GECo1 STATISTICAL ANALYSIS PLAN

Statistical Analysis PlanGECo1 - Case finding

Julie Barber & Federico Ricciardi Version
1.0: 25th February 2020

1. Introduction

This document reports the details of the main analyses planned for the GECo Case-finding study. Theaim is to pre-specify the analyses so that they are not influenced by the results when the final data are made available. This SAP does not prevent the analysis being adapted if justifiable situations ariseduring analysis, nor further analyses from being performed, which may become relevant while analysing the data or discussing the results with other stakeholders. This statistical analysis plan has been developed initially based on information from the study protocol (version 2) but has been combined with detail agreed through discussion with the team.

Writing of the SAP

This SAP was written by Julie Barber & Federico Ricciardi in collaboration with the core GECo Team (Checkley, Hurst, Pollard, Quaderi, Rykiel, Siddharthan).

Timing of analysis

Final analysis will be carried out when all data has been entered and checked and the analysis plan has been approved.

Data checking

Before analysis, checks for abnormal and inconsistent data will be performed. When inconsistenciesare found, data will be double checked with the study manager / data manager and corrected in thesource data if necessary, and may be set to missing otherwise.

Analysis

Analysis will be conducted in STATA (version 15) and R (version 3.5.0 or above) following the StandardOperating Procedures of the JRO Biostatistics group and in accordance with published guidelines.

Reporting

Reporting of results, will follow the STARD 2015 guidelines for reporting diagnostic accuracy studies(Cohen et al., 2016).

The reporting of analyses based on the 6-item COPD LMIC Assessment (COLA-6) score will be dependent on the final specification and testing of this new tool. COLA-6 results will not be publishedbefore publication of the reference paper, at least in abstract and/or pre-print form.

2. Study summary (from protocol)

Aims and objectives

The Global Excellence in COPD (GECo) Study is aimed at

(1) testing the validity of simple case-finding instruments for identifying individuals with COPD (GECo1);
(2) conducting a feasibility trial to evaluate COPD self-management plans facilitated bycommunity health workers both in Low and Middle Income (LMIC) countries (GECo2).

This Statistical Analysis Plan considers quantitative analyses for the case-finding study (GECo1).

Study population

We will enrol an age- and Sex-stratified random sample of full-time residents of the proposed studyareas in Nepal, Peru and Uganda using existing census data. Inclusion criteria are:

- aged ≥ 40 years;
- capable of performing spirometry;
- being a full-time resident of the community. Full-time residence will be defined as having lived inthe study area for more than 6 months.

Exclusion criteria are:

- self-reported pregnancy;
- having active pulmonary tuberculosis or being on medications for pulmonary tuberculosis;
- unable to do spirometry because of eye surgery, thoracic surgery, abdominal surgery, or myocardial infarction in the 3 months prior to study visit or a blood pressure > 180/100 mmHg.

Power calculation

Preliminary data from ongoing studies in Uganda (LiNK Study Cohort) have been analysed to assess the effectiveness of the LFQ as a proposed case-finding tool for COPD. Among 622 individuals, 32 tested positive for COPD. The sensitivity was 95% compared to standard spirometry. Receiver Operating Characteristic (ROC) analysis produced an area under the curve (AUC) equal to 0.69. Prior to initiation of GECo we anticipated modifications to the questionnaire based on formative research to improve the AUC to above 0.85. The sample size required to estimate the ROC area within 1.5% (based on a 95% confidence interval), assuming an anticipated sensitivity of 90% and specificity of 60%, and assuming 11% of those screened will have COPD is 9,669 participants. To ensure an adequatesample size is subsequently available for the feasibility trial, we plan to recruit a total of 10,500 subjects (3,500 at each site).

Original Statistical Analysis (as described in the Protocol)

Using a receiver operating characteristic (ROC) area analysis, we will examine diagnostic accuracy of the questionnaire scores in identifying cases of COPD (compared to spirometry). Curves will be obtained for the mLFQ and CAPTURE questionnaires alone and then each with addition of the PEF scores. Logistic regression models will be used to obtain the ROC curve and area (AUC), with 95% confidence interval. Estimates will be weighted based on census information from each site to betterreflect the population. A comparison between the ROC areas will be made by site.

3. Changes since publication of the protocol

The analysis plan specified in the published protocol has been updated with agreement from theresearch team and Steering committee in the following ways:

- In place of mLFQ +/- PEF, we are now using LFQ and COLA-6 scores
- The comparison between ROC areas will be made overall as well as by site

4. Available Data (from REDCap)

Data for the subjects included in the Case Finding study will be collected using a REDCap system on 6forms:

"Eligibility Criteria"

- Inclusion and exclusion criteria
- Age, Sex

"Demographic assessment form"

- Demographics
  - Contact Information
- Lifestyle assessment form
  - *Questions about smoking*
- Past Medical History assessment form
  - Family and personal illness history
- Treatment assessment form
  - Details of relevant medications being taken for breathing problems
- Patient Health Questionnaire (10 questions each allowing a choice of 4 responses)

"Socioeconomic survey form"

- Schooling
- Household information and use of fuels
- Exposure history form

"Pre and post-bronchodilator spirometry testing forms"

- Height
- Weight
- Pulse
- Systolic and diastolic blood pressure
- SpO2 – blood oxygen
- PEF (Peak expiratory flow)
- FEV_1 (forced expiratory volume in one second)
- FVC (forced vital capacity)
- FEV_1 post inhaler (forced expiratory volume in one second)
- FVC post inhaler (forced vital capacity)

[Notes:

- *All anthropometric measurements (standing / sitting height, weight, systolic and diastolic blood pressure, and pulse) will be measured 3 times. The triplicates are for quality control. The median of thethree measurements will be used for all measurements except for blood pressure. For systolic and diastolic blood pressure, we will disregard the first measurement and then take the average of the lasttwo measurements.*
- *The test acceptance columns in REDCap will be based on what the fieldworkers have deemed acceptable. For this reason, every patient will have their spirometry results 'over-read' by an expert locally who may check acceptability of the measurements. If unacceptable, the fieldworkers will go backand repeat the spirometry. If this happens, a second spirometry datasheet will be uploaded in REDCap.If a second spirometry data set comes through for a participant, then the first set of numbers will be replaced by the second set.*
- *The following steps will be followed to identify the best post-bronchodilator FEV1 and FVC measurements. The total number of acceptable tests will be calculated for each patient. If this numberis higher than 3, the maximum value of FEV1 and FVC of the acceptable tests will be used to calculate the FEV1/FVC ratio (and will be considered as best values). In cases where there are less than 3 acceptable tests, both best values and ratio will be missing.*]

"Respiratory symptoms form"

- Respiratory symptoms
- Lung Function Questionnaire (LFQ)
    - 5 item questionnaire about COPD symptoms/risk factors: rated 1 (never/low) 5, (very often/high).
- CAPTURE (COPD assessment in primary care to identify undiagnosed respiratory disease and exacerbation risk) Questionnaire
    - 5 item questionnaire – questions Y/N except question 5 (number of respiratory eventsrecorded as 0, 1 or 2+)
- Modified MRC dyspnea scale.
    - 1 item (records the degree of breathlessness related to activity and takes values from1 to 5)
- EQ-5D-3L (for Health Economics analysis)
    - 5 questions rated 1 to 3
- COPD Assessment test (CAT)
    - 8 questions score 0 (no) to 5 (bad)
- St. George's Respiratory Questionnaire (SGRQ)
    - 17 questions of varying form, covering 3 components (Symptoms, Activity, Impacts)[Note:

The items required for COLA-6 are collected as part of the above]

Details for scoring of these questionnaires are given below.

Census data will be used to provide weights for analysis. Numbers in the population by site, age andSex categories will be obtained from each site.

5. Scoring of questionnaires & definitions of key variables

Definition of Gold Standard diagnosis of COPD

For the primary analysis the gold standard COPD diagnosis is defined as a post-bronchodilator FEV1/FVC ratio that is below the lower limit of normal (LLN) for that population following ATS/ERS standardized guidelines. LLN will be calculated based on participants age, height (median of 3 measurements), Sex and ethnic group using published methodology (Quanjer et al 2012).

In secondary analyses a case of COPD will be redefined as:

- A post-bronchodilator FEV1/FVC ratio that is below 0.7 (rather than using LLN)(https://goldcopd.org/wp-content/uploads/2016/04/GOLD_Spirometry_2010.pdf).
- 2017 GOLD classification B, C or D (to focus on identification of clinically significant/symptomatic COPD) (https://goldcopd.org/wp-content/uploads/2017/02/wms-GOLD-2017-FINAL.pdf)
- Post-bronchodilator FEV1/FVC ratio that is below the lower limit of normal (LLN) and FEV1<60% predicted (Martinez et al 2017)

All other subjects will be considered as non-cases.

CAPTURE

The scoring of this questionnaire is based on a simple summation of patient responses to each of the5 items, yielding a questionnaire score ranging from 0 (no to all 5 questions) to 6 (yes to all questionsand > 2 respiratory events during the past year).

We will combine CAPTURE scores with PEF using published methods to define high and low COPD riskgroups (as in Martinez et al, 2017):

- a CAPTURE score <2 will be considered low risk;
- a CAPTURE score =2/3/4 with PEF≥250Lmin^{-1} for Females or PEFR≥350min^{-1} for Males will be considered low risk;
- a CAPTURE score =2/3/4 with PEF<250Lmin^{-1} for Females or PEFR<350min^{-1} for Males will be considered high risk;
- a CAPTURE score ≥5 will be considered high risk.

Lung Function Questionnaire (LFQ)

The LFQ questionnaire (Hanania et al, 2010) has 5 items. To calculate the total score, the item specificscore for the five items must be reversed. The coding is

- 01: Very often; 02: Often; 03: Sometimes; 04: Rarely and 05: Never for the first three items;
- 01: More than 30 years; 02: 21 to 30 years; 03: 11 to 20 years; 04: 10 years or more and 05:Never smoked for the item about smoking years;
- 01: 70 or older; 02: 60 to 69 years; 03: 50 to 59 years; 04:40 to 49; and 05: younger than 40years for the age.

If the sum is 18 or less, we will consider that the patient is at risk for COPD. There are no guidelines onhandling missing items for this questionnaire.

COLA-6 (*submitted for publication – reference to follow*)

The 6-item COLA questionnaire has 6 symptoms and function items which are combined with Age andPEF using the scoring set out below. This creates an overall score taking values from 0 to 9. There areno guidelines for handling missing items for this questionnaire.

Symptoms/Functional score	
Have you had whistling/wheezing in chest in last 12 months?	Yes = 1, No = 0
Have you brought up phlegm from your cheston most days or nights of the week during at least 3 months in a row in at least 2 years in a row?	
In the past 12 months, have you had to misswork or have your daily activities been impeded because of your respiratory problems?	
In the past 12 months, have you been hospitalized because of respiratory problems?	
Do you currently smoke?	
Do you use biomass fuel daily?	
Age score	
< 55 years	0
≥ 55 years	1
Peak expiratory flow score	
≥ 400 L/min	0
250 – 399 L/min	1
< 250 L/min	2

St. George's Hospital Respiratory Questionnaire (SGRQ)

This questionnaire (Jones et al, 1992) consists of 17 questions and 3 components (symptoms, activityand impacts component). There are specific weights for each question that can be found at the SGRQmanual. The Symptoms component is calculated from the summed weights for the positive responses to questions 1-8. The Activity component is calculated from the summed weights for the positive responses to questions 11 and 15. Finally, the Impacts component is calculated from the summed weights for the positive responses to questions 9-10, 12-14 and 16-17. The Total score is calculated bysumming all positive responses in the questionnaire and expressing the result as a percentage of the total weight for the questionnaire.

There are component specific guidelines for handling missing items. The Symptoms component will tolerate a maximum of 2 missed items. The weight for the missed item is subtracted from the total possible weight for the Symptoms component (662.5) and from the Total weight (3989.4). The Activitycomponent will tolerate a maximum of 4 missed items. The weight for the missed item is subtracted from the total possible weight for the Activity component (1209.1) and from the Total weight (3989.4). The Impacts component will tolerate a maximum of 6 missed items. The weight for the missed item issubtracted from the total possible weight for the Impacts component (2117.8) and from the Total weight (3989.4). (A reference manual is available on line - http://www.healthstatus.sgul.ac.uk/SGRQ_download/SGRQ%20Manual%20June%202009.pdf).

COPD Assessment Test (CAT)

The CAT questionnaire comprises 8 simple questions (score: 0-5) which are summed to provide a totalscore (Jones et al, 2019). It therefore has a scoring range of 0-40. Suggested grouping: >30: Very high;

>20: High; 10-20: Medium; <10: Low; 5: upper limit of normal (A user guide is available online - https://www.catestonline.org/content/dam/global/catestonline/documents/CAT_HCP%20User%20 Guide.pdf).

Patient Health Questionnaire (PHQ)

The Patient Health Questionnaire comprises 9 simple items (Kroenke et al, 2001). Each item is assigneda score of 0, 1, 2, and 3 for the response categories: not at all /several days/more than half the days/nearly every day. The PHQ total score is the sum for the first nine items and ranges from 0 to 27.Scores of 5, 10, 15, and 20 represent cut-points for mild, moderate, moderately severe and severe depression, respectively. Sensitivity to change has been confirmed.

6. Detailed analysis plan

6.1. Flow diagram

Using the STARD 2015 flow diagram format we will summarise by sites and overall the number ofsubjects:

- assessed for eligibility
- eligible (number (%))
- consenting (number (%))
- with and without COPD according to the gold standard (number (%)) (including indication ofnumbers with missing status)
- with and without CAPTURE/ PEF/ LFQ/ COLA-6 recorded (number (%)) (including indication of numbers with missing scores)

6.2 Subject characteristics

For each site and overall and by the presence or absence of COPD, we will summarise the following characteristics of participants:

- Sex
- Age
- Education level (No/ Primary school incomplete/Primary school complete/ Secondary/high schoolcomplete/ Any higher education)
- Employment (Y/N)
- Comorbidities (Y/N) and frequency of each comorbidity
- Smoking: the number of cigarettes smoked from LFQ will be used to calculate the pack years ofsmoking (1 pack year = 20 cigarettes smoked every day for 1 year)
- Ever smoked (Y/N)
- Biomass Exposure:
    - Ever exposed (Y/N)
    - Currently exposed (Y/N)
- Height, Weight, blood pressure
- BMI
- COPD GOLD status 1/2/3/4 & A/B/C/D
- Self-diagnosed COPD (Y/N)

- Self-diagnosed Asthma (Y/N)
- Lung function questionnaire score (LFQ) & LFQ risk group (high/low risk)
- CAPTURE questionnaire score & CAPTURE risk group (high/low risk)
- COLA-6 score
- Modified MRC dyspnoea score
- SGRQ score & subscores
- PEF, Pre max FEV1, Pre max FVC, Post max FEV1, Post max FVC, Post ratio (FEV1/FVC), % reversible(= post%-pre%)

All continuous variables will be described using the mean (standard deviation) or median (interquartilerange), as appropriate. Categorical variables will be described using frequencies and percentages.

A similar table of summary data will be constructed by diagnosis of COPD (Y/N)

The proportion of COPD cases diagnosed (Y/N, according to the defined gold standard) will be summarised overall and by country both as unweighted estimates and estimates weighted by censusdata to better reflect the population. Probability weights will be calculated in 8 age/gender categories(age categories: 40-44, 45-54, 55-64, 65-95) for each site as the number in category in the population(from census data) divided by the number in the category in the sample. The prevalence of restrictionon spirometry will also be calculated.

6.3 Screening tools

We will examine the distribution of LFQ, CAPTURE and COLA-6 questionnaire scores both overall andby COPD gold standard diagnosis group. We will tabulate high v low risk groups based on CAPTURE alone and in combination with PEF (see Section 5), and LFQ risk groups by COPD gold standard. This summary information will also be obtained by site.

Using spirometry results as the gold standard, we will construct ROC diagrams and calculate the areasunder the curve (AUCs) and associated 95% confidence intervals for the following continuous scores:

- LFQ
- CAPTURE
- COLA-6

Comparisons of AUC between the 3 pairs of diagnostic tools will be made using an algorithm suggestedby DeLong, DeLong, and Clarke-Pearson (1988) (and utilised in STATA command *roccomp*). Tests will be two sided and a P-value of 0.02 will be considered as the threshold for significance to allow for multiple testing.

In a sensitivity analysis we will compare the unadjusted ROC curves with those obtained afteradjustment for site, age and gender (through Stata's *roccurve* command) (Janes and Pepe (2008))

We will calculate sensitivity, specificity, PPV, NPV, odds ratio and % correctly classified with 95% confidence intervals for the following, based on published thresholds (Section 5 above)):

- CAPTURE + PEF
- LFQ

Comparison of diagnostic ability between these combinations will be made using McNemar's test andreported alongside estimates of absolute gain with 95% confidence intervals. All tests will be two sided.

In all analyses estimates will be weighted based on census information.

The extent of missing data for the diagnostic test information and spirometry gold standard will be reported and characteristics of those with missing values tabulated. Depending on the extent of missing data, imputation methods or other sensitivity analyses will be considered to assess the impactthis has had on the study results.

Estimates of diagnostic accuracy will also be reported by country.

6.4 Secondary analyses

The analyses described above will be repeated with a case of COPD redefined as:

- A post-bronchodilator FEV1/FVC ratio that is below 0.7 (rather than using LLN)
- 2017 GOLD classification B, C or D (to focus on identification of clinically significant/symptomatic COPD)
- Post-bronchodilator FEV1/FVC ratio that is below the lower limit of normal (LLN) and FEV1<60% predicted (Martinez et al 2017)

All other subjects will be considered as non-cases.

7. References

- Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracystudies: explanation and elaboration. BMJ Open 2016;6:e012799. doi:10.1136/bmjopen-2016- 012799.
- DeLong, E. R., D. M. DeLong, and D. L. Clarke-Pearson. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics 44: 837–845.
- Hanania NA, Mannino DM, Yawn BP, et al. Predicting risk of airflow obstruction in primary care: validation of the lung function questionnaire (LFQ). Respir Med. 2010;104:1160-1170.
- Janes H, Pepe M.S. Adjusting for Covariates in Studies of Diagnostic, Screening, or Prognostic Markers: An Old Concept in a New Setting . *American Journal of Epidemiology*, Volume 168, Issue 1, 1 July 2008, Pages 89–97, https://doi.org/10.1093/aje/kwn099
- Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure for chronic airflow limitation - the St George's Respiratory Questionnaire. Am Rev Respir Dis 1992;145:1321-1327.
- Jones PW, Harding G, Berry P, et al. Development and first validation of the COPD AssessmentTest. Eur Respir J 2009: 34: 648-654
- Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. JGen Intern Med. 2001;16(9):606–613. doi:10.1046/j.1525-1497.2001.016009606.x

- Martinez FJ, Mannino D, Leidy NK, et al. A New Approach for Identifying Patients with Undiagnosed Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2017;195(6):748–756. doi:10.1164/rccm.201603-0622OC.
- Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, Enright PL, Hankinson JL, Ip MSM,Zheng J, Stocks J, the ERS Global Lung Function Initiative. Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. *European Respiratory Journal* Dec 2012, 40 (6) 1324-1343; DOI: 10.1183/09031936.00080312