Manual for the Laboratory-based Surveillance of Measles, Rubella, and Congenital Rubella Syndrome

Chapter 7. Molecular epidemiology of measles and rubella

Overview

The genetic analysis of measles and rubella viruses has been well-documented as a valuable tool for characterizing outbreaks and tracking transmission pathways [1-12]. In addition, the identification of genotypes for confirmed cases is an important performance indicator for laboratory activities undertaken in support of national and global goals for measles and rubella elimination. The data collected can be used to assess the sensitivity of the surveillance to detect imported cases. In addition, the accumulated molecular surveillance data from the laboratory provide evidence for the impact of immunization programmes, including the interruption of endemic virus circulation, an essential criterion for the verification of elimination.

The descriptions of the temporal and geographic details of cases and outbreaks, as well as molecular characterization of the viruses are a vital component of the verification of elimination. The ability to identify a genotype is particularly important when the source of infection is unknown. However, there is an ongoing challenge to obtain adequate clinical samples to meet the goal for high quality virologic surveillance that is required for GMRLN laboratories. *WHO recommends that molecular surveillance be conducted during all phases of measles and rubella control to provide an accurate description of the global distribution of genotypes.*

There may be viruses with novel genotypes in circulation, particularly for rubella, since virologic surveillance for rubella has lagged that for measles. However, it is anticipated that as elimination programmes for measles and rubella progress, more of the recognized genotypes will join those considered to be inactive or extinct. The geographic associations that had initially proven to be epidemiologically useful for many of the measles genotypes are no longer valid since there are fewer genotypes circulating and most of these have a global distribution. For this reason, additional regions of the measles and rubella genomes have been identified or are under evaluation for inclusion in phylogenetic analyses to improve the ability to discern genetic differences among viruses within a genotype that reflect epidemiologic relationships.

The sections that follow will include the methodologies and links to protocols that are recommended for genotype determination by sequence analysis. In addition, the guidelines and limitations for the interpretation of molecular epidemiologic data will be discussed. Instructions for reporting genotypes, proposing new genotypes and nominating a virus variant as a named strain are included, along with detailed steps for accessing and navigating the measles nucleotide surveillance (MeaNS) and the rubella nucleotide surveillance (RubeNS) databases.

In this chapter:

Part A. Measles

7.1 Phylogenetic diversity and nomenclature for measles genotypes

7.2 Integration of measles molecular and epidemiological data

- 7.2.1 The named strain designation for circulating variants within a genotype

7.3 Overview and methods for determination of measles genotypes

7.4 Guidelines for reporting a new measles genotype and use of data from MeaNS

7.5 The measles nucleotide surveillance (MeaNS) database

- 7.5.1 Submission of sequences to the MeaNS database
- 7.5.2 Sequence data and analysis tools available in MeaNS
- 7.5.3 Sequence data obtained on MeaNS and determination of the source of virus

Part B. Rubella

7.6 Phylogenetic diversity and nomenclature for rubella genotypes

7.7 Integration of rubella molecular and epidemiological data

7.8 Overview and methods for determination of rubella genotypes

7.9 Guidelines for reporting a new rubella genotype

7.10 The rubella nucleotide surveillance (RubeNS) database

- 7.10.1 Submission of sequences to the RubeNS database
- 7.10.2 Sequence data and analysis tools available in RubeNS
- 7.10.3 Interpretation of information from RubeNS analyses

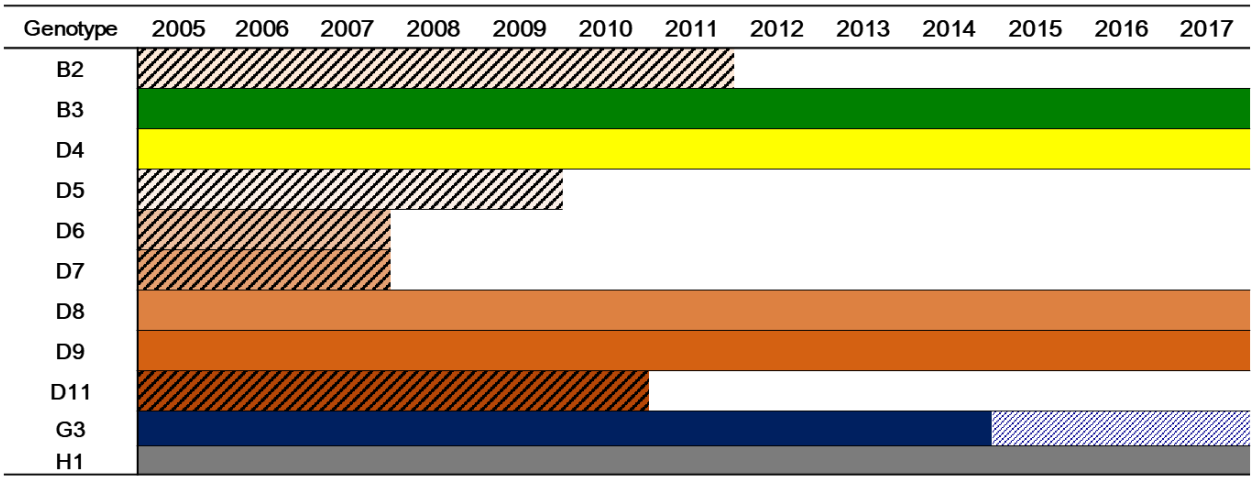
7. 11 Methods and prospects for enhancing resolution of sequence data for molecular epidemiology

7.1 Phylogenetic diversity and nomenclature for measles genotypes

A uniform method to describe the genetic diversity of wild-type measles viruses was adopted in 1998, following a meeting of experts hosted by WHO [13]. The published report of the meeting provided the recommended protocols for sequencing and phylogenetic analysis and included reference measles strains for each of the 15 genotypes that had been characterized at that time. The genotypes were assigned to one of the 8 major clades (A-H) according to phylogenetic analyses of the 450 nucleotides that code for the carboxyl-terminal 150 amino acids of the nucleoprotein (N-450). Clade A includes all vaccine strains of measles virus. Subsequent reports specified the requirements for a new genotype including the minimum nucleotide divergence from recognized genotypes. Each update of the nomenclature for measles genotypes has provided the status of known genotypes, along with the designated reference strain(s) [14-19].

The number of clades that have been identified and described has remained unchanged since 1998, however the number of genotypes within a clade has increased, particularly within clade D. Wild-type viruses in clade A are considered to be extinct. As of the 2015 update [19], 24 measles genotypes have been recognized. Clades A, E and F have continued to consist of only a single genotype. All of the recognized genotypes have one reference strain, except for B3, C2, D5 and D7, which have two reference strains. During 2005-2015, 11 wild-type genotypes were detected, and in 2017 only five genotypes in three clades (B3, D4, D8, D9, and H1) were detected/reported to MeaNS. Refer to Figure 7.1, Circulation of measles viruses 2005-2017.

Figure 7.1 Circulation of measles genotypes 2005-2017*



Solid color bar: Genotypes considered to be actively circulating
Hatched bar: Genotypes that have not been detected for at least 3 years

*Based on global surveillance for measles viruses which may be incomplete

The letter designation of the clade is used when referring to a genotype if it is the only member of a clade (e.g., genotype A). In 2001, when an additional genotype was recognized within clade H, the original viruses comprising genotype H were grouped together as genotype H1 and the new genotype H viruses were designated as genotype H2 [14]. A list of the measles genotypes with the year last detected, the designated reference strain(s), and GenBank accession number are provided in Table 7.

Table 7 Measles genotypes and reference strains

Genotype	Last Detected*	Reference strain	GenBank H	Genbank N
A	2008	MVi/Maryland.USA/0.54	JX436452	EU139076
A		Vaccine (Edmonston Zagreb)	U03669	U01987
B1[a]	1983	MVi/Yaounde.CMR./12.83	AF079552	U01998
B2	2011	MVi/Libreville.GAB/0.84	L46753	U01994

B3	ongoing	MVi/New York.USA/0.94	L46752	L46753
		MVi/Ibadan.NGA/0.97/1	AJ239133	AJ232203
C1 [a]	1992	MVi/Tokyo.JPN/0.84	AY047365	AY043459
C2	2007	MVi/Maryland.USA/0.77	M81898	M89921
		MVi/Erlangen.DEU/0.90	Z80808	X84872
D1 [a]	1986	MVi/Bristol.GBR/0.74	Z80805	D01005
D2[a]	2005	MVi/Johannesburg.ZAF/0.88/1	AF085498	U64582
D3[a]	2004	MVi/Illinois.USA/0.89/1	M81895	U01977
D4	ongoing	MVi/Montreal.CAN/0.89	AF079554	U01976
D5	2009	MVi/Palau.PLW/0.93	L46757	L46758
		MVi/Bangkok.THA/12.93/1	AF009575	AF079555
D6	2007	MVi/New Jersey.USA/0.94/1	L46749	L46750
D7	2007	MVi/Victoria.AUS/16.85	AF247202	AF243450
		MVi/Illinois.USA/50.99	AY043461	AY037020
D8	ongoing	MVi/Manchester.GBR/30.94	U29285	AF280803
D9	ongoing	MVi/Victoria.AUS/12.99	AY127853	AF481485
D10	2005	MVi/Kampala.UGA/51.00/1	AY923213	AY923185
D11	2010	MVi/Menglian.Yunnan.CHN/47.09	GU440576	GU440571
E[a]	1987	MVi/Goettingen.DEU/0.71	Z80797	X84879
F[a]	1994	MVs/Madrid.ESP/0.94 [SSPE]	Z80830	X84865
G1[a]	1983	MVi/Berkeley.USA/0.83	AF079553	U01974
G2[a]	2004	MVi/Amsterdam.NLD/49.97	AF171231	AF171232
G3	2014	MVi/Gresik.IDN/18.02	AY184218	AY184217
H1	ongoing	MVi/Hunan.CHN/0.93/7	AF045201	AF045212
H2[a]	2003	MVi/Beijing.CHN/0.94/1	AF045203	AF045217

* Adapted from Table 1 WER [19]. Based on search of the MeaNS database on Jan 4, 2018. Excluding vaccine-derived strains, and sequences derived from patients with SSPE. 'Ongoing' genotypes have been detected in the preceding 12 months.

[a] Considered to be inactive; no reports in MeaNS for the last 10 years.

The convention for naming measles sequences representative of wild-type measles viruses has remained unchanged since 1998. The source of the RNA used to sequence the strain is specified in the prefix of the strain name (also known as the WHO name). Viral RNA isolated from virus lysates prepared from propagation of virus in cell culture (isolates) is given the prefix MVi and RNA extracted directly from clinical material has the prefix MVs. The city (or state/province) and the country, indicated by the 3-letter ISO code where the case occurred and the week and the year when onset of rash occurred, follow the prefix. Acceptable substitutes for the week (in order of preference) if rash onset is unknown are (1), the week when the specimen was collected and (2), the week that the specimen arrived at the laboratory.

> **Epidemiologic week.** The weeks are entered as the epidemiological week (numerals 1-52) followed by a period and the year (2 digits). The epidemiologic week should be calculated from the first Monday of each week, with week 1 on or after the first Monday of the year. For example, January 1, 2012 (a Sunday) would be 52.11 and January 2, 2012 (a Monday) would be epidemiologic week 1 for 2012 (1.12). For historic samples where only the year is known, zero (0) should be used as the epidemiologic week.

If only the month and year are known, then the epidemiologic week number corresponding to the second full week in that month is used. Additional strains of the same genotype that are reported in the same epidemiologic week with the same prefix and location are distinguished by adding a slash and numeral following the 2-digit year.

The genotype determined for the measles virus can be placed in square brackets after the genotype designation but is not required. Measles sequences derived from cases with measles inclusion-body encephalitis or subacute sclerosing panencephalitis are denoted by adding (MIBE) and (SSPE), respectively, to the WHO name. Examples of the format for measles strain names are given below:

> MVi/London.GBR/3.12/2 [D4]
> MVs/NewYork.USA/17.11 [G3] (SSPE)

MVi/Maryland.USA/0.77

Measles strains within a genotype that meet criteria for epidemiologic relevance are designated as named strains (see section 7.2.1). The named strains provide a frame of reference for describing circulating viruses that are generally given this designation due to repeated identification of the strain in outbreaks of measles in several locations (i.e., epidemiologically significant). The named strain designation was introduced in the WER update in 2015 [19]. The nomenclature for named strains does not differ from standard genotype nomenclature.

7.2 Integration of measles molecular and epidemiological data

The virologic surveillance for measles virus has expanded through the efforts of the GMRLN laboratories and the sequence data has been made available in the MeaNS database (http://www.who-measles.org). The aggregate data can be accessed to produce a global snapshot of the molecular epidemiology of measles. However, gaps in virologic surveillance, particularly in the African Region and the South-East Asia Region, leave open the possibility of the detection of a novel genotype or a genotype that had been considered inactive or even extinct.

It should be emphasized that the accurate interpretation of data obtained from molecular surveillance of measles is dependent on the availability of epidemiologic data for individual cases and outbreaks. By integration of the measles molecular surveillance data with corresponding epidemiologic information, it has been possible to track global patterns of circulating genotypes, and to document the progress of programmes to eliminate transmission of endemic virus [1-12].

However, the progress that has been made to decrease global transmission of measles virus has not only reduced the number and diversity of circulating genotypes but also the diversity within genotypes. The phylogenetic analysis conducted using the N-450 region may be insufficient to discern between very similar, but epidemiologically unrelated, viruses within a genotype. In situations that could benefit from the ability to discriminate between closely related viruses within a genotype, particularly in elimination settings, sequencing of additional regions of the

viral genome may be required.

If, for example, a prolonged outbreak results from an imported case of measles, it may be difficult to demonstrate that a newly detected cluster of epidemiologically unrelated cases with the same genotype was the result of a new introduction of virus (repeated importation) from the same endemic source. The alternative hypothesis, that missed cases associated with the original outbreak were the source for the newly detected cases, would be difficult to rule out. Sequences that reveal genetic variation that supports the existence of separate transmission pathways among the similar viruses within a genotype would be required to distinguish between the two possible scenarios.

There is evidence to suggest that greater heterogeneity exists among virus strains within a genotype in settings with long-term endemic circulation compared to that associated with viruses from an outbreak that resulted from a single importation [23-27]. It is anticipated that the presence or absence of significant nucleotide variation observed by sequencing additional regions of the measles genome can help elucidate whether separate transmission pathways exist among contemporaneous outbreaks. Investigations are underway to optimize the sequencing methodologies for extended sequencing to provide increased resolution in genetic analyses. The methods and prospects for using extended window sequencing and whole genome sequencing to discriminate among genetically similar strains within a genotype for epidemiologic applications are described in section 7.11.

7.2.1 The named strain designation for circulating variants within a genotype

The 2015 WER update formalized the designation of "named strains" that describe phylogenetically similar strains observed within a genotype. The concept of a named strain provides a convenient means to identify sequences within a genotype that represent an epidemiologically significant viral lineage. A named strain offers an additional frame of reference within the recognized genotypes to describe and track lineages. The virus strain that is eligible for designation as a named strain must have been identified in numerous outbreaks in several countries over a 2-year period. A named strain can be proposed by any of the authorized MeaNS submitters but the sequences for the proposed named strains must be in the public

database in MeaNS and be available through GenBank.

The name of the genotype/virus sequence that is used to represent the named strain is derived from the earliest strain within the lineage that is available through GenBank. However, there is no epidemiologic significance implied for the representative strain. In addition, the source or location associated with the WHO name for the earliest strain is incidental and does not define the source for the lineage. As of July 2017, there were 7 named strains in genotype B3, 11 in genotype D4, 11 are genotype D8 viruses, 2 in genotype D9, and 6 named strains have been designated in genotype H1. For a current listing of named strains, refer to the information available in MeaNS.

7.3 Overview and methods for determination of measles genotypes

The minimum sequence region or 'window' required for determination of measles genotypes is the carboxy-terminal 450 nucleotides that code for the nucleoprotein (N-450). This region of the measles genome was selected because it was determined to be the most variable coding region. The protocol used to amplify the N-450 region and sequencing primers is available in Annex 7.1.

In addition to the N-450 window, the complete protein coding sequence of the N-gene as well as the full-length hemagglutinin gene (H-gene) can be useful to provide increased discrimination between sequences within a genotype. Sequencing of the H-gene (Annex 7.2) and full-length N gene (Annex 7.3) is required when a new genotype is proposed and is recommended for strains that are associated with a large outbreak. However, the initial efforts should be directed toward completion of the N-450 region of the measles genome prior to the pursuit of additional sequence data.

Following generation of the N-450 sequence, the measles genotype is determined by comparing the virus sequence with the reference sequences for the recognized genotypes that are available from the MeaNS website. Comparisons are usually made by phylogenetic analysis although a genotype can be assigned by sequence similarity with one of the reference strains designated for the genotype or a named strain within the genotype (as implemented within MeaNS).

Appropriate phylogenetic analyses can be accomplished by using bootstrapped neighbour-joining or maximum likelihood phylogenetic methods using a variety of programs (e.g. Mega version 7, http://megasoftware.net/).

7.4 Guidelines for reporting a new genotype of measles and use of data from MeaNS

The ongoing global efforts to eliminate measles together with expanded virus surveillance suggest that few new measles genotypes remain to be identified. Rather, it is expected that most sequence variants would fall within existing phylogenetic lineages. Therefore, designation of a new measles genotype must be based upon phylogenetic analysis of the complete measles sequence dataset and meet all of the following criteria:

- Sequences must be obtained from at least N-450 and the entire protein coding region of the H gene
- The new genotype must be based on sequences obtained from multiple cases and at least one viral isolate is available as a reference strain
- The genotype designation must be epidemiologically useful in that it can facilitate the identification of a source of infection, transmission pathway or characterize endemic transmission in a region or country
- Phylogenetic analyses must be performed using all of the available sequence data for N-450 and H such that the diversity within existing genotypes is captured, rather than phylogenies using reference sequences only
- The ancestral node of the putative genotype should fall within the range of genotype ancestral nodes defined by other genotypes within the same clade. The putative genotype must not form a cluster that has its ancestral node placed within sequences from an existing genotype
- The branch defining the putative genotype must be supported by high bootstrap values (>90%). Tree topologies generated by both N-450 and H phylogenies should be broadly concordant

All supporting evidence for a putative new genotype should be shared with WHO and the MeaNS and RubeNS Steering Committee*.

**The MeaNS and RubeNS Steering Committee (SC) provides overall governance of MeaNS and RubeNS. The SC works with WHO and the technical team at Public Health England to discuss methods to improve the functionality and accessibility of MeaNS and RubeNS. The SC provides input into the terms and conditions for access to MeaNS and RubeNS and controls access to the data. The SC members are from WHO GSLs and RRLs, representing each WHO region.*

Use of data in MeaNS for publications

Any publications using data provided from MeaNS should cite the database as directed on the website. In addition, apart from sequences that are in the public domain (i.e., submitted to GenBank) permission should be sought from the individual submitters of all the sequences and strains that are referred to in the planned publication. The following text box provides additional guidelines for compliance with this important aspect of utilizing genetic data.

The use of data derived from MeaNS in a published article must comply with stipulations regarding the use of that data. Failure to do so may result in loss of access to the MeaNS database. The information below addresses three similar, yet separate circumstances, that may require one or more actions prior to submission of the article for publication.

- **Reference is made only to sequence data from your laboratory in your draft article for publication**. MeaNS should be cited in both the text and the reference section of your paper using the citation on the home page: http://www.who-measles.org and thereafter. It is also advisable to contact the WHO regional laboratory coordinator to ensure that the data usage is appropriate.

  For example, MeaNS may be cited using the following publication:
  Rota PA et al. Global distribution of measles genotypes and measles molecular epidemiology. J. Infect. Dis. 2011 Jul;204 Suppl 1:S514-23.

- **In the draft article for publication, it is stated that your sequences are identical to a WHO 'named strain'**. In addition to the instructions above, the GenBank accession number assigned to that named strain must be included.

- **In the draft article for publication, there is a comparison or conclusions made regarding sequences reported by other submitters /countries to MeaNS**. In addition to the first requirement (always) and the second instance (when appropriate), the authorization of MeaNS submitters of sequences must be acquired prior to publication when their sequences are compared. Such authorization should be documented in the acknowledgment section of the paper. For example, if a statement such as, "identical sequences have been reported by X, Y, and Z authors /countries" is in the manuscript, you must get authorization of X, Y, and Z submitters prior to submission of the draft and acknowledge them (with reference to "submitters X, Y, Z") in the Acknowledgment section.

7.5 The measles nucleotide surveillance (MeaNS) database

MeaNS (http://www.who-measles.org/) was initiated as a joint project between the Health Protection Agency, London, United Kingdom (now Public Health England, PHE) and WHO. This database is comprised of measles sequences and associated data that are submitted primarily by the GMRLN or downloaded from GenBank. Most of the measles sequences that are submitted to MeaNS are N-450 sequences, which is the recommended minimum sequence window for genotyping measles viruses.

Initial navigation to the MeaNS website will display options that are available without logging in to the site. The login box and registration button are located on the top-right of the site and various information and aggregate data pages are displayed in a menu bar near the top of the page. Access to the data in MeaNS requires that users register with the website, whereupon registrants are promoted to users by system administrators. The default (public) level of access allows users to view data in MeaNS that is available in GenBank. GMRLN staff are granted submitter access which enables users to submit sequences and view sequences submitted by other users (non-public data). All users are required to acknowledge that they have read and

agreed to the terms and conditions of the site on initial login.

7.5.1 Submission of sequences to MeaNS

Submission of measles virus sequences to MeaNS can be completed through a simple web interface and requires an appropriate region of sequence and a minimal epidemiological dataset. Users with submitter access can submit wild-type measles sequences to the MeaNS database and are encouraged to submit all sequences rather than a representative subset. Users should not submit sequences from a case that had a rash due to a vaccine reaction (sequence identified as clade A, vaccine strain).

While most of the submissions consist of N-450 sequences, the sequences that encode the complete hemagglutinin gene (H-gene) and the complete N-gene are also commonly submitted. MeaNS is designed to accept coding sequence from any complete gene, as well as whole genomes. MVi and MVs sequences from different cases should have different isolate/sequence numbers in instances in which the date and location are identical. It is recommended that only one sequence for N and/or H genes be reported to the databases from each case even if sequence data are available from multiple specimens. Each sequence submitted is assigned a distinct sequence ID in MeaNS in addition to the WHO name.

The information required to generate a standard WHO name and additional epidemiologic data are entered into appropriate fields for each sequence. Guidelines and detailed instructions for entering the data for submission of sequences (new record) to MeaNS are available on the website home page in the form of a training video and FAQs and are included in Annex 7.4, Instructions for entering data and navigating MeaNS for sequence submission.

7.5.2 Sequence data and analysis tools available in MeaNS

Upon login to MeaNS, the menu bar expands to include additional tabs. In addition to submission of sequences (NEW RECORD), clicking the DATA tab reveals three other additional menu choices on tabs entitled LIST, SEARCH, and NAMED STRAINS.

The LIST tab presents users with a list of records stored within MeaNS, the top portion of this page contains a series of filters (geographical, temporal, genotype) that can be completed to narrow the selection of sequences that are displayed. The data fields in the listed table are sortable by clicking the table headers. The WHO name for each record serves as a hyperlink to the individual record entry.

The SEARCH tab presents the user with a series of labelled text boxes that allow users to enter fields that can be used to search for specific records within the database. It is worth noting that there is no fuzzy or wild-card matching and search fields must be entered as they are stored within the database. A successful search will return a single MeaNS sample record and all of the sequences associated with that sample.

The NAMED STRAINS tab links to a page that details WHO reference and named strains for each genotype. By clicking the menu tab, SEQUENCE ANALYSIS, several tools and options are available:

1. **SEQUENCE MATCHING tool**. The first option is to search for exact match. By selecting this option, a search is initiated for sequences identical to the input sequence. The search results are displayed in a table showing the WHO name and associated information about identical sequences. This list can be sorted, filtered and downloaded from the website. If the selected sequence matches a named strain, that record will be displayed with a red background at the top of the matching sequences list. A match with a WHO reference sequence will be shown in green at the top of the matching sequences list. Alternatively, users can also use the BLAST algorithm to search the MeaNS database for identical and similar sequences.

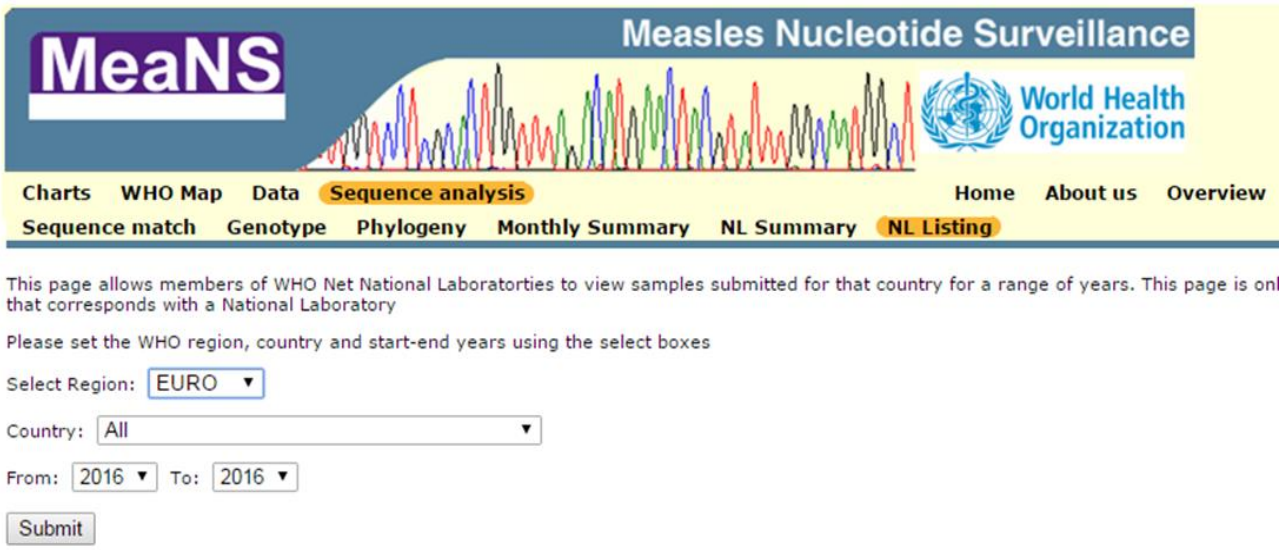
2. GENOTYPING tool. Users can check the genotype of their sequence by running the genotype tool within MeaNS. This option is only available for N-450 or H sequence. Genotype is assigned based on sequence similarity to the closest WHO reference sequence.

3. PHYLO tool. This option plots a phylogeny (tree) of the displayed sequence, the WHO reference set and sequences designated as named strains. The Sequence Analysis tab allows the

user to obtain aggregate data on submitted genotypes in different formats. The monthly summary tab provides genotype data on a monthly basis by year and WHO region chosen by the user. The remaining 2 options are only available to National Laboratories to summarize country data and for generation of annual reports.

4. NL LISTING tool. This tool allows producing annual tables of all sequences reported to MeaNS for a given country (see Figure 7.2 below). The output is a line listing of comprehensive information on genotypes, named strains, distinct sequence IDs provided in a useful format for the regional verification process, when applicable.

Figure 7.2. Screenshot for MeaNS, showing search fields for sequences by country



7.5.3 Sequence data obtained on MeaNS and determination of the source of virus

The analysis of sequence data available from the MeaNS database can be useful to describe the molecular epidemiology of circulating measles viruses, but it is important to understand the limitations that exist for inferring an epidemiologic relationship (e.g., a common source or transmission chain) based on identity of N-450 sequences. Some virus strains circulate over a wide geographic area, which limits the conclusions that can be drawn regarding the source of

virus. In addition, the travel history (origin or destination) of a case may be coincidental, as the exposure may have occurred while in transit.

If the travel history for a case suggests an importation from a specific country but a search in MeaNS reveals no similar sequences for that country, or the timeframe for the circulation of the matching sequences located in MeaNS makes the possibility of co-circulation unlikely, it may be advisable to contact the country laboratory to inquire whether additional information on the status of measles in their country is available to substantiate the possible importation.

Therefore, the source of infection for a case of measles may be classified as unknown even if *a genotype is identified* and similar sequences have been identified in MeaNS [1]. For the source of a case to be designated as either imported or importation-related, the investigation must identify an epidemiologic link to an imported case or association with a contemporaneous outbreak traced to an imported case (see text box below).

Categories for designation of the source for confirmed measles or rubella cases
(adapted from Table 1, *Framework for verifying elimination of measles and rubella*, Weekly Epidemiological Record, 2013, 88: 89-100)

Imported measles, or rubella, case
A case exposed to measles, or rubella, outside the region or country during the 7–21 days (12–23 days for rubella) prior to rash onset and supported by epidemiological or virologic evidence, or both
Note: For cases that were outside the region or country for only a part of the 7–21-day interval (12–23-day interval for rubella) prior to rash onset, additional evidence including a thorough investigation of contacts of the case is needed to exclude a local source of infection.

Importation-related measles, or rubella case
A locally acquired infection occurring as part of a chain of transmission originating from an imported case as supported by epidemiologic or virologic evidence, or both
Note: If transmission of measles, or rubella, from cases related to importation persists for ≥12 months, cases are no longer considered import-related but endemic.

Unknown source measles, or rubella, case
A confirmed case for which an epidemiological or virologic link to importation or to endemic transmission cannot be established after a thorough investigation.

However, demonstration of a genetic match to the outbreak and possible exposure through travel or work-related exposure may support an import-related classification.

A large number of virus sequences generated by a database search by country does not necessarily reflect the size of the outbreak or indicate high levels of measles transmission since some countries may report sequences from all cases from smaller outbreaks, while only a small proportion of sequences may be reported from a large, focalized outbreak. Although the recommendation is that all measles sequence information should be submitted to MeaNS, the actual practices depend on the approach to conducting virologic surveillance, the extent of measles activity and epidemiologic setting, and the capacity for collecting and processing specimens that exist in different countries and different laboratories.

There is a similar limitation due to timeliness of reporting, since there have been substantial delays (up to a year or more) in reporting sequences from some countries. Therefore, the data within MeaNS is incomplete and biased towards those countries with fewer/smaller outbreaks, laboratories that perform more intensive (comprehensive) sequencing, and those that report sequences in a timely manner.

Part B. Rubella

7.6 Phylogenetic diversity and nomenclature for rubella genotypes

In 2005, a standardized nomenclature for wild-type rubella viruses was established following a meeting of experts in 2005 [20]. This meeting and the subsequent publication also established the sequencing window of the envelope glycoprotein E1 coding region (739 nucleotides, E-739) for determination of rubella genotypes. Updates to the nomenclature were reported in 2007 and 2013 [21, 22]. The original two clades remain the major phylogenetic groups of rubella viruses. The clades differ by 8-10% (E-739) and are represented by numerals (1, 2) which precede the letter assigned to genotypes within the two clades. Uppercase letters are used for recognized genotypes of rubella that meet the specified criteria (section 7.9). Other genotypes that have been

identified but lack well-characterized progenitor wild-type viruses are designated with a lowercase letter.

The extent of genetic information available for rubella genotypes is limited compared to what has been obtained for measles. Many countries have not made significant progress in describing circulating rubella viruses. As of July 2017, there are 13 rubella genotypes. There are 10 genotypes within Clade 1(1a, 1B-1J) and Clade 2 consists of three genotypes (2A, 2B, 2C). Clade 1 includes the provisional genotype 1a. Genotype 1a includes vaccine strains from the 1960s. However, not all of the strains included in genotype 1a cluster together, making the phylogeny of this group unclear. Whole genome analysis of 30 rubella sequences from 8 of the 13 known genotypes generated a maximum likelihood phylogenetic tree with high bootstrap values for all genotypes except the provisional genotype 1a.

Provisional genotype 1a is highly diverse and was designated as a provisional genotype in part because of the historical importance of this group of viruses (for example, the RA27/3 vaccine strain is included among the 1a viruses). Vaccine viruses are not wild-type viruses but are included as reference sequences to allow vaccine viruses to be rapidly identified. Few viruses in this provisional genotype have been found since 2004, and genotype 1a remains provisional. It is important to note that the four reference viruses in this genotype often cluster in multiple groups, but the groups are distinct from other genotypes.

Clade 2 includes genotypes 2A, 2B, and 2C. Genotype 2A includes vaccine strains manufactured in China. All of the rubella genotypes have at least two reference strains. Genotypes 1B, 1C, 1G, and 2B have three reference strains and the provisional genotype 1a has four reference strains. Only four of the genotypes (1E, 1G, 1J, 2B) are commonly detected and reported. Of these four active genotypes, 1E and 2B are the most frequently detected and have a wide geographic distribution.

Four of the rubella genotypes, 1D, 1F, 1I, and 2A, have not been reported in circulation for over 10 years: the last report of 1D was in 1996, 1F was reported in 2002, 1I was reported in 1994 and wild-type 2A was reported in 1980 (although 2A vaccine-derived viruses have been reported

more recently). Thus, these four genotypes are considered inactive and are probably extinct. However, considerable gaps in molecular surveillance still exist globally.

The nomenclature for rubella genotypes is similar to that for measles genotypes [20-22]. Rubella virus RNA sequences can be generated from either viral isolates or RNA extracted directly from clinical material. Sequences are designated as either:

RVi – sequence derived from RNA extracted from a rubella virus isolate
RVs – sequence derived from RNA extracted directly from clinical material

The geographic association, dates, and numbering of additional isolates from the same location, year and epidemiologic week follow the measles nomenclature. For example, RVs/Hong Kong.CHN/20.12/2[1E] designates a genotype 1E rubella virus derived directly from a clinical sample collected in Hong Kong, China, in the 20th week of 2012. This is the second sequence reported from the same week and location. For WHO names of rubella genotypes, a slash is used after the year, but not after the sequence number if there is more than 1 sequence from the same location and week.

There are special designations for sequences derived from cases with congenital rubella syndrome (CRS) and from newborns with congenital rubella infection (CRI). In both instances the geographic location specified in the name is the place of birth and the onset date of disease is the date of birth. For example, RVi/Ho Chi Minh.VNM/41.11/[2B] (CRS) designates a genotype 2B rubella virus isolate from a CRS patient in Ho Chi Minh, Vietnam, in the 41st week of 2011.

Although reporting of rubella vaccine strains to RubeNS is not encouraged, reference strains from vaccines are designated by adding “VAC” to their names. For example, the BRD2 vaccine strain is written as RVi/Beijing.CHN/80[2A]VAC.

7.7 Integration of rubella molecular and epidemiological data

Epidemiological data should accompany viral surveillance specimens in order to increase the usefulness of the molecular information. If a virus can be identified as an importation using

standard epidemiological means and the country of origin is known, this information should be provided. The information for the corresponding viral sequence must be sufficient to provide the WHO name for the sequence.

The lack of comprehensive rubella surveillance in many parts of the world has resulted in a paucity of sequences available from many areas and an absence of representative rubella sequences from other areas. Consequently, the current database does not contain sufficient geographic diversity for a comprehensive depiction of global rubella genotype distribution. Continued efforts to strengthen rubella virologic surveillance are needed, especially in those countries where little or no virologic surveillance is carried out.

Even as virologic surveillance for rubella increases and more genetic data become available, the progress towards control and elimination of rubella will likely reduce the diversity of circulating rubella genotypes. Similar to the situation with measles, analyses of circulating viruses will require sequencing methodologies that can discriminate among closely related viruses within a genotype in order to track rubella viruses of genotypes 1E, 1G, and 2B. Extended window sequencing and next-generation sequencing are covered in section 7.11.

7.8 Overview and methods for determination of rubella genotypes

Most of the genetic studies of wild-type rubella viruses have been conducted by sequencing the full length or portions of the coding region of the E1 envelope protein. Sequencing other regions of the rubella virus genome for molecular epidemiology has been considered, but no compelling arguments for altering the target from the E1 region of the rubella genome have been forthcoming. However, different regions (windows) within the E1 coding region have been utilized for genetic characterization. Following an evaluation of the alternative windows within the E1 region, the currently recommended target region consists of 739 nucleotides (E-739) from 8731-9469. The CDC protocol for amplification of two overlapping fragments that can be used to sequence the E1 region for sequencing rubella viruses is provided in Annex 7.5.

7.9 Guidelines for reporting a new rubella genotype

The following criteria have been developed to standardize the requirements for a recognized genotype of rubella virus:

a. Two reference viruses in the genotype are available and have been submitted to a WHO recognized strain bank.

b. Sequences are available from the entire structural protein open reading frame (SPORF) (coding region for C, E2 and E1 proteins) of the two reference viruses in the genotype.

c. Phylogenetic analyses of the reference viruses in the genotype with the reference viruses from all other genotypes show that:

i. the reference viruses from the genotype cluster together with one another and separately from the reference viruses of all other genotypes with high confidence values (e.g. 80-90% bootstrap values). This clustering should be obtained with at least two different analysis methods;

ii. the phylogenetic tree topology obtained with full SP ORF sequences of the reference viruses of the genotype and other reference viruses should be the same when the coding sequences of C, E2 and E1 are analysed individually;

iii. the intragenotype and intergenotype distances for the reference viruses in the genotype are consistent with those for the existing genotypes; and

iv. the branching pattern obtained in the phylogenetic analysis versus other reference viruses is consistent with a new genotype.

d. Sequence data for reference viruses in the genotype have been submitted to GenBank with the virus name and associated epidemiological and clinical data

7.10 The rubella nucleotide surveillance (RubeNS) database

The rubella sequence database (www.who-rubella.org) is operated, as with MeaNS, through a cooperative activity under technical leadership from Public Health England (PHE), with a steering committee comprised of representatives from the Global Specialized and Regional Reference Laboratories. The corresponding database and web-application are maintained by PHE and sequence data are contributed by GMRLN member laboratories or downloaded from GenBank.

The primary objective of the RubeNS database is to collect sequences from rubella genotypes consisting of the minimum fragment required for genotyping (E-739). However, the complete E1 coding region, the complete structural polyprotein, or the complete genome are increasingly useful for comparison with other sequences collected globally. Bioinformatics tools in RubeNS allow users to find identical or similar sequences and identify a preliminary genotype designation.

7.10.1 Submission of sequences to the RubeNS database

The submission of rubella virus sequences to RubeNS, like MeaNS, is through a web interface. Although the website looks very different from MeaNS, it has the same basic functionality. Users need to register to use the website, which is only available to GMRLN members. As with MeaNS, users are encouraged to submit all rubella sequences, rather than what may be considered to provide a representative subset, and only wild-type, not vaccine, sequences should be submitted.

Rubella sequences of any length will be accepted into the RubeNS database, but analysis and genotype designation can only be carried out on the E-739 sequence or the complete E1 sequence. As is true for MeaNS, information can be submitted to RubeNS to generate a standard WHO name, or the user can supply the WHO name directly. Additional information, including more precise geographic information can also be added. Guidelines and details of the steps for entering data into the appropriate fields for submission of sequences (new record) to RubeNS are available in Annex 7.6.

7.10.2 Sequence data and analysis tools available in RubeNS

Upon login to RubeNS, the HOME page is displayed with information about the database and a series of tabs are available. In addition to the HOME page, USER EDIT allows users to change their details, including their password. The WHO Map displays the most recent WHO global map with rubella genotypes, SAMPLES and TOOLS are the main functional tabs in the database

and TANDC shows the Terms and Conditions of using the site.

The SAMPLES tab displays a list of all the samples within the database below a series of fields, many with drop down boxes. These fields can be filled in as needed to create a set of filters to narrow down the search. All the fields that allow free text can accept wild-card characters, i.e. the asterisk (*) or question mark (?), which is useful if the exact spelling is not known.

For each sequence that satisfies the search criteria, the genotype, country, date of submission and region of the gene are displayed, along with other basic information. By clicking on the WHO name, the sample record is displayed, including the sequence submitted. The tools for analysis of the sequence are provided along the foot of the page: EXACT MATCHING, BLAST MATCHING, GENOTYPING, PHYLOGENY and SUBMIT TO GENBANK. Similar tools can be found by clicking on the TOOLS tab, but that approach required that the sequence for analysis would have to be copied into the PASTE SEQUENCE box.

1. EXACT MATCHING tool. By selecting this option, a search is initiated for sequences identical to the input sequence. The search results are displayed in a table similar to the Samples listing. The results of the search can be downloaded for further analysis in Excel.

2. BLAST matching tool. As the RubeNS database is much smaller than MeaNS, and rubella virus sequences are more variable, the **Exact match** tool will often not find identical sequences and the BLAST matching tool may be more informative. Using the BLAST matching tool will search the RubeNS database and identify identical and similar sequences, with the percent identity and length of overlap provided.

3. GENOTYPING tool. The genotyping tool uses position-specific scoring matrices (PSSM) to not only provide the genotype corresponding to the input rubella sequence but also to give the confidence of the phylogenetic analysis which generated the genotype. Almost all currently circulating rubella sequences will be assigned a genotype with high confidence (Z score >3). This is not true of some of the historic rubella virus sequences which may require analysis with all the reference genotypes using a programme that uses Bayesian inference (e.g., MrBayes).

4. PHYLOGENY tool. This option plots a phylogeny of the displayed sequence (labelled input)

and the sequences of the WHO reference strains. This tool can also be used to confirm the genotyping prediction using the genotyping tool.

7.10.3 Interpretation of information from RubeNS analyses

As noted above, the RubeNS database contains far fewer sequences compared to MeaNS. There are many countries, including those with endemic rubella, for which there are no representative rubella sequences in RubeNS. Therefore, the use of the database for understanding the context of a particular sequence in terms of its association with geographic or epidemiologic significance, is challenging. Sequences with the same E-739 or complete E1 gene are rarely identified unless they are part of the same outbreak. However, identifying similar sequences may provide some insights into potential sources of an importation.

For rubella, the important aspect of molecular analysis is not so much for the purpose of identification of a particular genotype as much as it is to provide the sequences that can then be compared to other rubella sequences belonging to the same genotype in the database. Demonstration that sequences differ between or among some cases may also be useful in settings where separate introductions of virus may produce contemporaneous, but unrelated, outbreaks.

7.11 Methods and prospects for enhancing resolution of sequence data for molecular epidemiology

Sequence analysis of larger regions of the measles or rubella genome (extended sequencing), including whole genome sequencing (WGS) provides higher resolution for characterizing viral transmission pathways [23-27]. This increased resolution is needed to distinguish between local, ongoing transmission from repeated introductions of the same genotype in elimination settings.

The interpretation of the data that is generated by the extended sequencing requires an understanding of the expected rate of nucleotide substitutions from epidemiologically-linked viruses. The estimated rate would ideally be based upon studies of viruses collected from a range of settings, such as household transmission cases to outbreaks from an imported virus that

continue for several generations. These studies are necessary for estimating a threshold of sequence variability that would predict independent chains of transmission between similar strains, or conversely, would support the existence of epidemiologic links between outbreaks or clusters of cases in situations where the investigation failed to identify links [23-25]. It is important to note that the approach is to provide data that would be probabilistic and not definitive.

Both WGS and extended sequencing of hypervariable regions, such as the intergenic, or noncoding, region between the M and F genes (MF-NCR) of measles have been examined [27]. Sequencing of the MF-NCR can be achieved using standard Sanger sequencing methods, while practical application of WGS will require use of next generation sequencing methods. It is possible to perform WGS from viral RNA extracted directly from patient specimens, but greater success would be achieved by sequencing virus isolates.

The variability observed in the MF-NCR is roughly equivalent to the diversity present in the corresponding whole measles genome sequence [27]. Therefore, separate transmission pathways may be apparent by analysis of sequences of the MF-NCR among measles viruses that have identical N-450 sequences. However, more studies are needed to determine the utility of the extended sequencing and to establish the minimum number of nucleotide differences that can differentiate between separate chains of transmission.

The goals for the Next generation, Extended window and Whole genome sequencing working group (N.E.W.) of the GMRLN include standardization of the methods for obtaining sequences for the full genome of measles and rubella viruses and the MF-NCR region of measles for the routine application in the GMRLN [27]. The MeaNS and RubeNS databases have been modified for deposit of the appropriate extended sequencing or WGS and to allow a search for these sequences. The database for WGS will be continually updated as additional sequences become available for the whole genome for the genotypes and named strains of measles and rubella.

Bibliography to Chapter 7

1. Oliveira ML et al. Genetic homogeneity of measles viruses associated with a measles outbreak, São Paulo, Brazil, 1997. Emerging Infectious Diseases, 2002, 8:808-813.

2. Shulga SV et al. Genetic variability of wild-type measles viruses, circulating in the Russian Federation during the implementation of the National Measles Elimination Program, 2003-2007. Clinical Microbiology and Infection, 2009, 15:528-537.

3. Rota PA et al. Molecular epidemiology of measles virus. Current Topics in Microbiology and Immunology, 2009, 330:129-50.

4. Rota PA et al. Global distribution of measles genotypes and measles molecular epidemiology. Journal of Infectious Diseases, 2011, 204 Suppl 1:S514-23.

5. Rota PA et al. Improving global virologic surveillance for measles and rubella. Journal of Infectious Diseases, 2011, 204 Suppl 1:S506-13.

6. Vainio K et al. Measles virus genotyping an important tool in measles outbreak investigation in Norway, 2011. Eurosurveillance, 2012, 17(50): pii=20340. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20340

7. Pattamadilok S et al. Genetic characterization of measles viruses that circulated in Thailand from 1998 to 2008. Journal of Medical Virology, 2012, 84(5): 804-813.

8. Zhang Y et al. Monitoring progress toward measles elimination by genetic diversity analysis of measles viruses in China 2009-2010. Clinical Microbiology and Infection, 2014, 20(9): O566-577.

9. Cheng WY et al. Measles surveillance in Taiwan, 2012-2014: Changing epidemiology, immune response, and circulating genotypes. Journal of Medical Virology, 2016, 88(5):746-753.

10. Furuse Y, Oshitani H. Global transmission dynamics of measles in the measles elimination era. Viruses, 2017, 9,82: doi:10.3390/v9040082

11. Santibanez S et al. Molecular surveillance of measles and rubella in the WHO European Region: new challenges in the elimination phase. Clinical Microbiology and Infection, 2017, 23(8): 516-523.

12. Vaidya SR, Chowdhury DT. Measles virus genotypes circulating in India, 2011-2015. Journal of Medical Virology, 2017, 89(5): 753-758.

13. Expanded Programme on Immunization (EPI). Standardization of the nomenclature for describing the genetic characteristics of wild-type measles viruses. Weekly Epidemiological Record, 1998, 73(35):265-9.

14. WHO. Nomenclature for describing the genetic characteristics of wild-type measles viruses (update). Weekly Epidemiological Record, 2001, 76(33):249-51.

15. WHO. Update of the nomenclature for describing the genetic characteristics of wild-type measles viruses: new genotypes and reference strains. Weekly Epidemiological Record, 2003, 78(27):229-32.

16. WHO. New genotype of measles virus and update on global distribution of measles genotypes. Weekly Epidemiological Record, 2005, 80(40):347-51.

17. WHO. Global distribution of measles and rubella genotypes--update. Weekly Epidemiological Record, 2006, 81(51/52):474-9.

18. WHO. Measles virus nomenclature update: 2012. Weekly Epidemiological Record, 2012, 87(9):73-81.

19. WHO. Genetic diversity of wild-type measles viruses and the global measles nucleotide surveillance database (MeaNS). Weekly Epidemiological Record, 2015, 90(30):373-80.

20. WHO. Standardization of the nomenclature for genetic characteristics of wild-type rubella viruses. Weekly Epidemiological Record, 2005, 80(14):126-32.

21. WHO. Update of standard nomenclature for wild-type rubella viruses, 2007. Weekly Epidemiological Record, 2007, 82(24):216-22.

22. WHO. Rubella virus nomenclature update: 2013. Weekly Epidemiological Record, 2013, 88(32):337-43.

23. Thomas S et al. Measles outbreak with unique virus genotyping, Ontario, Canada, 2015. Emerging Infectious Diseases, 2017, 23(7):1063-1069.

24. Gardy JL et al. Whole-Genome sequencing of measles virus genotypes H1 and D8 during outbreaks of infection following the 2010 Olympic winter games reveals viral transmission routes. Journal of Infectious Diseases, 2015, 212(10):1574-8.

25. Harvala H et al et al. Role of sequencing the measles virus hemagglutinin gene and hypervariable region in the measles outbreak investigations in Sweden during 2013-2014. Journal of Infectious Diseases, 2016, 213(4):592-599.

26. Penedos AR et al. Assessment of the utility of whole genome sequencing of measles virus in the characterisation of outbreaks. PLoS One, 2015, 10(11):e0143081.

27. WHO. The role of extended and whole genome sequencing for tracking transmission of measles and rubella vriuses: report from the Global Measles and Rubella laboratory Network meeting, 2017. Weekly Epidemiological Record, 2018, 93(6):55-59.