Data is the new gold: Information flow during the 2014-2016 Ebola outbreak and the Coronavirus outbreak

Delphine Courvoisier
University of Geneva & Geneva University Hospitals
Pillars of epidemics response

1. Treating sick patients
2. Community engagement
3. Safe burial
4. Surveillance
   1. Information management
Outline

Epidemics, pandemics, and information flow

• A blood sample is not a person
• What’s behind the $R_0$?
• Garbage In, Garbage Out (GIGO)
• Infrastructure of data flow, data security, → who owns the data
• The example of COVID19 cases and contact tracing in Geneva
A blood sample is not a person

Labs provide the results of each test.
- The quality of the tube’s tag is crucial
- An example with Ebola, a week in January 2015
  - CDC (US): 82 cases
  - WHO: 51 cases
- What happened
  - Same source of data
  - Different algorithms
Different algorithm for date of onset

CDC used

- if dead (StatusAsOfCurrentDate), dateOnset=date of death - 4
- if dead and no date of death, dateOnset=SampleTest - 5
- if dead and no date of death and no SampleTest, dateOnset=date Report - 4
- if alive and no dateOnset, dateOnset=DateReport - 4
- if alive and no dateOnset and no DateReport, dateOnset=SampleTest - 7
- if missing StatusAsOfCurrentDate and no dateOnset, dateOnset=DateReport - 4
- if missing StatusAsOfCurrentDate and no dateOnset no DateReport, dateOnset=SampleTest - 6
- if DateOnset < 01/05/2015 (May 1\textsuperscript{st} 2015), dateOnset=DateReport - 4

WHO used:

- if DateOnset is missing, DateOnset=dateReport - 4
- if DateOnset and dateReport are missing, dateOnset=DateDeath - 5

We suggest that the procedure for imputing DateOnset should be decided as a policy by the leads of CDC and WHO teams.
Different algorithms for district

District name (DistrictRes or DistrictOnset variables): these district names are entered using various spelling resulting in more than 70 different values

• CDC used: district of onset, and correct the spelling.
• WHO used: district of residence and if missing, districtRes=districtOnset.
What’s behind the $R_0$?

\[
R_0 \propto \left( \frac{\text{infection}}{\text{contact}} \right) \cdot \left( \frac{\text{contact}}{\text{time}} \right) \cdot \left( \frac{\text{time}}{\text{infection}} \right)
\]

- **Transmissibility** (i.e., probability of infection given contact between a susceptible and infected individual)
- Average **rate of contact** between susceptible and infected individuals
- **Duration of infectiousness**
What’s behind the $R_0$?

$$R_0 \propto \left( \frac{\text{infection}}{\text{contact}} \right) \cdot \left( \frac{\text{contact}}{\text{time}} \right) \cdot \left( \frac{\text{time}}{\text{infection}} \right)$$

- **Transmissibility**: may decrease when temperature increases
- **Rate of contact**: underlies decisions for quarantine, confinement, self-isolation
- **Duration of infectiousness**: highlights the importance of finding a drug that would reduce disease duration
Garbage In, Garbage Out (GIGO)

- Data on patients with the disease or on the number of tests is necessary for:
  - Planning care resources
  - Helping patients
  - Monitoring disease progression and understanding when the disease is contained

→ Data quality input is crucial
# Garbage In, Garbage Out (GIGO)

<table>
<thead>
<tr>
<th>Aim</th>
<th>Examples of GI</th>
<th>Examples of GO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning care resources</td>
<td>- SL: missing name info</td>
<td>Samples were considered duplicates</td>
</tr>
<tr>
<td></td>
<td>- CH: Fax sent to FOPH may take time to be inputted</td>
<td>Less precise estimation of $R_0$</td>
</tr>
</tbody>
</table>
Switzerland epidemiological curve
## Garbage In, Garbage Out (GIGO)

<table>
<thead>
<tr>
<th>Aim</th>
<th>Examples of GI</th>
<th>Examples of GO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning care re ssources</td>
<td>- SL: missing name info</td>
<td>Samples were considered duplicates</td>
</tr>
<tr>
<td></td>
<td>- CH: Fax sent to FOPH may take time to be inputted</td>
<td>Less precise estimation of $R_0$</td>
</tr>
<tr>
<td>Helping patients</td>
<td>- GE: Labs do not provide accurate phone number</td>
<td>- Delay in contacting pts → isolation starts later</td>
</tr>
</tbody>
</table>
## Garbage In, Garbage Out (GIGO)

<table>
<thead>
<tr>
<th>Aim</th>
<th>Examples of GI</th>
<th>Examples of GO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning care ressources</td>
<td>- SL: missing name info</td>
<td>Samples were considered duplicates</td>
</tr>
<tr>
<td></td>
<td>- CH: Fax sent to FOPH may take time to be inputted</td>
<td>Less precise estimation of R₀</td>
</tr>
<tr>
<td>Helping patients</td>
<td>- GE: Labs do not provide accurate phone number</td>
<td>- Delay in contacting pts → isolation starts later</td>
</tr>
<tr>
<td>Monitoring disease progression and understanding when the disease is contained</td>
<td>- Only deaths at hospital are reported</td>
<td>- Underestimation of deaths and case fatality rate</td>
</tr>
<tr>
<td></td>
<td>- GE: Addresses often wrong or missing</td>
<td>- Incorrect reporting of clusters</td>
</tr>
</tbody>
</table>
Who owns the data?

- Patients own their data! Yes, but:
  - In the case of an epidemics, healthcare professionals need to have access to these data even if the patient has not given consent.
    - Ebola SL: data was hosted in the US because CDC provided a solution for case monitoring first.
  - Law of epidemics, art 12, obligation to declare. To paraphrase Alain Berset: “Security should be as strong as possible, but as lax as necessary.”
    - COVID Geneva: Office of the surgeon general required that the labs provide cellphone numbers to allow rapid monitoring and self-quarantine advice to new cases
Actionnable Registry of the Geneva Outpatient with SARS-Cov2 (ARGOS)
- Data flow

Delphine Courvoisier, Denis Mongin, Adriana Uribe, Jean-Luc Falcone, Sophia Vignard

Delphine.Courvoisier@hcuge.ch
ARGOS

• A blood sample is not a person → avoiding duplicates
• What’s behind the $R_0$?
• Garbage In, Garbage Out (GIGO)
• Infrastructure of data flow, data security, → who owns the data
Data flow: cases

Avoiding duplicates

LABORATORIES

HUG
Hospital, intensive care, deaths

COVICARE PARTNERS
Hospital data

FOPH

DCPM
Deaths, addresses

Policies

• CRIVE
• Unilabs
• Dianalabs
• SYNLAB
• MGD
• Espace-lab
• Vioiller

Transmission of lab data (8am, 12, 5pm) to secure email

10 min data management and integration in REDCap

Collect, clean, integrate data

Enter data for impossible to find (by nurses) patients

De-identified data extraction (solution for automatization with API)

Process OK

Process may change

COVICARE PARTNERS

Nurses / MD

COVICARE

Nurses / MD

Qlick Sense

REPORTING

MODELIZATION

DATA SECURITY

Schematic by Sophia Vignard
Data flow: contacts

1. Positive case records
   - Patient A record
     - Contacts list
     - contact A.1
     - contact A.2
     - contact A.3
     - Covid-19 positive
   - Patient ... record
     - Contacts list
     - contact ...
     - ...

2. Pre-Test: Waiting room
   - Patient A
     - Contacts list
     - contact A.1
     - contact A.2
     - contact A.3
   - Patient B
     - Contacts list
     - contact B.1
     - contact B.2
     - contact B.3

3. Contacts records
   - Contact A.1 record
     - Follow-ups
     - event 1
     - event 2
     - event 3
   - Contact A.2 record
     - Follow-ups
     - event 1
     - event 2
   - Contact A.3 record
     - Follow-ups
     - event 1
     - event 2

4. Labs: Test results
   - Covid-19 positive
   - Covid-19 negative

5. Negative case records
   - contact ...
   - ...
   - ...
What it really looks like...

~5000 lines of code developed in 6 weeks

```r
comp_names("11.08.1986","11.08.1975")
comp_names_vec(c("Delphine Sophie","denis","adriana alba","Hugo"),c("delphine","d@i:anis","Adriana","jean-pierre"))

# fonction une ligne recap ################
# a partir d'un dump, output une ligne par patient
# avec info labo, fin de suivi, premier suivi, et premiere hospitalisation
# a faire: tester, si plusieurs prenom, en inversant l'ordre des prenom
# idem nom. Ex:

oneline_recap = function(filepath){
  recap_pat <- fread(filepath,encoding = "UTF-8")

  # if positif project, set on one line:
  if("recap_event_name" %in% names(recap_pat)){
    # to determine which variable is in which event
    plouf <- recap_pat[,lapply(.SD,function(col){all(is.na(col)) & !all(col == "")}),by = .(recap_event_name,recap_repeat_instrument)]

    # looping on event
    eventdata_list <- lapply(c("s1_arm_1","fin_de_suivi_arm_1"),function(event){
      vars <- plouf[recap_event_name == event & (recap_repeat_instrument =="") | is.na(recap_repeat_instrument)],-c("recap_data_access_group","recap_data_access_group_id") %>%
        .[value == T,variable] %>%
        as.character() %>%
      if(length(vars) == 0){vars <- "record_id"}
      recap_pat[recap_event_name == event & (recap_repeat_instrument =="") | is.na(recap_repeat_instrument)]]$SD,SDcols = vars]
    }

    vars <- c("record_id","recap_event_name","recap_repeat_instrument","positif","nudnik","npot_hug","source","source_autre","prescripteur","facteurrisque_age","genre","telephone","mailert","composedemail","resend","resend_2","resend_3","telephone2","date_visite","date_","codepostal","adressecapanton","perdu_vue__1","res_test1","date_res1","res_test2","date_res2","res_test3","date_res3","date_res4"),
    vars <- vars[vars %in% names(recap_pat)]
    eventdata_list[[3]] <- recap_pat[recap_event_name == "labo_arm_1",.SD,SDcols = vars]

    # add hospitalisation if needed
    vars <- plouf[recap_event_name == "s1_arm_1" & recap_repeat_instrument =="hospitalisation",-c("recap_data_access_group")]
```
A multidisciplinary team effort
Thank you!

Nurses
Doctors
IT team
Data scientists
Biologists

Doctors / Nurses
Digital health experts
Epidemiologists
IT team / Business intelligence
Project managers
Business analysts
Police officers

IT developers
Data scientists