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)	Psychological (co)morbidity in patients with psoriasis: The impact of pruritus and anogenital involvement on symptoms of depression and anxiety and on body dysmorphic concerns – a cross-sectional study
AUTHORS	da Silva, Neuza; Augustin, Matthias; Hilbring, Caroline; von Stülpnagel, Catharina C.; Sommer, Rachel

VERSION 1 – REVIEW

REVIEWER	Langan, Sinead London School of Hygiene & Tropical Medicine
REVIEW RETURNED	28-Sep-2021

GENERAL COMMENTS	This is a cross-sectional study of adults with psoriasis from a single secondary care centre in Germany that aims to assess the effects of pruritus intensity and anogenital involvement on disease and mental health outcomes, and identify the variables associated with developing depression, anxiety, and body dysmorphia. Overall, the study design and methods used are appropriate to address the research objectives and the conclusions made are justified by the results. I have summarised what I feel are revisions that will improve the understanding of the paper. Please find my detailed comments below: 1. Introduction: The introduction provides a clear rationale for the study, and a detailed examination of the existing literature. However, you state that psoriasis affects “approximately 2.5% of the German population" without citing any literature that gives this figure. 2. Materials and methods: a. Study design and participants: The authors state that “patients were excluded if they presented any condition…that would confound the interpretation of the results”. It would be beneficial for the reader if the authors stated the other conditions that patients were excluded for and how these conditions may confound the interpretation of results, especially since people with comorbidities were included in the study. b. Outcome measures: The authors describe none/mild pruritus as an NRS score <=3 and moderate/severe pruritus as an NRS score>=4. It would be useful if the methods used to define cut-off points for none/mild and moderate/severe pruritus, as well as anogenital involvement were explained Were the categories decided arbitrarily, or where they based on previous studies/literature? The authors refer to other literature for the PROs. I think it would be of value to put a brief summary in the supplementary material, e.g,. the Patient Benefit Index will not be

	familiar to most- I had to look it up. I think adding some text so readers can know what the PBI is would be very useful. c. Statistical analyses: The authors present effect sizes for the comparison analyses, but do not describe how these numbers were calculated. d. Multivariable analyses: Only factors significant in univariable analyses were included in multivariable analyses. I wondered why a priori confounders were not included. 3. Results a. Were there any differences between those completing and not completing the questionnaires? b. For PROs, would median be more appropriate than mean, i.e., are these normally distributed? c. For tables 3-5, would have been useful to document what factors were adjusted for in multivariable analyses in a footnote. Was any assessment done to determine if there was collinearity in models? d. The conclusion about an interaction between pruritus and anogenital involvement seems to be based on univariate analyses and hasn't been explored in the multivariable model to our knowledge. 4. Discussion: a. The authors suggest that "the accumulative burden of anogenital psoriasis and moderate/severe pruritus may trigger coping strategies that have a protective effect on the patients' mental health", but do not posit examples of what these coping strategies may be. It would be useful for the reader if you gave further examples of the coping strategies and how you think they should be addressed in further research. b. As the authors mention, the sample size for this study is small, resulting in wide confidence intervals in multivariable analyses. For example, in the multivariable analysis the effect estimate reported for biological treatment was 80.30 [1.30–4950.81]. This study is likely underpowered to assess all of its objectives. Focusing on the OR of 80 in the discussion without mentioning that the CI ranged from 1. To almost 5000 is problematic. You need the CIs here and would suggest editing the language. c. Similarly, the statement that the multivariate analysis showed that clinical anxiety is 89 % less likely as the PBI increases by one unit is based on an OR and CI of 0.11 [0.02–0.66]. The statement is strong given the width of the CI and this should be reframed. There are other similar examples in the discussion. d. There is a lot of focus on statistical significance and p values. I would be more inclined to focus on the effect estimates and CIs, https://www.nature.com/articles/d41586-019-00857-9. e. It would be good to discuss generalisability, as this study was conducted in a single secondary care centre and findings may not be applicable to people seen in primary care or different geographical regions.

REVIEWER	Fethney, Judith University of Sydney, School of Nursing
REVIEW RETURNED	25-Nov-2021

GENERAL COMMENTS	Thankyou for the opportunity to review 'Psychological (co)morbidity in patients with psoriasis: The impact of pruritus and anogenital involvement on symptoms of depression and anxiety and on body dysmorphic concerns – a cross-sectional observational study'.

	Overall, I found the paper well written and comprehensive in its analysis of clinical and psychosocial variables. I do have some questions relating to the interpretation of a few of the ORs, as well as the conclusion that can be drawn from the very wide 95% confidence interval. Introduction First paragraph, 1st sentence, line 22 '…taking into account the patient' needs…' Change to '…taking into account the patient's needs…' Materials and Methods Study design and participants: 'Patients were excluded if they presented any condition, including other inflammatory diseases or dermatologic conditions, which would place them at unacceptable risk due to participation in the study or would confound the interpretation of the results'. Are the authors only referring to physical conditions here? There were no exclusion criteria based on psychological conditions? Outcome measures: 'A set of questionnaires were completed by the physician and by the patient.' Is the physician the usual physician who treats the patient or a physician engaged for the specific purposes of the research? Statistical analyses SPSS is no longer known as the Statistical Package for the Social Sciences . Please change to The statistical analyses were conducted using IBM SPSS Statistics (SPSS, version 23.0, IBM Corp., Armonk, NY). A point on definitions. 'Multivariate' typically means that there are multiple outcomes in the one model. Multivariable is the appropriate term to use when there are multiple predictors. See https://academic.oup.com/ntr/article/23/8/1446/5812038 I am also not a fan of using the notation Mean +/- SD, as the SD relates to the average distance of each observation from the mean (the dispersion of the data) and is not intended to convey the plausible range of values for the mean. I would prefer to see means and SDs reported as Mean (SD). Results Table 1: There are 2 major groups and within each of these there are also 2 groups, so 4 groups in total. Are the p values based on comparisons between the 4 groups, or the two larger groups? If based on the 4 groups, then some of these would have been post-hoc tests (continuous data) or assessed with ASRs (adjusted standardised residuals, chi square). For example, the p value for the higher mean %BSA in the patients with moderate/severe pruritus and with anogenital involvement is <0.001. So if based on a post hoc test, was the p value <.001 for all comparisons? The SD for this group is particularly large, indicating considerable variation in the response of this group to this outcome variable. Just want to make sure I understand Table 2. For example, for the DLQI, a model was built with pruritis (none mild; moderate severe) as a predictor and anogenital involvement (None or anogenital psoriasis) also as a predictor, and intertriginous psoriasis, biologic treatment, and PASI as covariates. There was also an interaction

	term of pruritis * anogenital in the model. There is a main effect for pruritis with a medium effect size, no main effect for anogenital involvement and no interaction. The authors have not addressed whether the assumptions for ANOVA were met. ANOVA is known to be robust under violations of normality, but less so when group sizes are unequal (as the case in this study). I apologise if I am misunderstanding Table 3. The way the predictors are reported, eg Pruritus (none/mild vs. moderate/severe) or Biological treatment (no vs. yes), the condition appearing first is the reference to which the other condition is compared, eg none/mild is the reference to which moderate/severe is compared, and the OR and 95% reflected the increase (or decrease) in odds compared to the reference. So for Pruritus (none/mild vs. moderate/severe) the OR for having clinically significant depression is 5.83 (1.74 – 19.53), therefore, as the authors report, moderate/severe pruritis is associated with an increased risk of clinical depression (5.83 times the risk or 483% increase). All good. For biological treatment (assuming that no is the reference and the statistics relate to the 'Yes' condition, the OR is 0.30 (0.10 – 0.91). This indicates that those undergoing biological treatment have a reduced risk of clinical depression. Yet the authors state that this is associated with an increased risk. Similar for disease duration – I would interpret the result as each increment (yearly?) in duration, the risk of clinical depression decreases by 5%. The multivariable model seems a bit odd to me. Nagelkerke is relatively high, but there are only 2 statistically significant variables in the model, and one of them (biological treatment) has an incredibly wide 95% CI. Very wide CIs can be due to sparse data, but according to Table 1, there were 66 patients undergoing biological treatment, and therefore 41 not having this treatment, so sparse data doesn't seem to be the problem here. It also seems unusual that biological treatment in univariate analysis reduces the risk of clinical depression (OR 0.30 (0.10 – 0.91), so reduced it by 70%, but in multivariable analysis, controlling for the other variables, it increases the risk by a massive amount. A 95% CI this wide also renders the result unreliable in that it is hard to conclude anything about the effect of this treatment, other than it causes an increase. This should be mentioned in the Limitations.Did the authors check for multicollinearity? This can be checked using the linear regression option in SPSS, even though the outcome in logistic regression is dichotomous it doesn't matter when doing this check as the multicollinearity check is only concerned with the predictors, not the dependent variable. Discussion Page 13, line 36 'The univariate associations between PASI and disease duration as well as PASI and PHQ ≥ 3 corroborate this hypothesis.' However, I don't see any univariate associations between PASI and disease duration, only between PASI and risk of clinical depression, and disease duration and clinical depression. Is this what the authors mean?

REVIEWER	Chiesa Fuxench, Zelma University of Pennsylvania Perelman School of Medicine, Dermatology

REVIEW RETURNED	14-Dec-2021

GENERAL COMMENTS	Thank you to the authors for their effort. I do have a few questions that hopefully can be addressed. First, in table 1. why was the cohort divided based on the intensity of the pruritus and not on the presence or absence of pruritus in the genital area? Would the results have been different if you had compared outcomes only among those with anogenital involvement according to level of itch intensity, meaning anogenital with mild/no itch vs. anogenital with mild/mode itch. I apologize if this was done and I missed it. For line 12, page 7 (introduction section): I would recommend rewording of the following sentence as this is a cross-sectional study: “..examine the influence of sociodemographic, clinical, and patient reported outcomes (PROs) of disease/treatment burden on the likelihood of developing clinically significant symptoms of depression, anxiety, and body dysmorphic concerns” Instead of “examine the influence” and “development of clinically significant symptoms” would restate as examine the association between risk factors and having symptoms of clinically significant depression/anxiety of BDD. Page 8, line 10 (section methods): I would recommend that the following sentence be further clarified: “…and those reporting itching also assessed its intensity using a Numeric Rating Scale (NRS) from 0 to 10” as 0=no itch and 10=worse itch Is this reporting worse or average itch in the past 24 hours? Page 8, line 56 (sections methods): I would be cautious about using the word prediction in this sentence since this is primarily an observation study “…of disease/treatment burden predicting the occurrence of clinically significant symptoms of...” Page 11, line 31 (results section) Can you please clarify further what "..increase of one unit in patient benefits..." means? Page 13 line 12 (discussion section) Can you please clarify further what do you mean by coping strategies in "...long-term efficacy of such coping strategies should be addressed in further research and in clinical." Are there any specific examples previously described in the literature? Line 31, page 13 "...Considering that biologics are not the first-line treatment for psoriasis, our results suggest that the combined effect of higher disease severity and longer disease duration, which qualifies the patients for biologic treatment, would be the explanatory factors for the higher likelihood of depression." Are you referring to confounding by indication in this statement?

VERSION 1 – AUTHOR RESPONSE

Reviewer #1:

This is a cross-sectional study of adults with psoriasis from a single secondary care centre in Germany that aims to assess the effects of pruritus intensity and anogenital involvement on disease and mental health outcomes, and identify the variables associated with developing depression, anxiety, and body dysmorphia. Overall, the study design and methods used are appropriate to address the research objectives and the conclusions made are justified by the results. I have summarised what I feel are revisions that will improve the understanding of the paper. Please find my detailed comments below:

1. Introduction: The introduction provides a clear rationale for the study, and a detailed examination of the existing literature. However, you state that psoriasis affects “approximately 2.5% of the German population" without citing any literature that gives this figure.
R1. Thanks for your comment. The following reference was added:
Augustin M, Reich K, Glaeske G, Schaefer I, Radtke M. Co-morbidity and age-related prevalence of psoriasis: Analysis of health insurance data in Germany. Acta Derm Venereol. 2010; 90:147–51.

2. Materials and methods:
a. Study design and participants: The authors state that “patients were excluded if they presented any condition…that would confound the interpretation of the results”. It would be beneficial for the reader if the authors stated the other conditions that patients were excluded for and how these conditions may confound the interpretation of results, especially since people with comorbidities were included in the study.
R2a. Thanks for your suggestion. We excluded patients with other inflammatory diseases or dermatologic conditions (e.g., xerosis cutis, eczema, chronic spontaneous urticaria), which could confound the interpretation of the results because of the inherent pruritus. However, psoriasis is recognised as a multisystemic disease, often associated with comorbid conditions such as cardiovascular diseases, arthritis, depression, etc. (Gottlieb & Dann, 2009). Since this was a real-world study that included patients from regular clinical practice, those with cardiovascular diseases (e.g., hypertension; n = 20), metabolic diseases (e.g., diabetes; n = 13), liver disease (e.g., hepatitis; n = 4), gastrointestinal diseases (e.g., inflammatory bowel disease; n = 1), kidney insufficiency (n = 1), pulmonary and allergic diseases (e.g., asthma; n = 13), rheumatoid/psoriatic arthritis (e.g., n = 13), and depression (n = 11) were included.
To improve clarity regarding the inclusion and exclusion of patients with comorbidities, the following corrections were made to the manuscript: “Patients were excluded if they presented any condition, including other inflammatory diseases or dermatologic conditions (e.g., xerosis cutis, eczema, chronic spontaneous urticaria), which would place them at unacceptable risk due to participation in the study or would confound the interpretation of the results because of the inherent pruritus. However, psoriasis is recognised as a multisystemic disease[20] and, thus, patients with common comorbid conditions to psoriasis, such as cardiovascular diseases, diabetes, psoriatic arthritis, and depression, were included.” (p. 6, subheading “Study design and participants”).

b. Outcome measures: The authors describe none/mild pruritus as an NRS score <=3 and moderate/severe pruritus as an NRS score>=4. It would be useful if the methods used to define cut-off points for none/mild and moderate/severe pruritus, as well as anogenital involvement were explained. Were the categories decided arbitrarily, or where they based on previous studies/literature? The authors refer to other literature for the PROs. I think it would be of value to put a brief summary in the supplementary material, e.g,. the Patient Benefit Index will not be familiar to most- I had to look it up. I think adding some text so readers can know what the PBI is would be very useful.
R2b. The cut-off points defined for mild and moderate/severe pruritus were based on previous psychometric studies on the numeric rating scales (NRS) for assessing the intensity of pruritus. To clarify our defined categories, we added the reference where these cut-off points were recommended to the manuscript: Reich A, Chatzigeorkidis E, Zeidler C, Osada N, Furue M, Takamori K, Ebata T, Augustin M, Szepietowski JC, Ständer S. Tailoring the Cut-off Values of the Visual Analogue Scale and Numeric Rating Scale in Itch Assessment. *Acta Derm Venereol* 2017; 97: 759–60.

The definition of groups regarding anogenital psoriasis was straightforward (yes/no) based on the patient report on a high-resolution grid scheme on topology of psoriasis (Augustin M, Sommer R, Kirsten N et al. Topology of psoriasis in routine care: Results from high-resolution analysis of 2009 patients. *Br J Dermatol* 2019; 181: 358–65.).
The authors also appreciate your suggestion to describe the PROs in supplementary material, which was done accordingly.

c. Statistical analyses: The authors present effect sizes for the comparison analyses, but do not describe how these numbers were calculated.
R2c. Thank you for this remark. We added the following text to the manuscript: "Partial eta squared (η_p^2), calculated from the sum of squares of the effect in relation to the sum of squares of the effect and the sum of squares of the error associated with the effect, were presented as measures of effect sizes for the comparison analyses, considering $\eta_p^2 \geq 0.01$, $\eta_p^2 \geq 0.06$, and $\eta_p^2 \geq 0.14$ as small, medium, and large effects, respectively." (p. 8, subheading "Statistical analyses").

d. Multivariable analyses: Only factors significant in univariable analyses were included in multivariable analyses. I wondered why a priori confounders were not included.
R2d. According to Lakens (Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front Psychol* 2013; 4: 863), we considered two types of confounders. First, we considered as a confounder the sociodemographic and clinical variables that were unequal distributed between the groups. Based on the comparison between the groups with none/mild vs. moderate/severe pruritus, and with vs. without anogenital involvement (one-way ANOVA or χ^2 tests presented in Table 1), we found significant differences for the presence of intertriginous psoriasis, being treated with biologic systemic therapy, PASI and %BSA. Except for %BSA that was excluded because of measurement overlap with PASI, these variables were included as covariates in the subsequent ANCOVAs and they were also tested as potential factors associated with the presence of clinically significant symptoms of depression, anxiety, and dysmorphic concerns in the logistic regression models. To improve clarity, we added the covariates included in the ANCOVAs as a footnote in table 2.
Second, we considered as confounders the factors predictive of the outcome, even in the absence of exposure. These were tested by the preliminary univariate logistic regression, and the factors significantly associated with the outcome were included in the multivariate models. The multivariate logistic regression models were not adjusted for any additional variables.

3. Results
a. Were there any differences between those completing and not completing the questionnaires?
R3a. Unfortunately, we were not able to compare the sociodemographic and/or clinical data between the participants and the patients that did not return the completed questionnaire, because we had no access to the data on a systematic basis; only paired patient-physician questionnaires were provided to the research team. The research team had also no direct access to the patient's clinical records, where such information could be retrieved because of ethical issues.

b. For PROs, would median be more appropriate than mean, i.e., are these normally distributed?
R3b. Normal distribution is, unfortunately, rarely observed in patient-reported outcomes. However, analysis of variance is known for its robustness to violations of normality in terms of Type-1 error. A recent study even showed that the robustness of the F-test to non-normality applies in different conditions, including small sample sizes and unequal group sample sizes (Blanca MJ, Alarcón R, Arnau J, Bono R, Bendayan R. Non-normal data: Is ANOVA still a valid option? *Psicothema* 2017; 29: 552–7), as in the case of our study. Considering also the disadvantages of using median and non-parametric tests, e.g., conversion of quantitative continuous data into rank-ordered data, with a consequent loss of information, we decided to use the mean and SD as measures of central tendency and to use parametric tests.

c. For tables 3-5, would have been useful to document what factors were adjusted for in multivariable analyses in a footnote. Was any assessment done to determine if there was collinearity in models?
R3c. As explained in our response R2d, the regression models presented in tables 3–5 were not adjusted for any additional variables.
The diagnosis of multicollinearity was made by examining the variance inflation factors (VIF) and considering VIF ≥ 10 as indicator of severe problems (this information was added in the methods

section). VIF were checked using the linear regression option in SPSS, as suggested by the Reviewer 2. Based on the VIF, no severe multicollinearity problems were identified in any of the multivariable regression model.

d. The conclusion about an interaction between pruritus and anogenital involvement seems to be based on univariate analyses and hasn't been explored in the multivariable model to our knowledge.
R3d. The interaction effects between intensity of pruritus and anogenital involvement on depression and perceived stigmatisation, that were presented in figure 1, were based on the plotting data provided by the ANCOVAs. These were univariate models (one dependent variable at a time), with both factors (intensity of pruritus and anogenital involvement) and covariates introduced at the same time. We opted for univariate models because the outcomes were assumed as independent (i.e., assessing different constructs, as opposed to the assessment of, for example, different dimensions of the same construct) and were not strongly correlated.
In the logistic regression models, we opted to exclude the examination of any interaction terms because of our small sample size. To examine interaction terms, we would also have to force the main effects into the model (even when the univariable effect was non-significant). This procedure to detect a 2 X 2 interaction in a regression model would require a sample size that is fourfold that to the detect a main effect of the same magnitude. However, if the Reviewer believes this is an important flaw of the manuscript, we will be happy to revise our analyses once more.

4. Discussion:
a. The authors suggest that "the accumulative burden of anogenital psoriasis and moderate/severe pruritus may trigger coping strategies that have a protective effect on the patients' mental health", but do not posit examples of what these coping strategies may be. It would be useful for the reader if you gave further examples of the coping strategies and how you think they should be addressed in further research.
R4a. Thank you for your comment. We are referring to avoidance coping mechanisms, including both cognitive/emotional strategies to reduce thoughts or feelings (e.g., mental disengagement/distraction; denial), and behavioural attempts to physically remove one's self from an aversive situation (e.g., social withdrawal). We have added these into the respective paragraph in the manuscript. Our explanatory hypothesis is that the avoidance of social situations would protect the patient from negative experiences of stigmatisation and, therefore, from feelings of anxiety and depression triggered by social interactions or social exposure. However, these coping strategies might have a protecting effect in the short-term, but at the long-term, they might have the exact opposite effect, leading the patient to isolation, severe depression and suicidal ideation. This hypothesis should be addressed in cohort studies, to evaluate the short- and long-term of avoidance coping on the patients' mental health outcomes.

b. As the authors mention, the sample size for this study is small, resulting in wide confidence intervals in multivariable analyses. For example, in the multivariable analysis the effect estimate reported for biological treatment was 80.30 [1.30–4950.81]. This study is likely underpowered to assess all of its objectives. Focusing on the OR of 80 in the discussion without mentioning that the CI ranged from 1. To almost 5000 is problematic. You need the CIs here and would suggest editing the language.
R4b. We agree with the reviewer and, accordingly, we reformulated the "quantitative" language used to present and discuss the results (i.e., the use of OR to quantify the increase/decrease of outcomes). Instead, we adopted now a more qualitative language, to interpret the results (e.g., positive/negative association between the variable). In addition, the following limitation was included in the discussion section (p. 13): "First, the small sample size diminishes the statistical power of analyses and resulted in wide confidence intervals, particularly in multivariable analyses. Consequently, conclusions based on effect estimates are unreliable and the results must be interpreted only qualitatively, in terms of positive/negative associations between the variables."

c. Similarly, the statement that the multivariate analysis showed that clinical anxiety is 89 % less likely as the PBI increases by one unit is based on an OR and CI of 0.11 [0.02–0.66]. The statement is strong given the width of the CI and this should be reframed. There are other similar examples in the discussion.
R4c. Thanks again for your comment. Accordingly, we reformulated our statement, which is now focused on the association, instead of the OR and its associated large CI: "The multivariable analysis

showed that the presence of clinical anxiety was less likely in patients with more patient-defined treatment benefits (Table 4).” Similar reformulations were made all over the manuscript.

d. There is a lot of focus on statistical significance and p values. I would be more inclined to focus on the effect estimates and CIs, https://www.nature.com/articles/d41586-019-00857-9.
R4d. Thanks for your comment and for sharing this thought-provoking article. For the presentation of comparative analyses (e.g., ANOVA), we opted to present the F-tests and the associated p-values, as these are the most commonly used; for the regression analyses, Wald statistics and its significance, but also OR and confidence intervals were presented. For the interpretation of results, and according to the reviewer’s previous comments, we adopted now a more qualitative language, as the interpretation of OR and their wide CI could also be misleading.

e. It would be good to discuss generalisability, as this study was conducted in a single secondary care centre and findings may not be applicable to people seen in primary care or different geographical regions.
R4e. According to your comment, we discussed further the generalisability of our results, which is a clear limitation of the study. The following sentence was added to the discussion of results (p. 13): “Second, the convenience sampling method in a single dermatology outpatient clinic based in an university hospital limits the generalisability of results, for example, to other geographic areas or to patients being cared by office-based dermatologists.”.

Reviewer #2:

Thank you for the opportunity to review ‘Psychological (co)morbidity in patients with psoriasis: The impact of pruritus and anogenital involvement on symptoms of depression and anxiety and on body dysmorphic concerns – a cross-sectional observational study’. Overall, I found the paper well written and comprehensive in its analysis of clinical and psychosocial variables. I do have some questions relating to the interpretation of a few of the ORs, as well as the conclusion that can be drawn from the very wide 95% confidence interval.

Introduction
First paragraph, 1st sentence, line 22 ‘…taking into account the patient’ needs…’
Change to ‘…taking into account the patient’s needs…’
R. Thank you for this correction, which was incorporated in the manuscript.

Materials and Methods
Study design and participants:
‘Patients were excluded if they presented any condition, including other inflammatory diseases or dermatologic conditions, which would place them at unacceptable risk due to participation in the study or would confound the interpretation of the results’.
Are the authors only referring to physical conditions here? There were no exclusion criteria based on psychological conditions?
R. As exclusion criteria, only physical conditions were considered (please see our response R2a. to reviewer 1). Eleven patients with previously diagnosed depression were also included, because depression is a common comorbidity of psoriasis and because the presence of clinical depression was an important study outcome and, thus, the exclusion of those patients would mitigate the results. Unfortunately, other comorbid psychological comorbidities (e.g., anxiety, suicidal ideation) could not be properly assessed in the course of this study.

Outcome measures: ‘A set of questionnaires were completed by the physician and by the patient.’ Is the physician the usual physician who treats the patient or a physician engaged for the specific purposes of the research?
R. The clinical questionnaire was completed by a physician assigned for this specific study, but that also belongs to the clinical team caring for the patient. This point was clarified in the manuscript. (p. 6f., subheading “Outcome measures”).

Statistical analyses
SPSS is no longer known as the Statistical Package for the Social Sciences . Please change to The statistical analyses were conducted using IBM SPSS Statistics (SPSS, version 23.0, IBM Corp.,

Armonk, NY).
R. Thank you for this correction, which was incorporated in the manuscript.

A point on definitions. 'Multivariate' typically means that there are multiple outcomes in the one model. Multivariable is the appropriate term to use when there are multiple predictors.
See https://academic.oup.com/ntr/article/23/8/1446/5812038
I am also not a fan of using the notation Mean +/- SD, as the SD relates to the average distance of each observation from the mean (the dispersion of the data) and is not intended to convey the plausible range of values for the mean. I would prefer to see means and SDs reported as Mean (SD).
R. Thank you for your remarks. The term "multivariate" was corrected to "multivariable" throughout the manuscript. For consistency of terms, univariate was also corrected to univariable, as we wanted to refer to one single predictor and not to one single outcome. Also following your recommendation, M ± SD was replaced by M (SD) in the text and tables.

Results
Table 1: There are 2 major groups and within each of these there are also 2 groups, so 4 groups in total. Are the p values based on comparisons between the 4 groups, or the two larger groups? If based on the 4 groups, then some of these would have been post-hoc tests (continuous data) or assessed with ASRs (adjusted standardised residuals, chi square). For example, the p value for the higher mean %BSA in the patients with moderate/severe pruritus and with anogenital involvement is <0.001. So if based on a post hoc test, was the p value <.001 for all comparisons? The SD for this group is particularly large, indicating considerable variation in the response of this group to this outcome variable.
R. Thanks for your comment. We had performed the pairwise post-hoc comparisons to identify which groups were significantly different, as we described in the methods section. However, we did not present these results because the aim of these analyses was only to identify potential confounders. But to improve clarity in our manuscript, we now added the following text, presenting the results for the significant post-hoc comparisons (p. 9, subheading "Sample characteristics"): "Comparative analyses revealed a lower frequency of intertriginous psoriasis among patients with none/mild pruritus and no anogenital psoriasis, compared to those with none/mild pruritus and anogenital involvement ($\chi2 = 9.98$, $p = 0.009$) and to those with moderate/severe pruritus and anogenital involvement ($\chi2 = 10.71$, $p = 0.004$). Patients with none/mild pruritus and no anogenital psoriasis were more often treated with biologics, compared to patients with moderate/severe pruritus and no anogenital psoriasis ($\chi2 = 5.51$, $p = 0.020$) or those with moderate/severe pruritus and anogenital involvement ($\chi2 = 7.70$, $p = 0.007$). Moreover, patients with moderate/severe pruritus and with anogenital involvement presented higher PASI and larger %BSA than those with none/mild pruritus and no anogenital involvement (mean difference [MD] = 5.77, standard error [SE] = 1.22, $p < 0.001$ for PASI and MD = 14.52, SE = 3.08, $p < 0.001$ for %BSA) and those with none/mild pruritus and anogenital psoriasis (MD = 5.63, SE = 1.68, $p = 0.007$ for PASI and MD = 14.98, SE = 4.45, $p = 0.007$ for %BSA).".

Just want to make sure I understand Table 2. For example, for the DLQI, a model was built with pruritis (none mild; moderate severe) as a predictor and anogenital involvement (None or anogenital psoriasis) also as a predictor, and intertriginous psoriasis, biologic treatment, and PASI as covariates. There was also an interaction term of pruritis * anogenital in the model. There is a main effect for pruritis with a medium effect size, no main effect for anogenital involvement and no interaction.
The authors have not addressed whether the assumptions for ANOVA were met. ANOVA is known to be robust under violations of normality, but less so when group sizes are unequal (as the case in this study).
R. That is exactly how Table 2 should be interpreted. We are glad that the results were clear. Regarding the assumptions of ANOVA, there is a recent study showing that the robustness of the F-test to non-normality applies in different conditions, including small sample sizes and unequal group sample sizes (Blanca MJ, Alarcón R, Arnau J, Bono R, Bendayan R. Non-normal data: Is ANOVA still a valid option? *Psicothema* 2017; 29: 552–7.), as in the case of our study. For this reason, we opted by ANOVA instead of the equivalent non-parametric tests.

I apologise if I am misunderstanding Table 3. The way the predictors are reported, eg Pruritus (none/mild vs. moderate/severe) or Biological treatment (no vs. yes), the condition appearing first is the reference to which the other condition is compared, eg none/mild is the reference to which moderate/severe is compared, and the OR and 95% reflected the increase (or decrease) in odds

compared to the reference.
So for Pruritus (none/mild vs. moderate/severe) the OR for having clinically significant depression is 5.83 (1.74 – 19.53), therefore, as the authors report, moderate/severe pruritis is associated with an increased risk of clinical depression (5.83 times the risk or 483% increase). All good.
R. Thanks for your comment. We are happy that the presentation of results from logistic regression analyses were clear but even so, to avoid misunderstanding from readers, we added the numeric values of each dichotomous predictor in tables 3–5, for example, biological treatment: 0 = no [reference] vs. 1 = yes), with the footnote that the reference category is always the first value.

For biological treatment (assuming that no is the reference and the statistics relate to the 'Yes' condition, the OR is 0.30 (0.10 – 0.91). This indicates that those undergoing biological treatment have a reduced risk of clinical depression. Yet the authors state that this is associated with an increased risk. Similar for disease duration – I would interpret the result as each increment (yearly?) in duration, the risk of clinical depression decreases by 5%.
R. Thank you very much for this comment! You are absolutely right in your interpretation of results and we apologize for our misinterpretation. Now, the correct text reads as follows (p. 10, subheading "Patient-reported outcomes of disease and treatment burden"): "Univariable analyses (Table 3) revealed that patients with moderate/severe pruritus, shorter disease duration, not prescribed with biologic treatment, higher PASI, more skin-generic and pruritus-specific QoL impairments, less patient benefits, sleeping problems, more stigmatisation experiences, and greater sexual dysfunction were more likely to present clinically significant symptoms of depression (PHQ-2 ≥ 3)".

The multivariable model seems a bit odd to me. Nagelkerke is relatively high, but there are only 2 statistically significant variables in the model, and one of them (biological treatment) has an incredibly wide 95% CI. Very wide CIs can be due to sparse data, but according to Table 1, there were 66 patients undergoing biological treatment, and therefore 41 not having this treatment, so sparse data doesn't seem to be the problem here. It also seems unusual that biological treatment in univariate analysis reduces the risk of clinical depression (OR 0.30 (0.10 – 0.91), so reduced it by 70%, but in multivariable analysis, controlling for the other variables, it increases the risk by a massive amount. A 95% CI this wide also renders the result unreliable in that it is hard to conclude anything about the effect of this treatment, other than it causes an increase. This should be mentioned in the Limitations. Did the authors check for multicollinearity? This can be checked using the linear regression option in SPSS, even though the outcome in logistic regression is dichotomous it doesn't matter when doing this check as the multicollinearity check is only concerned with the predictors, not the dependent variable.
R. We re-run all logistic regression models and the results remained unchanged. We also appreciate your suggestion for checking for multicollinearity among the independent variables in the multivariable models. The variance inflation factors (VIF) are now presented in tables 3–5 and, accordingly, we can say that our models were not limited by severe multicollinearity problems. Since multicollinearity or sparse data seems not explain the wide CI for biologic treatment, we assumed that the small sample size and consequent low power of our analyses could be the explanation. Therefore, and following the recommendations of all three reviewers, we reformulated the language for interpretation of results and mentioned the wide CI as an important study limitation (p. 13).

Discussion
Page 13, line 36 'The univariate associations between PASI and disease duration as well as PASI and PHQ ≥ 3 corroborate this hypothesis.' However, I don't see any univariate associations between PASI and disease duration, only between PASI and risk of clinical depression, and disease duration and clinical depression. Is this what the authors mean?
R. Thank you for this note. We meant the association between disease duration and depression. However, the direction of this association was misinterpreted and makes no sense. Therefore, we corrected our statement to "The univariable associations between higher PASI clinical depression, as well as between more DLQI impairments and clinical depression corroborate this hypothesis.".

Reviewer #3:

Thank you to the authors for their effort. I do have a few questions that hopefully can be addressed.

First, in table 1. why was the cohort divided based on the intensity of the pruritus and not on the presence or absence of pruritus in the genital area? Would the results have been different if you had compared outcomes only among those with anogenital involvement according to level of itch intensity, meaning anogenital with mild/no itch vs. anogenital with mild/mode itch. I apologize if this was done and I missed it.
R. Thank you very much for this comment. Unfortunately, the intensity of pruritus and the involvement of the anogenital area were assessed with independent instruments that could not be combined to create the groups as you suggested. More specifically, when the patient marked a square in the anogenital area, we assumed he/she had anogenital involvement. But often other body areas were also marked, as the prevalence of isolated genital psoriasis has been estimated between 2–5 % of all patients with psoriasis but generalised plaque-type or intertriginous psoriasis also affects the genital area in 29–40 % of psoriasis cases (Meeuwis KA, de Hullu JA, Massuger LF, van de Kerkhof PC, van Rossum MM. Genital psoriasis: A systematic literature review on this hidden skin disease. *Acta Derm Venereol* 2011; 91: 5–11.). Consequently, the patients' assessment of intensity of pruritus could not be assumed as anogenital pruritus, as they might be referring to pruritus in other body areas. However, we find your suggestion of enormous interest and we hope we can address this issue in further studies specifically designed to evaluate anogenital pruritus (and anogenital scratching) and its psychosocial impact in terms of feelings of embarrassment and stigmatization, as well as in the patient's general and sexual quality of life.

For line 12, page 7 (introduction section): I would recommend rewording of the following sentence as this is a cross-sectional study:
"..examine the influence of sociodemographic, clinical, and patient reported outcomes (PROs) of disease/treatment burden on the likelihood of developing clinically significant symptoms of depression, anxiety, and body dysmorphic concerns"
Instead of "examine the influence" and "development of clinically significant symptoms" would restate as examine the association between risk factors and having symptoms of clinically significant depression/anxiety of BDD.
R. Thank you for your comment. The wording suggesting directional associations, that are only possible to address in longitudinal studies, was corrected in the introduction and also throughout the manuscript.

Page 8, line 10 (section methods): I would recommend that the following sentence be further clarified:
"…and those reporting itching also assessed its intensity using a Numeric Rating Scale (NRS) from 0 to 10" as 0=no itch and 10=worse itch
Is this reporting worse or average itch in the past 24 hours?
R. The intensity of pruritus was assessed as average itching during the last 24 hours, using the question "How strong was the average itching during the last 24 h?". To improve clarity, this question was added to the manuscript.

Page 8, line 56 (sections methods): I would be cautious about using the word prediction in this sentence since this is primarily an observation study
"…of disease/treatment burden predicting the occurrence of clinically significant symptoms of..."
R. According to the reviewer's previous comment, the expression "predicting the occurrence of clinically significant symptoms" was replaced by "associated with the presence of clinically significant symptoms".

Page 11, line 31 (results section) Can you please clarify further what "..increase of one unit in patient benefits..." means?
R. Following the recommendations of the reviewers, we reframed the language used in the presentation and discussion of results, focusing on factors associated with an increase/decrease of the outcome, and less on the odd ratios because of the large confidence intervals. Consequently, this particular sentence was replaced by "patients prescribed with biologic treatment and reporting less patient benefits were more likely to present clinically significant symptoms of depression".

Page 13 line 12 (discussion section) Can you please clarify further what do you mean by coping strategies in "...long-term efficacy of such coping strategies should be addressed in further research and in clinical." Are there any specific examples previously described in the literature?

R. Examples of specific coping strategies were provided in the manuscript. For details, please, see our response R4a. to reviewer 1, regarding this topic.

Line 31, page 13 "...Considering that biologics are not the first-line treatment for psoriasis, our results suggest that the combined effect of higher disease severity and longer disease duration, which qualifies the patients for biologic treatment, would be the explanatory factors for the higher likelihood of depression." Are you referring to confounding by indication in this statement?
R. Yes, we are referring to a potential confounding by indication effect. More specifically, it could be that not the prescription of biologics itself would increase the likelihood of depression, but rather the patient and disease characteristics that qualifies the patient for biologic treatment, including longer disease course, higher severity, lower quality of life, etc. This confounding effect might actually be the reason for the contradictory associations in uni- and multivariable analyses, i.e., the association between biologic treatment and lower rates of depression in univariable analysis, and association between biologic treatment and higher rates of depression in multivariable analysis.

VERSION 2 – REVIEW

REVIEWER	Fethney, Judith University of Sydney, School of Nursing
REVIEW RETURNED	10-Feb-2022

GENERAL COMMENTS	Thankyou for the opportunity to review this paper. It is a much better read after the authors have addressed reviewers' concerns. I just have a few comments The authors should specify in the Statistical analyses section that for all tests, p values <0.05 were considered statistically significant. Personally, I would have selected predictors from the univariable models with p values of 0.10 or 0.20 for inclusion in a multivariable model, rather than 0.05. This is an interesting article about variable selection https://onlinelibrary.wiley.com/doi/pdf/10.1111/tri.12895 Underneath Table 2, the authors have reported the range of scores for the scales in that table. This would also be helpful for the PASI if possible in Table 1. I assume for the %BSA the minimum and maximum are 0 -100%? Would still be helpful to include this. Results, P9: It would be more helpful to report the 95% CIs around the mean differences (MD) than the SE. I am still finding the multivariable coefficients for Biological treatment and depression in Table 3 rather odd. It is reversed from the univariable and the CI is now incredibly wide. This requires some exploration. The VIFs for DLQI and ItchyQol, while < 10 specified by the authors, could still be problematic, as some literature mentions that VIFs > 2.5 can indicate a collinearity issue.

	These links might be of interest. https://stats.stackexchange.com/questions/116804/coefficient-changes-sign-when-adding-a-variable-in-logistic-regression https://statmodeling.stat.columbia.edu/2013/05/26/how-to-understand-coefficients-that-reverse-sign-when-you-start-controlling-for-things/ https://www.youtube.com/watch?v=Ysg6-puIR1s

REVIEWER	Chiesa Fuxench, Zelma University of Pennsylvania Perelman School of Medicine, Dermatology
REVIEW RETURNED	25-Feb-2022

GENERAL COMMENTS	Thank you for submitting a revised version and for the opportunity to review his work. Below, please find a few minor suggestions with the goal of making this a stronger manuscript. Section: Abstract Page: 4, Line 42, consider changing "but" for "and"line Page 5, Line 25, consider altering the order for the last 2 bullets as follows. Stress limitations of study design first followed by limitations of small sample size. Section: Materials and Methods Page 7, line 24, consider omitting the word "observational" as this is implied in cross-sectional studies Page 8, line 16, consider adding the word "the" in between "in" and "supplementary material"

VERSION 2 – AUTHOR RESPONSE

Reviewer #2:

Thank you for the opportunity to review this paper. It is a much better read after the authors have addressed reviewers' concerns. I just have a few comments

1. The authors should specify in the Statistical analyses section that for all tests, p values <0.05 were considered statistically significant. Personally, I would have selected predictors from the univariable models with p values of 0.10 or 0.20 for inclusion in a multivariable model, rather than 0.05. This is an interesting article about variable selection: https://onlinelibrary.wiley.com/doi/pdf/10.1111/tri.12895

R. Thank you for your comment. As suggested, the following sentence was added to the "Statistical analyses" section: "For all statistical tests, p-values < 0.05 were considered statistically significant".

2. Underneath Table 2, the authors have reported the range of scores for the scales in that table. This would also be helpful for the PASI if possible in Table 1. I assume for the %BSA the minimum and maximum are 0 -100%? Would still be helpful to include this.

R. Thank you for this suggestion. The following footnote was added to Table 1: "PASI: Psoriasis Area and Severity Index (range 0−72, with higher values indicating greater disease severity); %BSA: percentage of Body Surface Area affected by psoriasis (range 0%-100%)".

3. Results, P9: It would be more helpful to report the 95% CIs around the mean differences (MD) than the SE.

R. Following the Reviewer's suggestion, the standard errors (SE) were replaced by the 95% confidence intervals (CI).

4. I am still finding the multivariable coefficients for Biological treatment and depression in Table 3 rather odd. It is reversed from the univariable and the CI is now incredibly wide. This requires some exploration.

R. Thank you for insisting in this point. We totally agree that the extremely wide confidence intervals for biologic treatment are a cause for concern and we followed, once more, the Reviewer's suggestion of further exploration. Beyond the small sample size, which was already acknowledge in the study limitations, other reasons for wide 95%CI include the variability in the sample/ inconsistent data, and confounding/ multicollinearity issues. Although the independent variable "Biologic treatment" did not present high multicollinearity with the other predictors (VIF = 1.85), inter-relationships between variables, even if not approaching high collinearity, can have a substantial impact on regression model results. Because biologic treatment is not the first-line treatment for psoriasis, its prescription is dependent on a combination of other variables, namely disease severity (PASI), QoL impairments (DLQI) and also disease duration. And the inter-relations between these variables are far from being straightforward: for instance, a patient with newly diagnosed moderate to severe psoriasis could be first prescribed with conventional systemics, while a patient with long disease duration who was now prescribed with biologics, could present already improvements in disease severity (i.e., lower PASI) as a result of the biologic treatment. Therefore, it is not surprising that biologic treatment do not present a high correlation with PASI (r = -0.40), disease duration (r = 0.29) of DLQI (r = -0.39), but the interaction between all these variables might have contributed for the wide 95%CI.

To confirm this explanatory hypothesis, we conducted the logistic regression predicting depression (PHQ-2 ≥ 3), excluding PASI, disease duration and DLQI as independent variables, at a time:

- When disease duration was excluded, biologic treatment remained significantly associated with greater likelihood of depression, with an even wider CI (B = 4.55, SE = 2.10, Wald = 4.68, p = 0.03, OR [95 % CI] = 94.35 [1.54 – 5800.38]);
- When the DLQI was excluded, biologic treatment remained significantly associated with greater likelihood of depression, with a slightly narrower CI (B = 4.09, SE = 1.93, Wald = 4.50, p = 0.03, OR [95 % CI] = 59.68 [1.37 – 2609.08]);
- When PASI was excluded, the association between biologic treatment and depression was no longer statistically significant and the CI was significantly narrower: B = 1.79, SE = 1.25, Wald = 2.05, p = 0.15, OR [95 % CI] = 5.98 [0.52 – 69.01].

For all these 3 experimental models, the PBI remained significantly associated with depression.

Considering these results, we can assume that biologic treatment is instable as predictor of depression, with a probable reason of confounding by indication. Therefore, we revised the interpretation of the results regarding biologic treatment in the manuscript.

5. The VIFs for DLQI and ItchyQol, while < 10 specified by the authors, could still be problematic, as some literature mentions that VIFs > 2.5 can indicate a collinearity issue. These links might be of interest. https://stats.stackexchange.com/questions/116804/coefficient-changes-sign-when-adding-a-variable-in-logistic-regression https://statmodeling.stat.columbia.edu/2013/05/26/how-to-understand-coefficients-that-reverse-sign-when-you-start-controlling-for-things/ https://www.youtube.com/watch?v=Ysg6-puIR1s

R. Thanks for this consideration. In fact, there is no universal agreement for a cut-off point where a VIF value indicates multicollinearity. Most literature suggests that a VIF > 5 is cause for concern and VIF > 10 indicates a serious multicollinearity problem (e.g., Menard, 2001), but each case should be inspected individually. The problem of multicollinearity is that it can cause unstable estimates and inaccurate variances which affects confidence intervals and undermines the statistical significance of an independent variable. To further explore the consequences of multicollinearity for the DLQI and ItchyQoL, we repeated the regression analyses using the approach to drop one of the variables to reduce multicollinearity. The models were quite stable

and the main result (i.e., the significant association between less patient benefits and greater likelihood of depression and anxiety) remained unchanged. Therefore, we believe that the VIFs for the DLQI and ItchyQoL do not represent a serious problem in this case.

Reviewer #3:

Dear authors, Thank you for submitting a revised version and for the opportunity to review his work. Below, please find a few minor suggestions with the goal of making this a stronger manuscript.

1. Section: Abstract. Page: 4, Line 42, consider changing "but" for "and"line

R. Thank you for your comment. The abstract was corrected accordingly.

2. Page 5, Line 25, consider altering the order for the last 2 bullets as follows. Stress limitations of study design first followed by limitations of small sample size.

R. Thank you for your comment. The order of the two last bullet points was altered accordingly. To keep consistency, we also changed the order of this two limitations in the Discussion section.

3. Section: Materials and Methods. Page 7, line 24, consider omitting the word "observational" as this is implied in cross-sectional studies

R. Thank you for your comment. The word “observational” was deleted in the “Study design and participants” section, as well as in the title and abstract.

4. Page 8, line 16, consider adding the word "the" in between "in" and "supplementary material"

R. Thank you for this correction, which was incorporated in the manuscript.

VERSION 3 – REVIEW

REVIEWER	Fethney, Judith University of Sydney, School of Nursing
REVIEW RETURNED	25-Jul-2022

GENERAL COMMENTS	Manuscript is acceptable for publication