PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Real-world effectiveness of app-based treatment for urinary incontinence: a cohort study
AUTHORS	Rygh, Pontus; Asklund, Ina; Samuelsson, Eva

VERSION 1 – REVIEW

REVIEWER	Professor Douglas Tincello University of Leicester, UK
REVIEW RETURNED	14-Jul-2020

GENERAL COMMENTS	This is an interesting paper examining the "real world" effectiveness of an app based PFMT tool, previously shown to be effective in an RCT. In my view this paper requires minor revisions before it is ready for publication: Abstract The breakdown of your results is unclear; it would be better to present the overall change score first (as you do in the main text) and then indicate the score by severity. You should also include the definition of severity in your abstract. Methods p4, lines 54-56: the description of how you calculated downloads from Google play is not clear; please rewrite this p6, lines 34-35 (also p7, lines 51-52): the description of how you weighted the results to "match the results of the RCT" is not clear at all; particularly since the proportions of different categories of severity are very different (p 7, lines 12-13). A much fuller description of this process must be included and reviewed by a statistician before publication. Results Generally clear, except the section on weighting the results to the RCT and how this changed the data. I see no mention of how the results in the different severity groups compare to the minimal clinically important difference (MCID) of the ICIQ-UI SF. It seems essential to me to comment on how the degree of improvement recorded by respondents (especially the mildly affected group) relates to the MCID; without this you cannot confidently state the outcome is good! (page 9, line 36). The MCID of the ICIQ-SF is published here: https://doi.org/10.1002/nau.22533 Discussion This is over long and quite a lot of the content does not critically discuss the results; several paragraphs could be removed: paras 2-5 on page 8; para 2 on page 9. You need to discuss your results critically in the light of the MCID, and revise the bold conclusions accordingly.

REVIEWER	Marco Blanker, MD PhD dept. of General Practice and Elderly Care Medicine, University Medical Center Groningen, the Netherlands Please note my possible COI, being the project leader of a comparable mixed-method study on an app for stress, urgency and mixed UI in the Netherlands. Next, I have collaborated with Eva Samuelsson at two workshops on eHealth for incontinence at the annual meetings of the International Continence Societies. I feel that both aspects don't conflict my ability to provide sound feedback to the authors' work.
REVIEW RETURNED	22-Jul-2020

GENERAL COMMENTS	In this well-written manuscript authors report the real world experiences with the app Tät, that has been thoroughly developed and tested, and is to be used for urinary incontinence. I feel this report is important as for too many eHealth applications no studies are performed at all, or studies end with RCTs only. I do have some additional comments, meant to further improve this manuscript. The availability of the app has led to a huge number of downloads, showing the need for this kind of eHealth solutions. Unfortunately, authors experienced a very low response rate to the follow-up questionnaire. They have clearly provided information on the baseline differences between women with and without follow-up questionnaires. Notably, this showed that there were difference in demographic factors, but not in the incontinence related characteristics (Table 1). Could authors discuss this in more detail? Table 2 shows the possible discrepancy between app usage (frequency per day/week/month) and PFMT frequency. I feel that this finding needs more attention, as app usage itself is considered to be important, whereas for UI, the actual PFMT frequency is much more important. Some women appear to need only a few views on the app to perform PFMT. It would be of value to see if the main outcome (PGI-I) could be predicted at baseline, and how the PGI-I correlated with both app usage and PFMT frequency, although the latter may be biased towards no effect (need to continue PFMT is higher if UI is still present). Especially adding a prediction model to this study would help a lot for women that download the app, as it could be predicted if the UI will respond to the app-based treatment, or not (and if not, women could/should be advocated to seek care from their physician). Authors mention that the app has been downloaded from 94 countries. Could this be the result of presenting outcomes of the RCT on various international meetings, such as the annual meetings of the international continence society (ICS)? Do authors feel that these downloads really reflect patients that seek care, or could this also include care givers? The date of download for these different countries (not the main part from Sweden of course) could explain this, if this is associated to the international meeting dates.

	The app has been developed for the treatment of stressUI. Many of the women who have downloaded the app appeared to have urgency UI or mixed UI. How are these women served within the app? In the discussion, authors mention something about this. It would be good to see if the main outcome (PGI-I) differed between the UI-categories. It is to be expected that the effect for SUI will be larger than for mixed UI or urgency UI. Relevant differences might not be covered with the ICIQ-UI SF, so I suggest that PGI-I will be added here. (Figure 3 only shows the outcomes for different baseline severity categories, not type of UI.) Could authors explain why such a large number of downloads were from women without incontinence. Please correlate this to the reason for download (improve incontinence / preventative training), as this should also give some insight in how trustworthy the answers are. It is to be expected that these reasons would differ substantially between women with and without UI. Does the app make any reference to the preventative use of the app? If so, what statements are made with respect to the expected impact of app usage in this group? Is it advocated for preventative use, and what would be the basis for such claims?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Professor Douglas Tincello

Abstract
- The breakdown of your results is unclear; it would be better to present the overall change score first (as you do in the main text) and then indicate the score by severity.

Reply: To make the Result section in the Abstract clearer we have changed the text according to your advice by presenting the overall change score first and then the score by severity. The text has been changed from:

The ICIQ-UI SF score reduction at follow-up differed depending on the incontinence severity at baseline (slight 0.24 (95%CI: 0.06–0.42), moderate 1.41 (95%CI: 1.24–1.59), severe/very severe 3.23 (95%CI: 2.85– 3.61)). When the results were weighted to match the distribution of severity in the previous RCT, the overall ICIQ-UI SF score reduction was 2.2 compared to 3.9 in the RCT.

to:

The mean ICIQ-UI SF score reduction at follow-up was 1.31 (95%CI: 1.19-1.44) with a larger reduction in those with more severe incontinence at baseline (severe/very severe 3.23 (95%CI: 2.85–3.61), moderate 1.41 (95%CI: 1.24–1.59), slight 0.24 (95%CI 0.06–0.42)). When the results were weighted to match the distribution of severity in the RCT, the ICIQ-UI SF score reduction was 2.2 compared to 3.9 in the RCT.

- You should also include the definition of severity in your abstract.

Reply: Unfortunately, the number of words is very limited in the Abstract and we therefore choose to include the definition of severity only in the Methods section in the main article, where it is thoroughly described.

Methods
- p4, lines 54-56: the description of how you calculated downloads from Google play is not clear; please rewrite this

Reply: We have adjusted the text to make the description clearer. We changed the last sentence from:

From Google Play the exact number of downloads per month was available. This was adjusted by dividing the number of downloads in January 2018 by two and not counting the number of downloads in June 2019.

to:

Data regarding the exact number of downloads per month from Google Play were available from the Google Play services, we included half of the number of downloads for January 2018.

- p6, lines 34-35 (also p7, lines 51-52): the description of how you weighted the results to "match the results of the RCT" is not clear at all; particularly since the proportions of different categories of severity are very different (p 7, lines 12-13). A much fuller description of this process must be included and reviewed by a statistician before publication.

Reply: We have now included a more comprehensive description of how we weighted the results from the present study to match the severity and type of urinary incontinence in the previous RCT.
All changes have been discussed with an experienced statistician.

We have changed P6 lines 34-35 from:

The change in the ICIQ-UI SF score was compared with the findings from the previous RCT[13]. This score was the primary outcome in the RCT with data registered and reported at Clinical Trials (ID: NCT01848938). Using data from the RCT, the app results were weighted to match the distribution of severity and type of incontinence in the RCT (slight 2.4%, moderate 63.4%, severe 34.1%, all with SUI).

to:

The ICIQ-UI SF score was the primary outcome in the previous RCT, the data from which were registered and reported at Clinical Trials (ID: NCT01848938). The change in the ICIQ-UI SF score in the present study was compared with the findings from the previous RCT[13]. The distribution of severity in the RCT was: slight 2.4%, moderate 63.4%, severe 34.1%, and the type of incontinence was 100% SUI. In the present study, for women with SUI we weighted the results for each severity group to match this distribution.

Results
Generally clear, except the section on weighting the results to the RCT and how this changed the data.

Reply: We hope the clarification we have provided in the Method section on weighting the results to the RCT will make this clearer. We have also made changes in the Results section
p7 51-52 from:

When the results were weighted to match the distribution of type and severity of incontinence in the RCT, the score reduction was 2.2.

to:

When the results in the present study were weighted to match the distribution of type and severity of incontinence in the previous RCT, the ICIQ-UI SF score reduction was 2.2.

- I see no mention of how the results in the different severity groups compare to the minimal clinically important difference (MCID) of the ICIQ-UI SF. It seems essential to me to comment on how the degree of improvement recorded by respondents (especially the mildly affected group) relates to the MCID; without this you cannot confidently state the outcome is good! (page 9, line 36). The MCID of the ICIQ-SF is published here: https://doi.org/10.1002/nau.22533

Reply: We agree that the discussion about the MCID of the ICIQ-UI SF is very interesting, however, no MCID has been calculated either for the “real-world” setting that we are studying or for the mild types of incontinence dominant in our study. We are familiar with the reference by Sirls et al but we do not find that it is relevant as a comparison to our study since they calculated MCID for a completely different population that was undergoing surgery for SUI and had high pre-op scores for the ICIQ-UI SF. In our study, the major portion of the mildly affected group does not even reach a pre-treatment ICIQ-UI SF score of 5 (which is found to be the MCID in the article referred to above). We have now added the MCID for conservative treatment (Nyström et al) and we have changed the last paragraph in the Discussion from:

The ICIQ-UI SF score reduction differed depending on the severity of incontinence at baseline, with more severe incontinence being associated with a greater ICIQ-UI SF score reduction. However, the subjective experience of improvement according to the PGI-I was quite similar between the groups. This indicates that the level of the possible reduction in the ICIQ-UI SF score is highly dependent on the pre-treatment score which must be taken into account when comparing studies or interventions. A large proportion of the app users had slight incontinence with a very small reduction in their ICIQ-UI SF (0.24). The lowest level of frequency of leakage in the ICIQ-UI SF is "about once a week or less often" which makes the ICIQ-UI SF a poor measure of the effect in this group. The PGI-I correlates well with the reduction in the ICIQ-UI SF score in women with more severe incontinence[32]. Since 61.1% of users with slight incontinence answered that they experienced improvement on the PGI-I question, it is likely that they also had reduced symptoms even if this is not shown in the ICIQ-UI SF score.

To:

The ICIQ-UI SF score reduction differed depending on the severity of incontinence at baseline, with more severe incontinence being associated with a greater ICIQ-UI SF score reduction. However, the subjective experience of improvement according to the PGI-I was quite similar between the groups. This indicates that the level of the possible reduction in the ICIQ-UI SF score is highly dependent on the pre-treatment score which must be taken into account when comparing studies or interventions. The minimal clinically important difference (MCID) for the ICIQ-UI SF was found to be 2.5 in a previous study of a PFMT-based intervention. This was established by using the mean ICIQ-UI SF score reduction of those who answered “a little better” on the PGI-I[30]. However, this group had more severe incontinence with a mean pre-treatment ICIQ-UI SF score of 10.2. In our study, a large proportion of the app users had slight incontinence with a very small reduction in their ICIQ-UI SF (0.24). The lowest level of frequency of leakage in the ICIQ-UI SF is "about once a week or less often" which makes the ICIQ-UI SF a poor measure of the effect in this group. Since 61.1% of the users with slight incontinence answered that they experienced improvement on the PGI-I question, it is likely that they also had reduced symptoms even if this is not shown in the ICIQ-UI SF score. To our knowledge, there is no other comparable study, and the best estimate of clinically relevent improvement in the

population of this study is probably that 65% reported on the PGI-I that they experienced improvement.

We agree that the statement on page 9, line 36 was a little bold and out of context there so we have removed it from that paragraph and have instead tried to interpret our results under Conclusion where we have changed the text from:

App-based self-management of UI is an effective first-line treatment even in the real world. Larger reductions in symptom scores are expected the more severe the incontinence is at baseline. App-based treatment is easy accessible and can be used for women that want to manage their symptoms on their own, as well as a complement to other actions within regular health care.

to:

App-based self-management of UI is an effective first-line treatment even in the real world and the symptoms of two-thirds of the users improve. However, the improvement in the ICIQ-UI SF score reduction, as well as the percentage that experienced subjective improvement, is lower in the real world than in the RCT. Larger reductions in symptom scores are expected the more severe the incontinence is at baseline. App-based treatment is easy accessible and can be used for women that want to manage their symptoms on their own, and also as a complement to other actions within regular health care.

Discussion

- This is over long and quite a lot of the content does not critically discuss the results; several paragraphs could be removed: paras 2-5 on page 8;

Reply: These paragraphs include strengths and limitations of the study which we prefer to keep in this form.

- para 2 on page 9.

Reply: We have revised and shortened this paragraph from:

In a general Swedish population, about 30% of women have university as their highest level of education[28], but among the app users, 65% had university education and among those included in follow-up, as many as 75% were university educated. It is known that highly educated people use health apps more[29]. In a previous study, the level of education was not a factor in performing PFMT correctly after brief instructions[30]. The proportion of Swedish app users who lived in rural areas was similar to the proportion in the general population of Sweden[31].

to:

Among the app users, 65% had university education compared to the general Swedish population in which about 30% of women have university as their highest level of education[28]. It is known that highly educated people use health apps more[29].

- You need to discuss your results critically in the light of the MCID, and revise the bold conclusions accordingly.

Reply: We have revised our somewhat bold conclusions as mentioned earlier and changed the Conclusion from:

App-based self-management of UI is an effective first-line treatment even in the real world. Larger reductions in symptom scores are expected the more severe the incontinence is at baseline. App-based treatment is easy accessible and can be used for women that want to manage their symptoms on their own, as well as a complement to other actions within regular health care.

to:

App-based self-management of UI is an effective first-line treatment even in the real world and the symptoms of two-thirds of the users improve. However, the improvement in the ICIQ-UI SF score reduction, as well as the percentage that experienced subjective improvement, is lower in the real world than in the RCT. Larger reductions in symptom scores are expected the more severe the incontinence is at baseline. App-based treatment is easy accessible and can be used for women that want to manage their symptoms on their own, and also as a complement to other actions within regular health care.

We have also made some changes in the Discussion according to the comments about comparison with RCT results.
p9, line 34-6 from:

When the results were adjusted for the type of incontinence and severity, app users had a mean reduction in the ICIQ-UI SF score of 2.2, this compares with a mean reduction of 3.9 in the RCT. We consider this a good result.

to:

When the app results were weighted to match the type of incontinence and severity in the RCT, app users had a mean reduction in the ICIQ-UI SF score of 2.2, this compares with a mean reduction of 3.9 in the RCT.

Reviewer: 2
Marco Blanker, MD PhD

- The availability of the app has led to a huge number of downloads, showing the need for this kind of eHealth solutions. Unfortunately, authors experienced a very low response rate to the follow-up questionnaire. They have clearly provided information on the baseline differences between women with and without follow-up questionnaires. Notably, this showed that there were difference in demographic factors, but not in the incontinence related characteristics (Table 1). Could authors discuss this in more detail?

Reply: We have not discussed this finding further since it was not the major focus of this article. We are however interested in the factors that can predict completion of the program or successful treatment and this will be analysed in a separate article.

- Table 2 shows the possible discrepancy between app usage (frequency per day/week/month) and PFMT frequency. I feel that this finding needs more attention, as app usage itself is considered to be important, whereas for UI, the actual PFMT frequency is much more important. Some women appear to need only a few views on the app to perform PFMT. It would be of value to see if the main outcome (PGI-I) could be predicted at baseline, and how the PGI-I correlated with both app usage and PFMT frequency, although the latter may be biased towards no effect (need to continue PFMT is higher if UI

is still present). Especially adding a prediction model to this study would help a lot for women that download the app, as it could be predicted if the UI will respond to the app-based treatment, or not (and if not, women could/should be advocated to seek care from their physician).

Reply: The main objective of this article was to describe the user characteristics and change in symptoms in an unselected group of women that had downloaded the freely available Tät app, and compare these results with data from the RCT. We agree that a prediction model would be interesting and such analyses and discussions are being prepared for a separate article.

- Authors mention that the app has been downloaded from 94 countries. Could this be the result of presenting outcomes of the RCT on various international meetings, such as the annual meetings of the international continence society (ICS)? Do authors feel that these downloads really reflect patients that seek care, or could this also include care givers? The date of download for these different countries (not the main part from Sweden of course) could explain this, if this is associated to the international meeting dates.

Reply: We have not looked at any such association. However, we have excluded the app users who stated that they downloaded the app because they were researchers, health care providers or just curious about the app (see the flowchart of the study).

- The app has been developed for the treatment of stress UI. Many of the women who have downloaded the app appeared to have urgency UI or mixed UI. How are these women served within the app? In the discussion, authors mention something about this.

Reply: The app has been designed and evaluated solely for SUI. To clarify this, we have added the following text in the first paragraph of the Method:

Before the users of the app could answer the questionnaires, they had to provide their consent by reading a short text and ticking a box to confirm that they were aware that their answers would be sent anonymously to our research database and that results from previous research on the app were based on women with SUI.

We have also changed the text in paragraph 7 in the Discussion from:

Although the app has been designed for SUI, PFMT is recommended as first-line treatment for all types of incontinence[1] and many women with UUI and MUI used the app.

to:
Although the app has been designed and evaluated solely for SUI, PFMT is recommended as first-line treatment for all types of incontinence[1] and many women with UUI and MUI used the app.

- It would be good to see if the main outcome (PGI-I) differed between the UI-categories. It is to be expected that the effect for SUI will be larger than for mixed UI or urgency UI. Relevant differences might not be covered with the ICIQ-UI SF, so I suggest that PGI-I will be added here. (Figure 3 only shows the outcomes for different baseline severity categories, not type of UI.)

Reply: We agree that this is an interesting aspect and we have also added the results for this analysis at the end of the Results section by adding the following text:

No major differences in improvement of the PGI-I were seen in the different types of incontinence (SUI: 65.0% UUI: 66.2% MUI: 66.3%).

To comment on these results, we also revised the following text in paragraph 7 under Discussion from:

PFMT has been shown to be effective primarily in SUI[11], but in this study, we found significant effects on all types of incontinence with no major differences in effect between different types of incontinence.
to:
PFMT has been shown to be effective primarily in SUI[11], but in this study, we found significant effects on all types of incontinence with no major differences in the ICIQ-UI SF score reduction or subjective improvement between different types of incontinence.

- Could authors explain why such a large number of downloads were from women without incontinence. Please correlate this to the reason for download (improve incontinence / preventative training), as this should also give some insight in how trustworthy the answers are. It is to be expected that these reasons would differ substantially between women with and without UI. Does the app make any reference to the preventative use of the app? If so, what statements are made with respect to the expected impact of app usage in this group? Is it advocated for preventative use, and what would be the basis for such claims?

Reply: We can not explain why such a large number of downloads were from women without incontinence other than the suggestion that there seems to be a big demand for such an app. There is no reference to preventive use in the app and it is clearly stated upon download that the app has been designed and evaluated for women with SUI.

We chose to define incontinence as those who responded that they had both some frequency of incontinence and some amount of leakage on the ICIQ-UI SF questionnaire. We believe it is natural that this definition can differ somewhat from how women define themselves, as incontinent or not, especially if they have mild symptoms.

VERSION 2 – REVIEW

REVIEWER	Marco Blanker dept. of General Practice and Elderly Care Medicine, University Medical Center Groningen, the Netherlands Same as in previous review: Please note my possible COI, being the project leader of a comparable mixed-method study on an app for stress, urgency and mixed UI in the Netherlands. Next, I have collaborated with Eva Samuelsson at two workshops on eHealth for incontinence at the annual meetings of the International Continence Societies. I feel that both aspects don't conflict my ability to provide sound feedback to the authors' work.
REVIEW RETURNED	22-Nov-2020

GENERAL COMMENTS	Authors have properly revised their manuscript in respons to the reviewers comments. Although I regret that some responses included referring to future manuscripts, and authors decided not to add new information, I feel that the manuscript is suitable for publication, but of course the editors will decide upon that.