can absolutely be questioned and cause possible loss of information. We also tried another version with the following three categories:
We've added a paragraph about this in the discussion part on page 18.
It is correct that within the 9 patients reporting no measurements the last 14 days, 1 patient used CGM. Among the remaining 8 patients 5 did not use CGM and the last 3 had a note in the medical records stating that the patient did not perform self-monitoring.

- We agree that this matter should be in the scope of the article. It is an important methodological limitation and it is also important to discuss possible underlying reasons for the attrition such as the comprehensive and demanding intervention for the individual, especially with regard to pre-intervention work sheets that they were encouraged to fill in before each group session. We have added more text to further address this based on our data (see p. 17-18).

- Indeed, we were not explicitly comparing our study to those in references 17-20. We have deleted this text.

VERSION 3 – REVIEW

REVIEWER	Dominic Ehrmann Research Institute of the Diabetes Academy Mergentheim (FIDAM)
REVIEW RETURNED	04-Apr-2017

GENERAL COMMENTS	The authors did a very good revision and addressed each point satisfactorily. I have no further comments.