MENINGITIS CASE-BASED DATA VALIDATION CHECKLIST

This tool is designed to serve as a guide for data management and surveillance staff for the validation of case-based meningitis data. Validation of case-based surveillance (CBS) data significantly enhances the completeness and quality of data collected and is an effective method to assess sensitivity of CBS data systems and inform improvements to surveillance and data management activities. All variable names and coding in this guide are defined in the MenAfriNet data dictionary and may need to be modified depending on the variable names and coding used in each country.

This guide has two components: 1) a minimum checklist of data validation standards including priority variables and formats to check and 2) a second, more in-depth checklist of standards. A combination of both can be used and they are meant to be adapted to country-specific contexts.

MINIMUM STANDARDS

1. COMPARISON OF CASE-BASED DATA TO AGGREGATE

Ensuring that you have the most recent aggregate and case-based datasets, compare the totals below using the same time period (either by month or by week) to determine whether they match. Investigate discrepancies to ensure errors are corrected as possible:

- ☐ Total suspect cases
- ☐ Total deaths
- ☐ Total confirmed cases (if available in aggregate dataset)
- ☐ Total probable cases (if available in aggregate dataset)
- ☐ Totals of each confirmed pathogen (if available in aggregate dataset)

Consider any contexts that might cause discrepancies or alter the totals and ensure the data are reflective of these (e.g. if there was an outbreak during a certain time period, do the total suspect and confirmed cases and confirmed pathogens above reflect the outbreak?)

2. GENERAL DATA CHECKS

☐ All dates recorded as DD/MM/YYYY

- ☐ 'DateCreated'
- ☐ ‘DateOnset’
- ☐ ‘DateConsultation’
- ☐ ‘DateOfBirth’
- ☐ ‘DateMenA’
- ☐ ‘DateHib_LASTDOSE’
- ☐ ‘DatePCV_LASTDOSE’
- ☐ ‘DatePenta’
- ☐ ‘DateSpecimenCollected’
- ☐ ‘DateHFNotifiedDistrict’
- ☐ ‘DateResultSentHF’
- ☐ ‘DateResultSentDistrict’
- ☐ ‘DateResultSentMOH’
- ☐ ‘DateSpecimenReceivedDistrict’
- ☐ ‘DateSpecimenReceivedRegion’
- ☐ ‘DateSpecimenReceivedNRL’
- ☐ ‘DatePCR’
- ☐ ‘DateResultReceivedDistrict’
- ☐ ‘DateFormSentDistrict’
- ☐ ‘DateFormReceivedDistrict’

☐ 'DateOfBirth’ is the earliest date. No other dates can occur before 'DateOfBirth'

☐ After ‘DateOfBirth’, the following priority dates may occur on the same date, or in logical chronological order from earliest to most recent: ‘DateOnset,’ ‘DateConsultation,’ ‘DateSpecimenCollected’

☐ ‘CountryCode,’ ‘RegCode,’ and ‘DistCode’ variables include only 3 relevant letters (Not included in data dictionary)

☐ Accents and acronyms are not used for laboratory names and geographic areas (such as ‘District’ or ‘Region’)

3. VALIDATION OF PRIORITY VARIABLES

Ensure that the following priority variables are completed correctly by comparing the data to the case reporting form and consultation register, if possible, and that the formatting is consistent.

☐ ‘EpidNumber’ is complete and follows the format of: COU/REG/DIS/YY/‘MEN’/####-By order: CCC is 3 letters for country code; RRR is 3 letters for region code; DDD is 3 letters for district code; YY is 2 last numbers for year of consultation; MEN for disease (does not change), and 4 numbers for case number

☐ ‘EpidNumber’ - unique for all cases (there are no duplicates)

☐ ‘Country’- completed for all cases and fully spelled out

☐ ‘Region’- completed for all cases and fully spelled out

☐ ‘District’- completed for all cases and fully spelled out

☐ Only one of the following is completed numerically: ‘AgeDays’, ‘AgeMonths’, ‘Ageyears’. If patient’s age is <1 month, enter age in ‘AgeDays’ (0-31); if patient is ≥ 1 month, enter age in ‘AgeMonths’ (1-11); if patient is ≥ 1 year, enter age in ‘AgeYears’ (1-120)

☐ ‘Sex’- formatted as ‘F’ for female, ‘M’ for male or ‘9’ for unknown

☐ ‘Outcome’ is completed for all cases as ‘1’ for alive (in treatment /recovered), ‘2’ for died or ‘9’ for unknown

☐ After ‘DateOfBirth’, the following priority dates may occur on the same date, or in logical chronological order from earliest to most recent: ‘DateOnset,’ ‘DateConsultation,’ ‘DateSpecimenCollected’

- ☐ ‘SpecimenCollected’ is completed for all cases as ‘1’ - yes, ‘2’ - no, or ‘9’ – Unknown‘ in accordance to laboratory database or case report form. The following is completed for the validation of laboratory data for all cases that had a specimen collected ('SpecimenCollected' = 1):
  - ☐ ‘DateSpecimenCollected’ is completed
  - ☐ ‘CultureNRL’ and ‘CultureOtherNRL’ (including if contaminated or not done) are completed according to the CRF and/or internal laboratory database, if separate from case-based surveillance database for all cases. Coded as 1-NmA 2-NmC 3-NmW 4-NmY 5-NmB 6-NmX 7-Nm Indeterminate 8-S. pneumoniae 9-Hib 10-H. influenzae (non-b) 11-Streptococcus B 12-Other organism 13-Contaminated 14-Negative 15-Not done
  - ☐ ‘LabNameGram’ is completed
  - ☐ ‘GramResult’ is completed (including if not done) according to the case reporting. Coded as 1-GPD 2-GND 3-GPB 4-GNB 5-Other germs 6-No organisms 9-Not done.
  - ☐ ‘ContainerType’ is completed for all cases. Coded as 1- TI, 2- Cryotube, 3- Dry tube, 4- Other
- ☐ ‘DateOnset’ is completed for all cases in accordance with the consultation register
- ☐ If at least one of the following variables is complete: ‘DateSpecimenReceivedNRL’, ‘DatePCR’, ‘PCR’, then all three need to be complete, with ‘PCR’ coded as : 1-NmA 2-NmC 3-NmW 4-NmY 5-NmB 6-NmX 7-Nm Indeterminate 8-S. pneumoniae 9-Hib 10-H. influenzae (non-b) 11-Streptococcus B 12-Other organism 13-Contaminated 14-Negative 15-Not done. (If coded as 12, ‘PCROther’ should be completed)
- ☐ ‘FinalClassification’ is completed for all cases, coded as 1- Confirmed, 2-Probable, 3-Suspect (only 1 category should be selected per case; case should never be classified as discarded; see annex 1 for case definitions)
- ☐ ‘FinalResultNRL’ should match results of tests done in dataset. If other test results differ from PCR results, use the PCR results to match/complete FinalResultsNRL variable. Code the variable as 1-NmA 2-NmC 3-NmW 4-NmY 5-NmB 6-NmX 7-Nm Indeterminate 8-S. pneumoniae 9-Hib 10-H. influenzae (non-b) 11-Streptococcus B 12-Other organism 13-Contaminated 14-Negative 15-Not done. ‘FinalResultNRL’ is considered the final classification of pathogens for confirmed cases and should always be completed when a case is confirmed

GOLD STANDARD PROCESS

1. COMPARISON OF CASE-BASED DATA TO AGGREGATE

Minimum standards included in this section are in orange box, and all additional standards are part of the gold standard process.

Ensuring that you have the most recent aggregate and case-based datasets, compare the totals below using the same time period (either by month or by week) to determine whether they match. Investigate discrepancies to ensure errors are corrected as possible:

- ☐ Total suspect cases
- ☐ Total deaths
- ☐ Total confirmed cases (if available in aggregate dataset)
- ☐ Total probable cases (if available in aggregate dataset)
- ☐ Totals of each confirmed pathogen (if available in aggregate dataset)

Consider any contexts that might cause discrepancies or alter the totals and ensure the data are reflective of these (e.g. if there was an outbreak during a certain time period, do the total suspect and confirmed cases and confirmed pathogens above reflect the outbreak?)

2. GENERAL FORMATTING OF CASE-BASED DATA

Ensure that all case-based data are recorded in the same manner for each quarter:

- ☐ All required variables are present in the dataset (see data dictionary)
- ☐ All dates recorded as DD/MM/YYYY
  - ☐ 'DateCreated'
  - ☐ ‘DateOnset’
  - ☐ ‘DateConsultation’
  - ☐ ‘DateOfBirth’
  - ☐ ‘DateMenA’
  - ☐ ‘DateHib_LASTDOSE’
  - ☐ ‘DatePCV_LASTDOSE’
  - ☐ ‘DatePenta’
  - ☐ ‘DateSpecimenCollected’
  - ☐ ‘DateHFNotifiedDistrict’
  - ☐ ‘DateResultSentHF’
  - ☐ ‘DateResultSentDistrict’
  - ☐ ‘DateResultSentMOH’
  - ☐ ‘DateSpecimenReceivedDistrict’
  - ☐ ‘DateSpecimenReceivedRegion’
  - ☐ ‘DateSpecimenReceivedNRL’
  - ☐ ‘DatePCR’
  - ☐ ‘DateResultReceivedDistrict’
  - ☐ ‘DateFormSentDistrict’
  - ☐ ‘DateFormReceivedDistrict’
- ☐ Numeric variables include only numeric data (see variable format in data dictionary)
- ☐ Character variables include only characters and are all capitalized with no accents (see variable format in data dictionary)
- ☐ CountryCode, RegCode, and DistCode variables include only 3 relevant letters

3. CROSS-CHECKING REGISTRY

Using health facility registers to confirm recorded data, ensure that the following variables are complete and accurate in the case- based dataset:

- ☐ 'FirstName' and 'FamilyName' of patient - recorded in all capitalization and no accents
- ☐ 'DateOfbirth' of patient
- ☐ 'Sex'- formatted as 'F' for female, 'M' for male or '9' for unknown
- ☐ 'DistrictofResidence' and 'Village' - recorded as patient's place of main residence
- ☐ If vaccination information available in registry for patient, confirm that patient information matches vaccination data recorded in dataset
- ☐ Only one of the following variables is completed for age: 'AgeYears', 'AgeDays', or 'AgeMonths' (completed when DateOfBirth is missing)
- ☐ 'AgeYears' - should be less than 120
- ☐ 'AgeMonths' - should be less than 12
- ☐ 'AgeDays' - should be less than 31

4. CLEANING OF CLINICAL DATA

Review format and data coding of the following variables as detailed below:

- ☐ 'Outcome' variable is completed - recorded as 1-Alive (in treatment /recovered), 2-Died or 9-Unknown
- ☐ 'InOutPatient' variable is completed - recorded as 1-Inpatient, 2-Outpatient or 9-Unknown
- ☐ Optional: If symptom variables (fever, headache, etc.) are in use, ensure each is coded as 1-Yes, 2-No, or 9-Unknown (See data dictionary for complete list of optional symptom variables)

5. CLEANING OF VACCINATION DATA

Review date format and data coding of the following variables as detailed below:

- ☐ 'VaccinationStatus' completed as 1-Yes, 2-No or 9-Unknown
- ☐ 'MenA', 'Penta', 'Hib', 'PCV' are completed as 1-Yes by card or register, 2-Yes by verbal history, 3-No, or 9-Unknown
- ☐ 'DateMenA', 'DatePenta', 'DateHib_LASTDOSE', and 'DatePCV_LASTDOSE' are completed, if respective vaccine variables (above) were recorded as '1'
- ☐ If 'Hib' completed as '1', DoseHib is completed as 1,2, or 3 (equaling total doses given)
- ☐ If 'PCV' completed as '1', DosePCV is completed as 1,2, or 3 (equaling total doses given)
- ☐ If 'OtherVaccine' completed as '1', 'NameOtherVaccine' and 'DateOtherVaccine' are completed

6. VALIDATION AND CLEANING OF LABORATORY DATA

- ☐ 'SpecimenCollected' is completed for all cases as '1' - yes, '2' - no, or '9' – Unknown' in accordance to laboratory database or case report form. The following is completed for the validation of laboratory data for all cases that had a specimen collected ('SpecimenCollected' = 1):
  - ☐ 'DateSpecimenCollected' is completed
  - ☐ 'SpecimenType' is completed numerically as 1-CSF, 2-blood or 3-other (If result of 3, 'SpecimenTypeOther' must also be completed)
  - ☐ 'Aspect' completed numerically for all CSF specimens collected as 1-Clear, 2-Turbid, 3-Bloody, 4- Xanthochromic, 5-Citrin, 6-Cloudy, Purulent or 9- not done for all cases with specimens collected
  - ☐ 'ContainerType' is completed numerically with 1-TI, 2-Cryotube, 3-Dry tube or 4-Other (note: multiple selection is possible)
  - ☐ 'ConditionDrytubeNRL', 'ConditionTINRL' and 'ConditionCryotubeNRL' is completed in all cases with 'SpecimenCollected' coded as 1 and in alignment with selection of 'ContainerType'
  - ☐ District, regional (if applicable) and NRL (if applicable) test data and results recorded for each case
  - ☐ 'Districtlab', 'RegionalLab' and 'LabNameNRL' are completed with the corresponding lab names in all capitalization with no accents
  - ☐ 'GramResult' is completed numerically as 1-GPD, 2-GND, 3-GPB, 4-GNB, 5-Other germs, 6-No organisms or 9-Not done for all cases in which gram test was performed. (If result of 5, 'GramResultOther' must also be filled out)
  - ☐ 'LabNameGram' is completed in all capitalization without accents if 'GramResult' completed
  - ☐ If latex agglutination performed, 'LatexResult' must be completed numerically as 1-NmA, 2-NmC, 3- NmW/Y, 4-NmB/E. coliK1, 5-NmX, 6-S.pneumoniae, 7-Hib, 8-Streptococcus B, 9-Negative or 10-Not done
  - ☐ If rapid diagnostic test performed, 'RDT' must be completed numerically as 1-NmA, 2-NmC, 3-NmW, 4-NmY, 5-NmB, 6-NmX, 7-Nm Indeterminate, 8-S. pneumoniae, 9-Hib, 10-H. influenzae (non-b), 11- Streptococcus B, 12-Other organism, 13-Contaminated, 14-Negative or 15-Not done. (If result of 12, 'RDTResultOther' must also be filled out)

☐ ‘CulturePeripheral’ is completed numerically as 1-NmA, 2-NmC, 3-NmW, 4-NmY, 5-NmB, 6-NmX, 7-Nm Indeterminate, 8-S. pneumoniae, 9-Hib, 10-H. influenzae (non-b), 11-Streptococcus B, 12-Other organism, 13-Contaminated, 14-Negative or 15-Not done. (If result of 12, ‘CultureOtherPeripheral’ must also be filled out)

☐ If ‘OtherTest’ coded as 1, enter test name and results in ‘OtherTestName’ and ‘OtherTestResults’

☐ ‘CultureNRL’ is completed as 1-NmA, 2-NmC, 3-NmW, 4-NmY, 5- NmB, 6-NmX, 7-Nm Indeterminate, 8-S. pneumoniae, 9-Hib, 10-H. influenzae (non-b), 11- Streptococcus B, 12-Other organism, 13-Contaminated, 14-Negative or 15-Not done. (If 12 recorded, ‘CultureResultOther’ completed)

☐ If at least one of the following variables is complete: ‘DateSpecimenReceivedNRL’, ‘DatePCR’, ‘PCR’, then all three need to be complete, with PCR coded as : 1-NmA 2-NmC 3-NmW 4-NmY 5-NmB 6-NmX 7-Nm Indeterminate 8-S. pneumoniae 9-Hib 10-H. influenzae (non-b) 11-Streptococcus B 12-Other organism 13-Contaminated 14-Negative 15-Not done. (If coded as 12, ‘PCROther’ should be completed) If SPN serotyping performed, ‘SerotypeSpn’ completed as text in all capitalization with no accents.

☐ If white cell count performed, ‘WhiteCellCount’ completed as numeric in thousands cells/mm^3 (i.e. 20000). All unknowns should be coded as ‘9’

☐ ‘FinalResultNRL’ should match results of tests done in dataset. If other test results differ from PCR results, use the PCR results to match/complete FinalResultsNRL variable. Code the variable as 1-NmA 2-NmC 3-NmW 4-NmY 5-NmB 6-NmX 7-Nm Indeterminate 8-S. pneumoniae 9-Hib 10-H. influenzae (non-b) 11-Streptococcus B 12-Other organism 13-Contaminated 14-Negative 15-Not done. ‘FinalResultNRL’ is considered the final classification of pathogens for confirmed cases and should always be completed when a case is confirmed.

☐ If culture, PCR, or final results listed for a case do not match, results need to be re-confirmed. If still not matching record, provide explanation in ‘Observations’ variable.

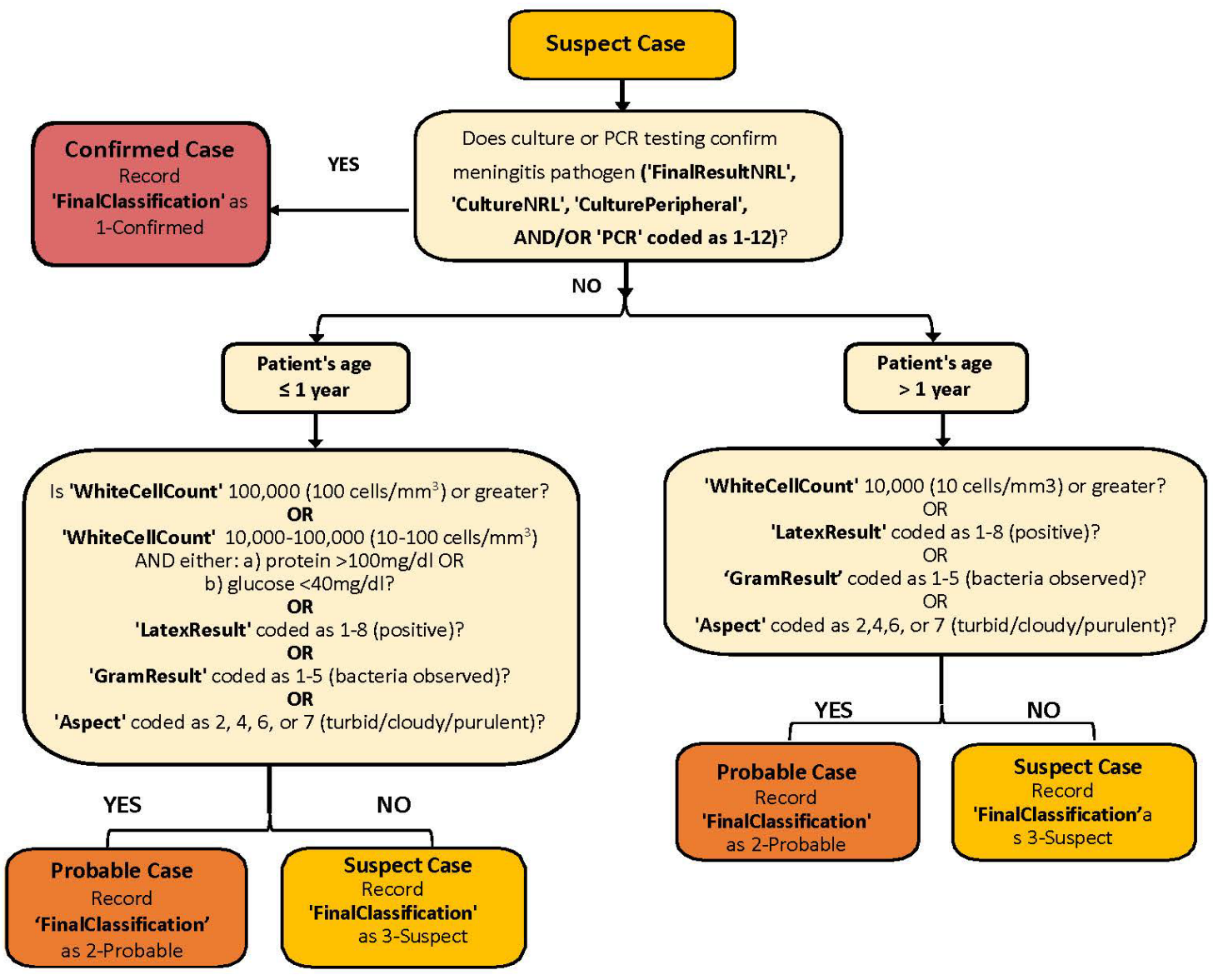
☐ If susceptibility testing performed, results recorded as S-Sensitive, R-Resistant, I-Intermediate or 9- Not done for any completed antibiotics (‘Penicillin’, ‘Ceftriaxone’, ‘Ciprofloxacin’ and ‘Oxacillin’)

☐ If other antibiotic not listed is tested, name of the antibiotic is provided in the ‘OtherAntibiotic’ variable as text in full capitalization with no accents and the result of the test as S-Sensitive, R- Resistant, I-Intermediate or 9-Not done

7. VALIDATION OF CASE CLASSIFICATION

☐ ‘FinalClassification’ variable is completed numerically as 1-Confirmed, 2-Probable OR 3-Suspect (only 1 category should be selected per case; case should never be classified as discarded; see annex 1 for case definitions).

Use the following diagram for the validation of case classification data:



8. VALIDATION OF TRACKING AND CHRONOLOGICAL DATA

Ensure the following tracking and date variables are completed and use the formats and coding indicated below:

- ☐ 'EpidNumber' is complete and follows the format of: COU/REG/DIS/YY/'MEN'/####-By order: CCC is 3 letters for country code; RRR is 3 letters for region code; DDD is 3 letters for district code; YY is 2 last numbers for year of consultation; MEN for disease (*does not change*), and 4 numbers for case number
- ☐ 'EpidNumber' - unique for all cases (there are no duplicates)
- ☐ 'Country'- completed for all cases and fully spelled out
- ☐ 'Region'- completed for all cases and fully spelled out
- ☐ 'District'- completed for all cases and fully spelled out
- ☐ Case Number is recorded as numeric and matches the case number at the end of the 'EpidNumber'
- ☐ 'ReporterHF' is completed as capitalized text without accents
- ☐ 'ReporterPhone' is recorded as numeric only and excludes dashes or symbols identification
- ☐ 'SpecimenLabelID' numerically coded as 1-Yes, 2-No or 9-Unknown and completed for all cases in which 'SpecimenCollected' is coded as '1'

Check the following dates for the correct chronology as indicated below (see data dictionary for all date variables):

- ☐ No recorded dates are after the present date
- ☐ 'DateOfBirth' is the earliest date. No other dates can occur before 'DateOfBirth'
- ☐ After date of birth, the following priority dates may occur on the same date, or in logical chronological order from earliest to most recent: 'DateOnset', 'DateConsultation', 'DateSpecimenCollected' 'DateCreated', and 'DateReceivedNRL'
- ☐ All dates of vaccination 'DateMenA', 'DatePCV', 'DatePenta' and 'DateHib' cannot be after 'DateConsultation' or before 'DateOfBirth'
- ☐ 'DateSpecimenCollected' occurs on or after 'DateConsultation' and is completed for all cases in which 'SpecimenCollected' is coded as '1'
- ☐ 'DateSpecimenReceivedDistrict', 'DateSpecimenReceivedRegion', 'DateSpecimenReceivedNRL' are completed as applicable for all cases with 'SpecimenCollected' coded '1' and occur on or after 'DateSpecimenCollected'
- ☐ 'DatePCR' occurs on or after 'DateSpecimenReceivedNRL'
- ☐ 'DateHFNotifiedDistrict' and 'DateFormSentDistrict' occur on or after 'DateConsultation'
- ☐ 'DateFormReceivedDistrict' occurs on or after 'DateFormSentDistrict'
- ☐ 'DateSpecimenReceivedDistrict', 'DateSpecimenReceivedRegion' and 'DateSpecimenReceivedNRL' occur on or after 'DateSpecimenCollected'
- ☐ 'DateSpecimenReceivedRegion' occurs on or after 'DateSpecimenReceivedDistrict' (or on or after 'DateSpecimenCollected' if 'DateSpecimenReceivedDistrict' is blank)
- ☐ 'DateSpecimenReceivedNRL' occurs on or after 'DateSpecimenReceivedRegion' (or on or after 'DateSpecimenCollected' if 'DateSpecimenReceivedRegion' is blank)
- ☐ 'DateResultReceivedDistrict', 'DateResultSentHF', 'DateResultSentDistrict' and DateResultSentMOH' completed for all cases in which 'SpecimenCollected' is coded as '1'
- ☐ 'DateResultReceivedDistrict' occurs on or after 'DateResultSentDistrict'
- ☐ 'DateResultSentMOH', 'DateResultSentDistrict' and 'DateResultReceivedDistrict' occurs on or after 'DatePCR'
- ☐ 'DateResultSentHF' occurs on or after 'DateResultReceivedDistrict'

9. FINALIZATION, FEEDBACK, AND DISSEMINATION OF DATASET (National)

Implement the following steps to finalize the validation process:

- ☐ Dataset saved separately as final, validated dataset (consider having 'FinalValidated' in dataset file name)
- ☐ Use the final validated dataset for bulletin and epidemic reports
- ☐ Final, validated data sent to all regions for further dissemination (regions and districts informed of corrected data)
- ☐ Generate a list of recommended changes in data collection or entry based on common data entry errors identified (nationally and by region/district). Share this feedback with all regions and districts.
- ☐ Final, validated dataset shared with Ministry of Health, relevant in-country partners, WHO colleagues and other relevant partner

ANNEX 1

(Ref: Standard operating procedures for surveillance of meningitis preparedness and response to epidemics in Africa)

Suspected meningitis case: Any person with sudden onset of fever (> 38.5 °C rectal or 38. °C axillary), and neck stiffness or other meningeal signs, including bulging fontanel in infants
Probable meningitis case: Any suspected case with macroscopic aspect of cerebeospinal fluid (CSF) turbid, cloudy, or purulent; or with a CSF leukocyte count >10 cells/mm3 or with bacteria identified by Gram stain in CSF; or positive antigen detection (for example, by latex agglutination testing) in CSF. **In infants:** CSF leukocyte count >100 cells/mm3; or CSF leukocyte count 10-100 cells/mm3 and either an elevated protein (>100 mg/dl) or decreased glucose (<40 mg/dl) level.
Confirmed meningitis case: Any suspected or probable case that is laboratory confirmed by culturing or identifying (i.e. polymerase chain reaction) a bacterial pathogen (Neisseria meningitidis, Streptococcus pneumoniae, Haemophilus influenzae type b) in the CSF or blood