Statistical Analysis Plan: Developing diagnostic criteria for psoriasis in children (DIPSOC): A case control diagnostic accuracy study

V0.11 February 2019

1. A description of the study data sources, linkage methods, and study design including intended study population, inclusion and exclusion criteria and study period with discussion of strengths and weaknesses.

The study design is a case-control diagnostic accuracy study in paediatric dermatology clinics. Children under the age of 18 years with a confirmed dermatologist's diagnosis of psoriasis (cases) or another scaly inflammatory rash (controls) will be included in the study. All participants will be required to have active (visible) disease at the time of assessment. Children with pustular psoriasis and erythrodermic psoriasis will be excluded from the study. The study period for each participant will be a single research visit lasting approximately 30-45 minutes, during which a diagnostic criteria assessment will take place. Data will be collected at the research visit and extracted from the medical record.

The study will include an assessment of interobserver variability and variability in the reference standard for psoriasis. There will also be a nested study investigating the predictive ability of the diagnostic criteria in participants who currently have indeterminate psoriasis. The total study duration is estimated to be 18 months.

The diagnostic criteria are intended to be used in clinical practice, with particular utility in primary care and paediatric, and research. The case-control study is the first diagnostic test accuracy study of these criteria, and therefore specific diagnostic accuracy testing and validation will be needed in the populations and settings that the cirteria are intended to be used.

A case-control design for testing diagnostic accuracy is known to over-estimate the diagnostic ability of the diagnostic tool. However, a case-control study is an efficient design to investigate the diagnostic accuracy in a disease with a moderately low clinic prevalence (4-7%) (Burden-Teh, Thomas et al. 2016). The decision to undertake this study in a secondary population will introduce spectrum bias (ie children are more likely to have severe and persistent skin disease), but this is a feasible population for both recruitment and obtaining a reference standard (a dermatologist's diagnosis). The index test (diagnostic criteria) will be independent of the reference standard by blinding the assessors to the case or control status of the participant. All participants will receive the index test and the reference standard ensuring complete verification.

2. Formal definitions of exposure including transformations to determine duration and quantity of exposure.

The exposure variables are the 16 diagnostic features identified as important for the diagnosis of plaque psoriasis in children by an eDelphi consensus study with the International Psoriasis Council. Two diagnostic features that almost reached consensus will also be included in the study.

Major criteria (the presence of 1 or more of the following features allow a diagnosis of psoriasis to be made)

- scaly erythematous plaques on the extensor surfaces of the elbows and knees

- scaly erythematous plaques on the trunk triggered by a sore throat or other infection
- raindrop plaques typical of guttate disease on the trunk or limbs

Minor criteria (the presence of 3 or more of the following features, in the absence of a major criterion, allow a diagnosis of psoriasis to be made)

- scale and erythema in the scalp involving the hairline
- retro-auricular erythema (including behind the earlobes)
- scaly erythema inside the external auditory meatus
- persistent well-demarcated erythematous scaly rash anywhere on the body
- fine scaly patches involving the upper thighs and buttocks
- well-demarcated erythematous rash in the napkin area involving the crural folds
- persistent erythema in the umbilicus
- nail pitting
- onycholysis of the nail(s)
- subungal hyperkeratosis of the nail(s)
- positive family history of psoriasis
- koebner phenomenon
- fusiform swelling of a toe or a finger suggestive of dactylitis

Criteria almost reaching consensus

- Persistent well-demarcated facial rash with fine or absent scale
- Natal cleft erythema and/or skin splitting

Each exposure will be dichotomised to 'present' or 'absent' at the time of the diagnostic criteria assessment.

All study investigators undertaking the diagnostic criteria assessment will undergo standardised face-to-face or teleconference training using Powerpoint slides. The training covers how to approach assessment for each criterion, and provides plain language explanations and clinical photographs. All training will be given by Dr Esther Burden-Teh, DIPSOC study coordinator and Dermatology registrar with an interest in paediatric psoriasis. The training was developed with Dr Murphy, international expert in adult and paediatric psoriasis, and other trained dermatologists. The training was piloted for ease of use and understanding with non-dermatology trained investigators. Throughout the training, investigators will be given the opportunity to ask questions. Investigators will also be provided with a printed training manual to use as a reference aid during the diagnostic criteria assessment.

At the end of the training session the investigators understanding of the training will be assessed using clinical photographs in a short test, during which Dr Burden-Teh will ask questions about the decision making process for a number of the criterion. If the test is passed (90%) a certificate of training will be provided.

3. Definition of follow-up and censoring if applicable.

This study does not involve any follow-up

4. **Formal definitions of any outcomes, for example 'fatal myocardial infarction' that might be defined as 'death within 30 days of a myocardial infarction'. Outcome variables based on historical data may involve complex transformations to approximate clinical variables not explicitly measured in the dataset used. These transformations should be discriminated from those made to improve the fit of a statistical model. In either case the rationale should be given. In the latter case this will include which tests of fit will be used and under what conditions a transformation will be used.**

The outcome variables are the disease status of the participants. For cases the outcome variable is a dermatologist's confirmed diagnosis of plaque psoriasis (where plaques are the main feature) and for controls the outcome variable is a dermatologist's confirmed diagnosis of a disease presenting with a scaly inflammatory rash (excluding psoriasis or indeterminate psoriasis). Examples of skin diseases which may fulfil the eligibility criteria for a control participant are eczema, pityriasis rosea, pityriasis rubra pilaris, tinea infection, viral exanthem, but this is not an exhaustive list. The diagnosis for both is a clinical diagnosis, which may be supported by a skin histology. The outcome variables will be extracted from the medical record from the most recent consultation.

5. **Formal definitions for other variables – e.g. thresholds for abnormal levels of blood parameters. When values of variables for a subject vary with time, care should be given to explaining how the values will be determined at each time point and recorded in the dataset for use in a statistical model.**

Data on the variables below will be collected at the time of the research visit or extracted from the medical record. There are no time variable variables, as all data are collected at one time point and there are no follow-up visits.

- Age at the time of assessment (continuous) – years

Calculated as the number of days between the date of birth and date of the research visit. Date of birth collected at the research visit.

- Age at the time of first symptoms (continuous) – years

Extracted as the age in years of first symptoms (eg rash) from the GP referral letter or the first dermatology consultatation. If under 1 year of age then recorded as 0 years. If not recorded in the medical record then there is an option for 'not documented' (missing data).

- Disease duration calculated from date of first diagnosis (continuous) – years

Extracted as the date first received current skin diagnosis from a dermatologist. This will often be the first consultation date as a new patient. If clearly stated in the medical history that a dermatologist's diagnosis was received at another centre, then this date sill be used.

- Sex (categorical) – male/female/other/prefer not to say

Collected at the research visit - participants/parents asked which category the child/young person identifies with.

- Ethnicity (categorical) – groups as per UK census

Collected at the research visit - participants/parents asked which category the child/young person identifies with.

- Socioeconomic group (ordered categorical) – groups as per UK census

Collected at the research visit – participants/parents asked what occupation the adults in the household hold. These data are the categorised into the five groups for socioeconomic status as per the institute for national statistics.

- New or follow-up patient (categorical) – binary

Extracted from the most recent consultation. New patient is defined as new presentation of skin disease and not currently undergoing dermatology follow-up (ie new referral has been made from primary care). Follow-up patient is defined as currently undergoing dermatology follow-up.

- Disease severity (ordered categorical) – mild, moderate, severe, not documented

Extracted from the most recent consultation where this is documented.

A new variable of categorised disease severity will be created. If severity not documented then the free text description of severity and/or PASI score will be used, if possible, to categorise severity.

- Treatment (ordered categorical) – topical, phototherapy, systemic, biologic

Extracted from the most recent consultation ie medications the participant is on at the time of assessment.

- Quality of life – CDLQI and CHU9D (continuous)

Collected at the research visit. Quality of life impairment will then be categorised as per guidance provided by the authors of Children's Dermatology Life Quality Index (CDLQI) or Child Health Utility Index 9D (CHU9D). The CDLQI is a dermatology specific quality of life instrument and the CHU9D is a generic preference-based health related quality of life instrument.

- Assessor type (ordered categorical) – groups as per CRF

Collected at the research visit. Investigators self-report their assessor type.

- Blinding of assessor(categorical) – binary

Collected at the research visit. Investigators document if they were unblinded before or during the diagnostic criteria assessment.

Variation of the diagnostic accuracy by the following variables will be explored. This analysis will compare the performance of the diagnostic criteria across different clinical contexts. No minumim data in each strata will be required, but the results will be presented with confidence intervals.

- Age at the time of assessment – 1) 9 years and younger; 2) 10 years and older

Defined as per WHO guidance for child vs adolescent (ie the onset of puberty).

- Sex – 1) male; 2) female; 3) other or PNTS
- Assessor type – 1) dermatology trained (derm cons, paed cons, derm registrar, derm nurse); 2) dermatology untrained (other doctor, non-derm nurse, other investigator)
- New of follow-up – 1) new; 2) follow-up

A sensitivity analysis will be performed using the following variables. This analysis will help interpretation of whether the performance of the diagnostic criteria differs when a part of the study population is removed.

- Ethncitity – remove the category 'white'
- Blinding of assessor – remove unblinded 'no'
- Disease severity - remove all categories except mild

6. The effect measures and statistical methods used to address each primary and secondary objective.

Primary objective

To determine the diagnostic accuracy (sensivity and specificity) of the consensus agreed diagnostic criteria (either 1 major criterion and/or 3 minor criteria) for plaque psoriasis in children/young people and develop the best predictive criteria using multivariate analysis.

Diagnostic accuracy of the consensus agreed criteria. Sensitivity will be calculated as the proportion of people with psoriasis who were identified by the consensus agreed diagnostic criteria with psoriasis. Specificity will be calculated as the proportion of people without psoriasis who were excluded from a diagnosis of psoriasis by the consensus diagnostic criteria. Likelihood ratios will also be calculated.

Developing the best predictive criteria. The best predictive criteria are defined as the list of major and/or minor criteria (including determing the minimum coefficient threshold for minor criteria), that best predict psoriasis in the multivariate model. In interpreting best, the criteria will need to reach a threshold of 80% sensitivity and 80% specificity based on discussion with an expert group.

The sensitivity and specificity of each of the major criteria will be calculated. Those criteria reaching the sensitivity and specificity threshold (80% sensitivity and 80% specificity), and therefore the individual criteria will support a diagnosis of psoriasis, will be kept as major criteria. Those criteria which don't meet this threshold will be included in the analysis for minor criteria

The sensitivity and specificity of each of the minor criteria (and the 2 criteria that almost reached consensus) will be calculated. All minor criteria will be entered in to a backward regression model to develop the best predictive minor criteria. Likelihood ratios will be presented. The ROC curve will be used to determine a coefficient threshold, above which the score will support a diagnosis of psoriasis. The minimum sensitivity and specificity threshold for the minor criteria model is equal to the threshold specified for the major criteria (80% sensitivity and 80% specificity).

The best predictive criteria will be applied to the study population and the sensitivity and specificity (diagnostic accuracy) determined. A receiver operator characteristic (ROC) curve will be drawn and the area under the curve will be calculated.

Internal validation will be undertaken using bootstrapping. This approach will quantify the over estimation of the predictive performance of the modelled criteria (Pavlou, Ambler et al. 2015).

Exploratory analysis

Latent class analysis (LCA) may be used as an exploratory technique to create diagnostic groups (latent classes) using the diagnostic criteria (exposure variables).

Different thresholds may be set to explore the effect of these on the best predictive criteria. These thresholds will mirror those discussed with the expert group for use in future validation studies in primary care/community (90% sensitivity and 80% specificity), research studies (80% sensitivity and 95% specificity) and for documentation in clinical records (80% sensitivity and 95% specificity).

Secondary objective

1. To compare the diagnostic performance of the consensus agreed diagnostic criteria and the best predictive criteria for plaque psoriasis in children.

The consensus agreed diagnostic criteria and the best predictive criteria will be compared using receiver operator characteristic (ROC) curves.

2. To evaluate the inter-observer variability in the assessment of the consensus agreed diagnostic criteria.

The inter-observer variability will be evaluated using the Kappa statistic (95% CI)

3. To assess the variability in the reference standard for psoriasis.

The variability will be evaluated using the Kappa statistic (95% CI)

4. To assess the performance of the best predictive diagnostic criteria to identify psoriasis in children/young people currently diagnosed with indeterminate disease (sub-study).

The sensitivity and specificity of the best predictive diagnostic criteria in the sub-study population (indeterminate psoriasis) will be calculated. The reference standard will be the patient reported dermatologist's diagnosis. This data will only become available at a minimum of 2 years after the study has finished recruitment.

7. Blinding to exposure variables of evaluators making subjective judgments about the study.

The investigator performing the diagnostic criteria assessment (ie checking for the presence and absence of each criterion) will be blinded, where possible, to the disease status of the participant ie whether the participant is a case or control. Participants will be asked by their normal dermatology team, study investigator and in the participant information sheets not to share their disease status with the assessor.

Participants will be identified from previous clinic letters by the recruiting centre administrator or by the participant's normal dermatology team in clinic. Consecutive patients will be approached.

The 'other variables' data will be collected either directly from the participant (where the data will not unblind the assessor) or by the centre administrator from the medical record.

These measures will ensure that the diagnostic assessor remains blinded to the disease status of the participant and the data extractor has no involvement in the diagnostic criteria assessment.

8. **Methods of dealing with confounding, such as:**

    a. **Which confounders will be considered and how they will be defined**

This is a diagnostic accuracy study. Therefore, instead of confounders (alternative explanations for an observed association), predictors of diagnosis will be considered. These variables will be assessed in stratification (age, sex, assessor type, consultation type) or sensitivity analysus (ethnicity, blinding, disease severity).

    b. **Adjustment for confounders in statistical models** -NA

    c. **Restriction in analysis** – NA

    d. **Matching, including PS matching** - No

The cases and controls will not be matched. The rationale for not matching is that matching itself introduces a selection bias and that efficiency may be lost through a need for sufficient discordant pairs. It is also not possible to investigate the effect of a variable on the outcome for any variables matched at the study design stage (Rose and Laan 2009).

    e. **Self-controlled study designs - NA**

    f. **Statistical approach for any selection of a subset of confounders** - NA

    g. **Methods for assessing the level of confounding adjustment achieved** - NA

    h. **Sensitivity analyses for residual confounding** – NA

9. **Handling of missing data, including:**

    a. **How missing data will be reported;**

Diagnostic accuracy data will be collected prospectively at the time of assessment and recorded on the case report form. A confirmed dermatologist's diagnosis is part of the eligibility criteria. Therefore, it is anticipated that the proportion of missing data for the primary objective will be low.

Data on the other variables will be collected retrospectively from the medical records and therefore the proportion of missing data may be greater, depending on the completeness of documentation.

Missing or unclear data will be checked through central monitoring and additional details requested from the recruiting site if data is identified as missing.

The missing data will be reported as a percentage for each variable for cases and controls.

Indeterminate results are not anticipated within the study, as due to the index test and reference standard it is not possible to have an indeterminate results.

Multiple imputation will not be possible.

b. Sensitivity analyses for handling missing data;

Variables required for the analysis are the diagnostic criteria assessment and the dermatologist's diagnosis. A complete case analysis and all participant analysis will be calculated and compared in a sensitivity analysis. Two all participant analyses will be performed, one where all missing criteria will be coded as 'no' and one where all missing criteria are coded as 'yes'. This will allow the size of the effect from missing data to be assessed.

c. **How censored data will be treated, with rationale.** - NA

10. Fit of the model – if considered for a predictive model, including:

a. Criteria for assessing fit;

Prediction performance of the model will be assessed in terms of calibration and discrimination. Calibration assesses the agreement between the predicted outcome and the observed outcome. This will be assessed using the Hosmer-Lemeshow statistic. Discrimination is the ability to separate participants with and without the outcome of interest (ie those with and without psoriasis). Discrimination will be assessed using area under the receiver operator characteristic (ROC) curve. The adjusted R^2 will also be used as an overall measure of goodness of fit.

b. Alternative models in the event of clear lack of fit.

The study aims to develop the best predictive criteria based on the data obtained. If the model has a lack of a clear fit then it will be necessary to reconsider the exposure variables (diagnostic features) included in the model.

11. Interim analyses – if considered: No

a. Criteria, circumstances and possible drawbacks for performing an interim analysis and possible actions (including stopping rules) that can be taken on the basis of such an analysis

12. How the achieved patient population will be characterised:

a. Description of target population;

Percentages, means (standard deviations) or medians (interquartile range) will be used to describe the demographics, disease severity, current treatment and disease duration of the study population.

b. **Description of the analysis population if different, e.g. after PS matching or in IV analyses.** – NA

13. **Treatment of multiplicity issues not elsewhere covered.** – NA

14. **Sample size considerations should be presented, making explicit the data source from which the expected variation of relevant quantities and the clinically relevant differences are derived. It should be noted that in observational studies on data that already exist and where no additional data can be collected, sample size is not preclusive and the ethical injunction against 'underpowered' studies has no obvious force provided the results, in particular the 'absence of effect' and 'insufficient evidence', are properly presented and interpreted.**

The sample size is based on the primary objective.

Reporting guidance for risk prediction models (TRIPOD) have stated that there are no clear methods for calculating an adequate sample size. The guidance supports the current rule of thumb for sample size calculations of 10 events per variable (Moons, Altman et al. 2015). As there are 16 diagnostic features in the consensus agreed diagnostic criteria a sample size of 160 cases and 160 controls has been calculated. For a sample size of 320 participants, the precision to which a sensitivity and specificity of 80% could be estimated is calculated to be 73.6% to 86.4% (95% CI)(Hajian-Tilaki 2014).

No drop-out is anticipated as participant involvement is limited to one visit.

The sample size for the secondary objectives of inter-observer variability and variability in the reference standard is 20 cases and 20 controls. The nested substudy will recruit a convenience sample of up to 50 participants.

Minimising bias

Selection bias – A sample of consecutive patients will be recruited. All those approached but not recruited will be included on a screening log to demonstrate a non-selective approach. The study is inclusive of all children/ young people with plaque psoriasis who have active disease and are able to provide consent.

Index test – The diagnostic criteria assessment will be undertaken by an investigator who is, where possible, blinded to the diagnosis of the participant. In the standardised traininig for assessors it is emphasised that they should concentrate on assessing the presence and absence of each of the criteria, and not try to form a diagnosis. The study will test a two pre-specified thresholds i) a threshold for the consensus agreed criteria suggested through the eDelphi study (one of three major criteria, or in the absence of a major criterion, three of more more minor criteria), ii) a minimum threshold for the best predictive criteria based on discussion with a group of experts (80% sensitivity and 80% specificity).

Reference standard – The eligibility criteria specifies that only children/young people with a confirmed diagnosis of psoriasis/other inflammatory scaly skin disease can be recruited to the case control study. Therefore the reference standard is likely to correctly indentify participants. The diagnosis of skin disease is made without knowledge of the diagnostic criteria (index test) as this is a case-control study.

Flow and timing – The study as been designed to enable on the day recruitment directly from clinic, therefore the time between the reference standard and index test for most participants will be short. All participants receive the reference standard (a dermatologsist's diagnosis), therefore complete verification will be achieved. All participants will be included within the analysis and a complete data set sensitivity analysis is planned.

Limitations

A case-control design has been used which will overestimate the diagnostic ability. However, this is a feasible and suitable study design for developing diagnostic criteria. Furture studies will be planned to test the diagnostic performance of the refined criteria in a non case-control design, in the setting and population they are intended to be used. Spectrum bias has been introduced through conducting the study in a secondary care setting. This is a feasible setting to recruit sufficient patients with psoriasis and obtain the reference standard of a dermatologists diagnosis. All participants receive the same reference standard, a dermatologists diagnosis, however there is anticipated variation in how each clinicians reaches their diagnosis. The variation in the reference standard for psoriasis will be estimated through assessing agreement in diagnosis from a sample of clinical photographs.

Ensuring the end utility of the criteria

All participants recruited are required to have active (visible) disease at the time of assessment. To check whether the performance of the criteria differed in those with mild disease (using the new variable of categorised disease severity), a sensitivity analysis will be performed excluding all those with disease severity greater than mild.

DIPSOC study lead and coordinator

Dr. Esther Burden-Teh _____ Date 21st Feb 2019

DIPSOC Statistician

Dr. Sonia Gran _____ Date 21st Feb 2019

DIPSOC Chief Investigator

Prof. Kim Thomas _____ Date 21st Feb 2019.

References

Burden-Teh, E., K. S. Thomas, S. Ratib, D. Grindlay, E. Adaji and R. Murphy (2016). "The epidemiology of childhood psoriasis: a scoping review." British Journal of Dermatology **174**(6): 1242-1257.

Hajian-Tilaki, K. (2014). "Sample size estimation in diagnostic test studies of biomedical informatics." J Biomed Inform **48**: 193-204.

Moons, K. G., D. G. Altman, J. B. Reitsma, J. P. Ioannidis, P. Macaskill, E. W. Steyerberg, A. J. Vickers, D. F. Ransohoff and G. S. Collins (2015). "Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration." Ann Intern Med **162**(1): W1-73.

Pavlou, M., G. Ambler, S. R. Seaman, O. Guttmann, P. Elliott, M. King and R. Z. Omar (2015). "How to develop a more accurate risk prediction model when there are few events." BMJ **351**: h3868.

Rose, S. and M. J. Laan (2009). "Why match? Investigating matched case-control study designs with causal effect estimation." Int J Biostat **5**(1): Article 1.