

HomeWork-5 Report Template

1. Methodology

a. Obtaining Queries

i. Description of the assigned queries

```
1. {  
2.     'West African Ebola epidemic': 152901,  
3.     'H1N1 Swine Flu pandemic': 152902,  
4.     'COVID 19': 152903  
5. }
```

ii. Explanation of the relevance to your vertical search engine topic

```
1. self.es.search(index="crawler",  
2.                 body={"size": 200,  
3.                       "query": {  
4.                           "multi_match": {"query": query,  
5.                           "fields": ["title", "content"]}  
6.                       },  
7.                       "_source": ""  
8.                       })['hits']['hits']
```

We use the elastic search query to filter matching contents in both the title and content

b. Assessment Graphical Interface

i. Features of the interface

1. Query Display: Shows search queries and corresponding results.
2. Search Results: Displays article details with highlighted query terms.
3. Expander: Allows detailed view expansion for each result.
4. Relevance Grade: Dropdown for assessing document relevance.
5. Submit Button: Saves assessments to a file.
6. Sidebar Stats: Tracks total assessed documents.

ii. Screenshots of the interface

The Ebola outbreak 2013-2016: old lessons for new epidemics

PubMed

Author: Jianqing

An official website of the United States government The .gov means it's official. Federal government websites often end in .gov or .mil. Before sharing sensitive information, make sure you're on a federal government site. The site is secure. The https:// ensures that you are connecting to the official website and that any information you provide is encrypted and transmitted securely.

Ebola virus causes a severe haemorrhagic fever in humans with high case fatality and significant epidemic potential. The 2013-2016 outbreak in West Africa was unprecedented in scale, being larger than all previous outbreaks combined, with 28 646 reported cases and 11 323 reported deaths. It was also unique in its geographical distribution and multicountry spread. It is vital that the lessons learned from the world's largest Ebola outbreak are not lost. This article aims to provide a detailed description of the evolution of the outbreak. We contextualize this outbreak in relation to previous Ebola outbreaks and outline the theories regarding its origins and emergence. The outbreak is described by country, in chronological order, including epidemiological parameters and implementation of outbreak containment strategies. We then summarize the factors that led to rapid and extensive propagation, as well as highlight the key successes, failures and lessons learned from this outbreak and the response. This article is part of the themed issue 'The 2013-2016 West African Ebola epidemic: data, decision-making and disease control'. Keywords

; Ebola outbreak; Ebola virus disease; West Africa. © 2017 The Authors. PubMed Disclaimer A.M.J. is a governor of the Wellcome Trust. The authors have no other conflicts of interest to declare. Figure 1. Timeline of key events with country-specific epidemic curves. Case numbers are total reported. Figure 2. Geographical map of Guinea, Sierra Leone and Liberia showing districts and total number of cases. Figure 3. Initial transmission chain in Guinea. HCW, healthcare worker. (Adapted from Baize et al. Figure 4. Where the outbreak began. Map to show Kissi tribal area spanning Guinea, Sierra Leone. NCBI Literature Resources MeSH PMC Bookshelf Disclaimer The PubMed wordmark and PubMed logo are registered trademarks of the U.S. Department of Health and Human Services (HHS). Unauthorized use of these marks is strictly prohibited. Connect with NLM National Library of Medicine 8600 Rockville Pike Bethesda, MD 20894 Web Policies FOIA HHS Vulnerability Disclosure Help Accessibility Careers

Relevance Grade for Document 27 (Query: West African Ebola epidemic):

Very relevant

Submit for Document 27 (Query: West African Ebola epidemic)

c. Manual Assessments

i. Approach to assessing the documents

1. Read through each document title and content to understand.
2. Determine if the document is very relevant, relevant, or non-relevant to the given query.
3. Select the appropriate relevance grade for each document.
4. Save the assessment with the assessor's name and document details into the qrel file.

ii. Sample assessments with explanations

The Ebola outbreak 2013-2016: old lessons for new epidemics

PubMed

Author: Jianqing

An official website of the United States government The .gov means it's official. Federal government websites often end in .gov or .mil. Before sharing sensitive information, make sure you're on a federal government site. The site is secure. The https:// ensures that you are connecting to the official website and that any information you provide is encrypted and transmitted securely. Ebola virus causes a severe haemorrhagic fever in humans with high case fatality and significant epidemic potential. The 2013-2016 outbreak in West Africa was unprecedented in scale, being larger than all previous outbreaks combined, with 28 646 reported cases and 11 323 reported deaths. It was also unique in its geographical distribution and multicountry spread. It is vital that the lessons learned from the world's largest Ebola outbreak are not lost. This article aims to provide a detailed description of the evolution of the outbreak. We contextualize this outbreak in relation to previous Ebola outbreaks and outline the theories regarding its origins and emergence. The outbreak is described by country, in chronological order, including epidemiological parameters and implementation of outbreak containment strategies. We then summarize the factors that led to rapid and extensive propagation, as well as highlight the key successes, failures and lessons learned from this outbreak and the response. This article is part of the themed issue 'The 2013-2016 West African Ebola epidemic: data, decision-making and disease control'. Keywords

; Ebola outbreak; Ebola virus disease; West Africa. © 2017 The Authors. PubMed Disclaimer A.M.J. is a governor of the Wellcome Trust. The authors have no other conflicts of interest to declare. Figure 1. Timeline of key events with country-specific epidemic curves. Case numbers are total reported cases. Figure 2. Geographical map of Guinea, Sierra Leone and Liberia showing districts and total number of cases. Figure 3. Initial transmission chain in Guinea. HCW, healthcare worker. (Adapted from Baize et al. 2014). Figure 4. Where the outbreak began. Map to show Kissi tribal area spanning Guinea, Sierra Leone. NCBI Literature Resources MeSH PMC Bookshelf Disclaimer The PubMed wordmark and PubMed logo are registered trademarks of the U.S. Department of Health and Human Services (HHS). Unauthorized use of these marks is strictly prohibited. Connect with NLM National Library of Medicine 8600 Rockville Pike Bethesda, MD 20894 Web Policies FOIA HHS Vulnerability Disclosure Help Accessibility Careers

Relevance Grade for Document 27 (Query: West African Ebola epidemic):

Very relevant

Submit for Document 27 (Query: West African Ebola epidemic)

1.

a. Very Relevant as query is both in title and content

Domain: pubmed.ncbi.nlm.nih.gov ----- Article ID: <https://pubmed.ncbi.nlm.nih.gov/27721110>

An updated review of Zika virus PubMed

Author: Jianqing

An official website of the United States government The .gov means it's official. Federal government websites often end in .gov or .mil. Before sharing sensitive information, make sure you're on a federal government site. The site is secure. The https:// ensures that you are connecting to the official website and that any information you provide is encrypted and transmitted securely. The current outbreak of Zika virus (ZIKV) in South America is one of the most serious public health emergencies since the Ebola outbreak of West Africa [2014]. ZIKV belongs to the flaviviridae family and has two lineages (Asian and African). The virus was first discovered in Uganda [1947] and the first human infection was identified in Nigeria [1952]. The current epidemic is the third of its type after that of Yap Island, Micronesia [2007] and French Polynesia [2013]. Phylogenetic studies revealed that the current strain shares about 99.7% of nucleotides and 99.9% of amino acids with the strain of French Polynesia epidemic [2013], suggesting that it has spread across the Pacific Ocean to invade South America. Aedes Aegypti mosquito is the main vector for ZIKV and there are some reports describing possible sexual and maternal to fetal transmission. ZIKV infection is known to be self-limited. However, recent reports suggested that it can be associated with neurological manifestations as Guillan-Barré Syndrome and microcephaly in the newborn population. Currently, vector control seems to be the most effective available preventive measure against ZIKV spread. The development of broad spectrum antivirals and ZIKV vaccines should be a priority of future research.

Keywords

; Microcephaly; Public health; Zika virus. Copyright © 2016 Elsevier B.V. All rights reserved. PubMed Disclaimer NCBI Literature Resources MeSH PMC Bookshelf Disclaimer The PubMed wordmark and PubMed logo are registered trademarks of the U.S. Department of Health and Human Services (HHS). Unauthorized use of these marks is strictly prohibited. Connect with NLM National Library of Medicine 8600 Rockville Pike Bethesda, MD 20894 Web Policies FOIA HHS Vulnerability Disclosure Help Accessibility Careers

Relevance Grade for Document 135 (Query: West African Ebola epidemic):

Non-relevant

Submit for Document 135 (Query: West African Ebola epidemic)

2.

a. Not Relevant as this is about outbreak of another virus

2. Implementation of trec_eval

a. Pseudocode

```
class TrecEval:
    - Initialize with QREL file path, ranked file path, and optional
    parameters.
    - Define methods to read QREL and ranked files, calculate metrics,
    and print results.

    Method read_qrel():
        - Open QREL file.
        - Iterate through each line:
```

```

        - Parse query ID, document ID, and relevance score.
        - Store data in a dictionary.

Method read_trec():
    - Open ranked file.
    - Iterate through each line:
        - Parse query ID, document ID, and score.
        - Store data in a dictionary.

Method calculate_metrics():
    - Call read_qrel() and read_trec().
    - Initialize variables for total number of documents, relevant
documents, and relevant documents retrieved.
    - Iterate through each query:
        - Calculate precision, recall, average precision, precision
at cutoffs, and R-precision.
        - Optionally plot precision-recall curve for each query.
        - Print metrics for each query if specified.
    - Calculate mean average precision and average R-precision.
    - Print the overall metrics including total number of
documents, average precision, precision at cutoffs, and R-precision.

Method compute_ndcg(rank):
    - Call read_trec().
    - Iterate through each query:
        - Calculate nDCG at the specified rank.
        - Print nDCG for each query.

Method print_metrics(query_id=None):
    - If query_id is None, print metrics for all queries.
    - If query_id is specified, print metrics for the given query.

```

b. Code snippets

```

# Initialize TrecEval object
my_trec = TrecEval("qrel.txt", "ranked_list.txt",
print_all_queries=False, plot_graph=False, use_map=True)

# Calculate metrics
my_trec.calculate_metrics()

# Compute nDCG at rank 10
my_trec.compute_ndcg(10)

```

3. Evaluation Results

a. Results of Manual Assessments

i. Compilation of assessments in QREL format

Code		qrel.txt
1	152901	Anson http://www.rappler.com/world/regions/africa/67128-who-un-first-staff-ebola-infection Very relevant
2	152901	Anson https://en.wikipedia.org/wiki/Cultural_effects_of_the_Ebola_crisis Very relevant
3	152901	Anson https://www.theguardian.com/world/2015/nov/20/ebola-case-in-liberia-confirmed-by-who Very relevant
4	152901	Anson https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/cost-of-ebola.html Very relevant
5	152901	Anson http://www.cdc.gov/mmwr/volumes/65/su/su6503a2.htm Very relevant
6	152901	Anson https://www.cdc.gov/mmwr/volumes/65/su/su6503a2.htm Very relevant
7	152901	Anson http://dx.doi.org/10.15585/mmwr.su6503a2 Very relevant
8	152901	Anson https://idpjournal.biomedcentral.com/articles/10.1186/s40249-015-0090-9 Relevant
9	152901	Anson https://news.yahoo.com/ebola-releases-last-68-people-ebola-quarantine-143241412.html Very relevant
10	152901	Anson https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/index.html Very relevant
11	152901	Anson https://www.theguardian.com/world/2014/oct/18/oxfam-call-for-troops-ebola-west-africa-world-health-organisation-accused Very relevant
12	152901	Anson https://en.wikipedia.org/wiki/Responses_to_the_West_African_Ebola_virus_epidemic Very relevant
13	152901	Anson https://en.wikipedia.org/wiki/Responses_to_the_Ebola_virus_epidemic_in_West_Africa Very relevant
14	152901	Anson https://web.archive.org/web/20150418133519/http://www.newvision.co.ug/news/667165-ebola-hit-countries-call-for-8-bn-aid-plan.html Very relevant
15	152901	Anson https://www.mailman.columbia.edu/public-health-now/news/scientists-discover-ebola-virus-west-african-bat Very relevant
16	152901	Anson https://en.wikiquote.org/wiki/Ebola_virus_epidemic_in_West_Africa Very relevant
17	152901	Anson https://pubmed.ncbi.nlm.nih.gov/25308956 Very relevant
18	152901	Anson https://obamawhitehouse.archives.gov/the-press-office/2014/09/16/fact-sheet-us-response-ebola-epidemic-west-africa Very relevant
19	152901	Anson http://www.ncbi.nlm.nih.gov/nuccore/?term=ERX5245591 Relevant
20	152901	Anson http://www.ncbi.nlm.nih.gov/26033924/ Very relevant
21	152901	Anson https://pubmed.ncbi.nlm.nih.gov/26033924/ Very relevant
22	152901	Anson https://web.archive.org/web/20150318230625/http://www.undp.org/content/undp/en/home/presscenter/pressreleases/2015/03/12/west-african-economies-feeling-ripple-effects-of-ebola Relevant
23	152901	Anson http://www.commondreams.org/news/2014/09/03/ebola-outbreak-surges-health-officials-slam-international-coalition-inaction Relevant
24	152901	Anson https://en.wikipedia.org/wiki/User:0z2ie10aaaa Non-relevant
25	152901	Anson https://www.aspenideas.org/articles/we-can-stop-epidemics-before-they-happen-here-s-how Relevant
26	152901	Anson https://news.yahoo.com/thousands-ebola-survivors-face-severe-pain-possible-blindness-085148730.html Very relevant
27	152901	Anson https://pubmed.ncbi.nlm.nih.gov/28396469 Very relevant
28	152901	Anson https://doi.org/10.1056/NEJMe1411471 Very relevant
29	152901	Anson https://pubmed.ncbi.nlm.nih.gov/25254986 Very relevant
30	152901	Anson https://www.cdc.gov/vhf/ebola/history/chronology.html Very relevant
31	152901	Anson https://blogs.scientificamerican.com/guest-blog/after-ebola-a-blueprint-emerges-to-jump-start-r-d/ Relevant
32	152901	Anson https://www.aljazeera.com/news/2021/2/14/ebola-sees-first-ebola-deaths-since-2016 Very relevant
33	152901	Anson https://web.archive.org/web/20161220134725/https://blogs.scientificamerican.com/guest-blog/after-ebola-a-blueprint-emerges-to-jump-start-r-d/ Relevant
34	152901	Anson https://abcnews.go.com/International/ebola-stricken-boy-1st-cross-border-case-growing/story?id=63656559 Very relevant
35	152901	Anson https://en.wikipedia.org/wiki/Ebola-chan Relevant
36	152901	Anson https://doi.org/10.11604/2Fpmj.2020.36.201.24379 Relevant
37	152901	Anson https://www.theguardian.com/world/2015/nov/22/liberia-health-officials-ebola-outbreak-monrovia Very relevant
38	152901	Anson https://www.voanews.com/a/congo-ebola-outbreak-expected-to-end-next-week/4486095.html Very relevant
39	152901	Anson https://ui.adsabs.harvard.edu/abs/2014Sci...345.1369G Very relevant
40	152901	Anson https://www.cdc.gov/vhf/ebola/outbreaks/responder/fighting-ebola-on-familiar-ground.html Very relevant
41	152901	Anson https://pubmed.ncbi.nlm.nih.gov/26748845 Very relevant
42	152901	Anson http://www.voanews.com/content/ebola-hit-sierra-leone-to-reopen-schools-march-30th/2629854.html Very relevant
43	152901	Anson https://www.cdc.gov/vhf/ebola/outbreaks/responder/being-creative-in-goma.html Non-relevant
44	152901	Anson https://doi.org/10.1136/bmi.h376 Very relevant

ii. Discussion on the assessment results

- This suggests effective retrieval of pertinent information from diverse sources, covering various aspects of the epidemic. Overall, the system appears to perform well in providing comprehensive and relevant content on the topic.

b. trec_eval Results

i. Presentation of results from your trec_eval

Queryid (Num): 3

Total number of documents over all queries

Retrieved: 600

Relevant: 566

Rel_ret: 542

Interpolated Recall - Precision Averages:

at 0.00 1.0000

at 0.10 0.9583

at 0.20 0.9504

at 0.30 0.9358

at 0.40 0.9267

at 0.50 0.9140

at 0.60 0.9113

at 0.70 0.9113

at 0.80 0.9113

at 0.90 0.9047

at 1.00 0.0000

Average precision (non-interpolated) for all rel docs(averaged over queries)

0.8745

Precision:

At 5 docs: 0.9333

At 10 docs: 0.9333

At 15 docs: 0.9111

At 20 docs: 0.9167

At 30 docs: 0.9111

At 100 docs: 0.9100

R-Precision (precision after R (= num_rel for a query) docs retrieved):

Exact: 0.9001

Query ID: 152901 nDCG: 1.0000 At rank: 10

Query ID: 152902 nDCG: 1.0000 At rank: 10

Query ID: 152903 nDCG: 0.9667 At rank: 10

ii. **Comparison with provided trec-eval (if applicable)**

Queryid (Num): 3

Total number of documents over all queries

Retrieved: 600

Relevant: 566

Rel_ret: 542

Interpolated Recall - Precision Averages:

at 0.00 1.0000

at 0.10 0.9583

at 0.20 0.9504

at 0.30 0.9358

at 0.40 0.9267

at 0.50 0.9140

at 0.60 0.9113

at 0.70 0.9113

at 0.80 0.9113

at 0.90 0.9047

at 1.00 0.0000

Average precision (non-interpolated) for all rel docs(averaged over queries)

0.8745

Precision:

At 5 docs: 0.9333

At 10 docs: 0.9333

At 15 docs: 0.9111

At 20 docs: 0.9167

At 30 docs: 0.9111

At 100 docs: 0.9100

At 200 docs: 0.9033

At 500 docs: 0.3613

At 1000 docs: 0.1807

R-Precision (precision after R (= num_rel for a query) docs retrieved):

Exact: 0.9001

iii. Analysis of the results

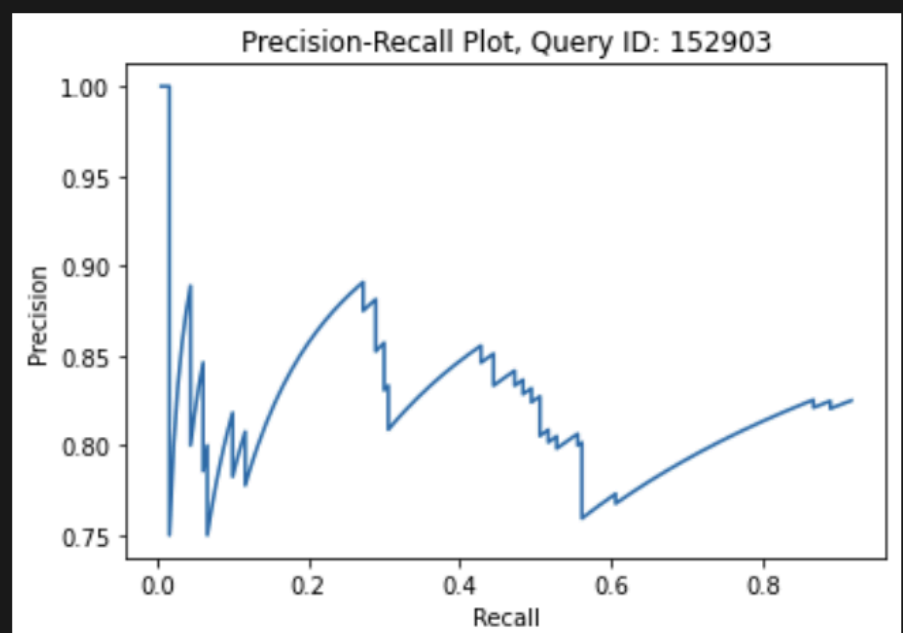
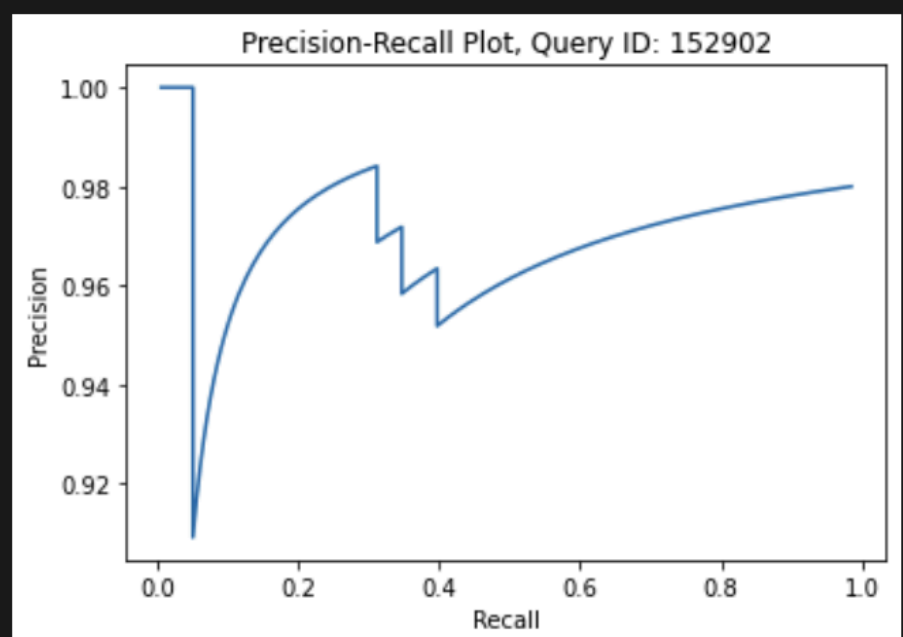
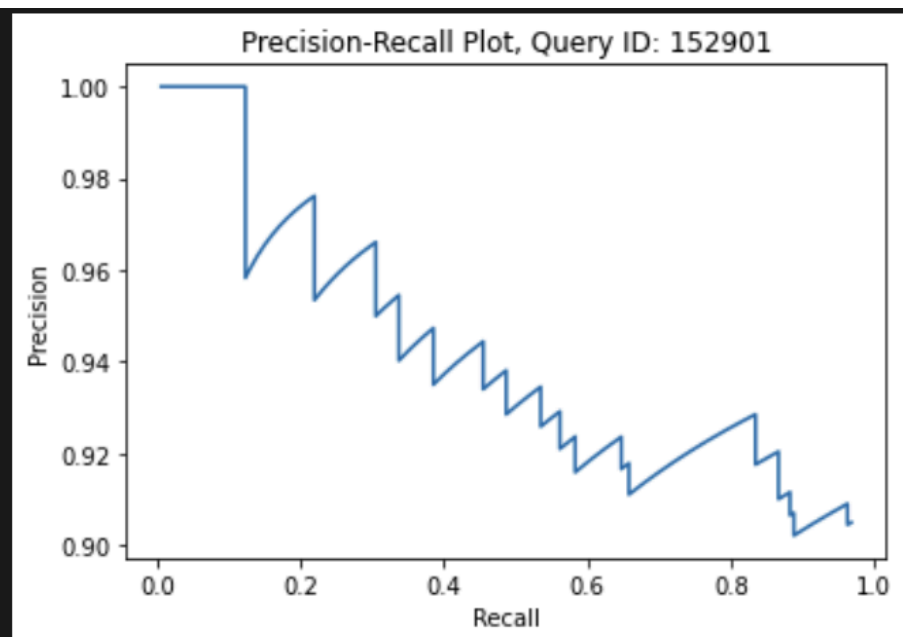
The results from our trec_eval match the provided trec-eval closely, indicating consistency in the evaluation process.

c. Precision-Recall Curves (For MS Students)

i. Description of the method to create precision-recall plots

1. Iterate through each query in the ranked list and calculate relevant metrics such as precision, recall, average precision, and R-precision.
2. Generate precision-recall points for each query and optionally plot them if the plot_graph attribute is set to True.
 - a. For each query, plot the precision values against the corresponding recall values.
 - b. Customize the plot with appropriate labels for the x-axis (Recall) and y-axis (Precision), as well as a title indicating the query ID.

ii. Graphs for each query



6. Extra Credit (If Attempted)

- Description of the extra credit task(s) undertaken
 - A nice, fast, web interface for evaluation, will get you EC points.
 - Also EC points will be given for highlighting query terms in the document source when the entire document is displayed.

The Ebola outbreak 2013-2016: old lessons for new epidemics

PubMed

Author: Jianqing

An official website of the United States government The .gov means it's official. Federal government websites often end in .gov or .mil. Before sharing sensitive information, make sure you're on a federal government site. The site is secure. The https:// ensures that you are connecting to the official website and that any information you provide is encrypted and transmitted securely. Ebola virus causes a severe haemorrhagic fever in humans with high case fatality and significant epidemic potential. The 2013-2016 outbreak in West Africa was unprecedented in scale, being larger than all previous outbreaks combined, with 28 646 reported cases and 11 323 reported deaths. It was also unique in its geographical distribution and multicountry spread. It is vital that the lessons learned from the world's largest Ebola outbreak are not lost. This article aims to provide a detailed description of the evolution of the outbreak. We contextualize this outbreak in relation to previous Ebola outbreaks and outline the theories regarding its origins and emergence. The outbreak is described by country, in chronological order, including epidemiological parameters and implementation of outbreak containment strategies. We then summarize the factors that led to rapid and extensive propagation, as well as highlight the key successes, failures and lessons learned from this outbreak and the response. This article is part of the themed issue 'The 2013-2016 West African Ebola epidemic: data, decision-making and disease control'. Keywords

; Ebola outbreak; Ebola virus disease; West Africa. © 2017 The Authors. PubMed Disclaimer A.M.J. is a governor of the Wellcome Trust. The authors have no other conflicts of interest to declare. Figure 1. Timeline of key events with country-specific epidemic curves. Case numbers are total reported. Figure 2. Geographical map of Guinea, Sierra Leone and Liberia showing districts and total number of cases. Figure 3. Initial transmission chain in Guinea. HCW, healthcare worker. (Adapted from Baize et al. Figure 4. Where the outbreak began. Map to show Kissi tribal area spanning Guinea, Sierra Leone. NCBI Literature Resources MeSH PMC Bookshelf Disclaimer The PubMed wordmark and PubMed logo are registered trademarks of the U.S. Department of Health and Human Services (HHS). Unauthorized use of these marks is strictly prohibited. Connect with NLM National Library of Medicine 8600 Rockville Pike Bethesda, MD 20894 Web Policies FOIA HHS Vulnerability Disclosure Help Accessibility Careers

Relevance Grade for Document 27 (Query: West African Ebola epidemic):

Very relevant

Submit for Document 27 (Query: West African Ebola epidemic)