Risk Scoring in Digital Contact Tracing Apps

Michael Höhle[1]

[1]Department of Mathematics, Stockholm University, Sweden
m_hoehle

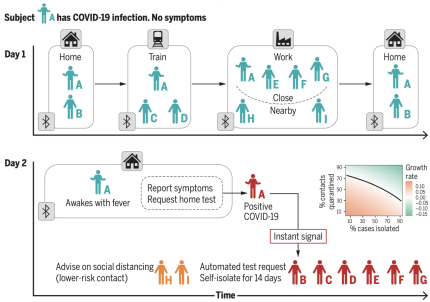
MatStat Seminar, SU
09 Dec 2020

Overview

Outline

Contact tracing

- Contact tracing is the process of identifying persons who had contact with an infected person in order to test, isolate and trace onwards
- Contact tracing is an important non-pharmaceutical control measure
- In an influential paper, Ferretti et al. (2020) argued that a digital contact tracing app could speed up contact tracing s.t. $R(t) < 1$ would be possible



Source: Fig. 4 of Ferretti et al. (2020)

Outline

GAENF (1)

- The Google and Apple exposure notification framework (GAENF) enables privacy-preserving contact tracing on mobile devices based on Bluetooth Low Energy
- Each device continuously emits so called Rolling Proximity Identifiers (RPI, changing every $\sim$20 min) derived cryptographically from a daily temporary exposure key (TEK)
- Each device regularly (every $\sim$5 min) scans all visible RPIs and stores these locally together with their signal strength
- If a user is tested positive for COVID-19, they can volunteer to upload their TEKs of the past 14 days to a central server

GAENF (2)

- An uploaded TEK is called a diagnosis key
- All users download the available diagnosis keys daily and compare the corresponding RPIs to the locally stored RPIs
- The GAENF is used by country specific COVID-19 Apps in several European countries, e.g., Switzerland, Latvia, Denmark, Italy, Poland, Ireland, Germany, UK, ...
- Since the summer the digital contact tracing apps also work between (many of) the countries

Outline

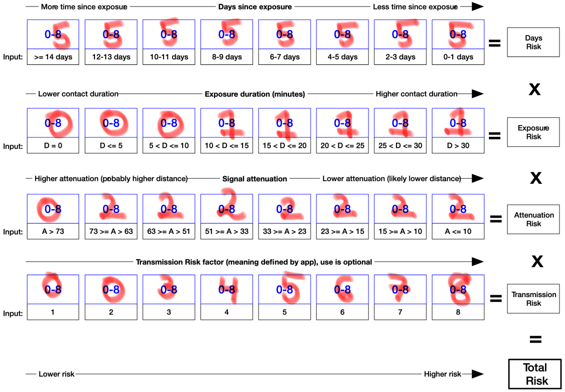
Risk Scoring in the GAENF (v1)

- Consider a specific user. Let I be the user's subset of all stored exposures within the past 14 days which originate from keys known to be COVID-19 test positive
- In the GAENF (v1) we compute the **total risk** for each exposure with a test positive, i.e. for each element $i \in I$, by

$$\text{TR}(i) = \text{duration}(i) \times \text{attenuation}(i) \times \text{TRL}(i) \times \text{days}(i),$$

  where $\text{TRL}(i)$ is the so called **transmission risk level**.
- The GAENF operationalizes all RHS functions by an integer value between 0 and 8

Risk Calculation per Exposure Key



Source: Corona Warn App, Solution Architecture, Fig. 12 (Apache License 2.0) + current config annotated

Transmission Risk Level

- The transmission risk level can be used to enhance the classification with epidemiological information about the potential source

- The following derivations are closely following the document *Epidemiological Motivation of the Transmission Risk Level* (CWA Team, 2020) written as part of the German CWA development[1]

[1] Reflects COVID-19 knowledge as of May-Jun 2020

Sequence of events for an infected person A

- $T_E = T_{\text{infection}}$: transmission of SARS-CoV-2 to an exposed person A from some unknown source
- $T_I = T_{\text{infectious}}$: start of the infectious period in person A, i.e. A is able to infect others
- $T_S = T_{\text{symptoms}}$: onset of symptoms in person A (also referred to as DSO, day of symptom onset)
- $T_P = T_{\text{sampling}}$: time of sampling of person A
- $T_R = T_{\text{result}}$: time of A obtaining the positive test result
- $T_U = T_{\text{upload}}$: time where person A uploads the positive test result to the system (aka. DU, day of upload)

Information about Symptoms in A at DU

Four cases:

1. Knowledge of DSO in A
2. Knowledge that A was symptomatic, but DSO not available, only DU
3. Knowledge that A was not symptomatic at or before DU (completely asymptomatic or pre-symptomatic)
4. Unknown symptom status at DU

Note:

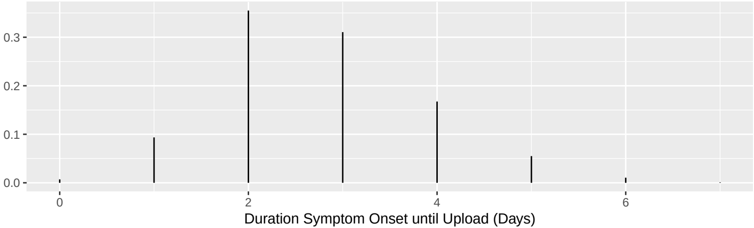
The previously shown TRL corresponds to Case 4, which is computed as a marginalization over Cases 1-3.

Case 1: Duration between DSO and DU

- Consider the durations $T_U - T_R$, $T_R - T_P$ and $T_P - T_S$
- We find the distribution of

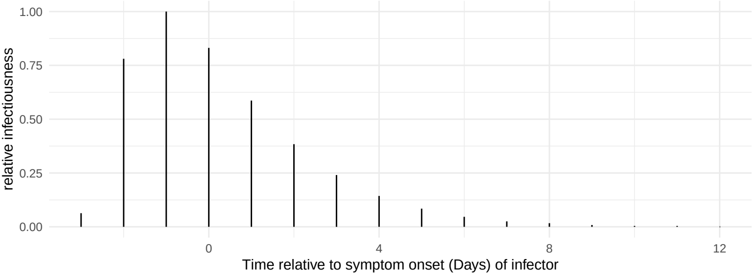
$$T_U - T_S = (T_U - T_R) + (T_R - T_P) + (T_P - T_S)$$

  by a convolution of the 3 durations
- Given informed assumptions we obtain the following PMF



Infectiousness (1)

- He et al. (2020a) provides a distribution for time of exposure in B relative to symptom onset in A
- Distribution is used as proxy for the infectiousness of A, $v_A(d)$



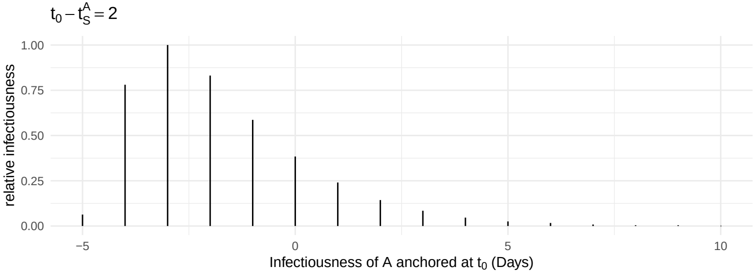
- Recently, a correction appeared moving the graph further to the left (He et al., 2020b).

Infectiousness (2)

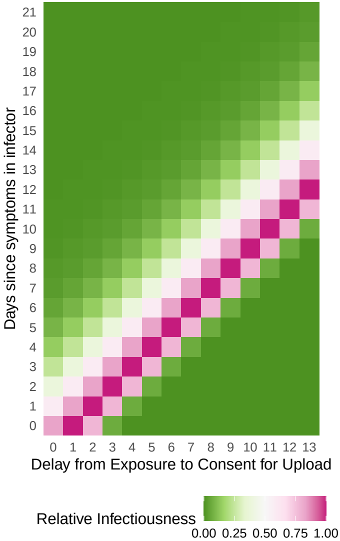
- At time now, t_0, the onset of symptoms in the primary case A happened $t_0 - t_S^A \geq 0$ days ago.
- Relative infectiousness of A anchored s.t. time zero is at t_0 and a function of how many days d ago the contact between A and B was:

$$\lambda_A^*(d) = v_A(-d + (t_0 - t_S^A))$$

- Example: Plot of the infectiousness profile, if DSO happened 2 days before upload.



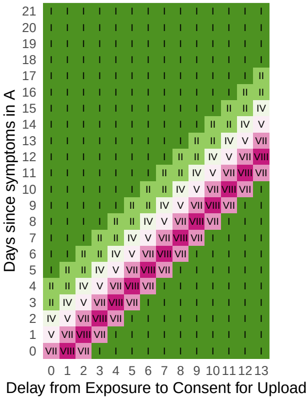
Infectiousness (3)



Infectiousness (4) - Transmission risk level

Discretization of the $[0, 1]$-scale into 8 equal-sized intervals:

$$\lambda_A(d) = l \quad \Leftrightarrow \quad \frac{l-1}{8} \leq \lambda_A^{(*)}(d) < \frac{l}{8}, \quad \text{where } l \in \{1, \ldots, 8\}.$$



Overall Risk (1)

- Exposures with a total risk below the minimum risk score of 11 are set to zero
- A weighted summation of time spent in 3 attenuation buckets (distance: far, intermediate and close) is computed

$$wt_j = \max\left(30, \sum_{i \in I^*} \mathcal{I}\left(l_j \leq \text{attenuation}(i) < u_j\right) \times \text{duration}(i) \times w_j\right),$$

where $I^* = \{i \in I : \text{TR}(i) > 0\}$, $w_1 = 0, w_2 = 0.5$ and $w_3 = 1$.

- The limits (l_j, u_j) of the three classes are chosen at $[63, \infty)$, $[55, 63)$ and $[0, 55)$ dB.
- The **Exposure Score** is then calculated as as $\sum_{j=1}^{3} wt_j$.

Overall Risk (2)

- Next, a normalized total risk score defined as

$$\text{Normalized Total Risk Score} := \frac{\max_{i \in I} \mathrm{TR}(i)}{50}.$$

- The combined risk score is now given as

$$\text{Combined Risk Score} = \text{Exposure Score} \times \text{Normalized Total Risk Score}$$

- A warning is given by the app, if

$$\text{Combined Risk Score} > 15.$$

Outline

Summary...

- Mathematical parametrization of the GAENF risk classifier in an app downloaded 23.5 mio times[2]
- Contact tracing apps are not the silver bullet for COVID-19 control. However, they provide an important supplement to manual contact tracing and speed up the turn-around time for test results
- The success of the app heavily depends on the widespread voluntary use in the population
- One component of winning as many users as possible is **transparancy**
  - Source code and documentation of the CWA app and server back-end is available from GitHub
  - Description of the risk scoring is available at several audience levels (user, epidemiologist, mathematician)

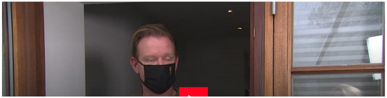
[2] As of 2020-12-04

A not so good example... (from Denmark)

PENGE

Jens fik corona, men smittestop-app gav ikke hans kone besked

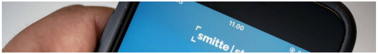
Problem med appen Smittestop er et tillidsbrud, mener De Konservative.



PENGE

Fejl i Smittestop vokser i omfang: Kollegaer og venner kan have misset beskeder

Der er en masse smittekæder, der ikke bliver brudt på grund af fejl i appen, vurderer epidemiolog.



TEKNOLOGI

Myndigheder erkender fejl i Smittestop-app

De allernærmeste kontakter har ikke i alle tilfælde fået besked om smittefare. Uvist, om fejlen har været der fra starten af.

... it turns out that NetCompany, which created the app for the Ministry of Health and the Elderly, has either misunderstood or been misinformed about a key element of the technology developed by Apple and Google - an element that determines who is notified of the corona infection.

Source: DR Nyheter Sep 2020

Opinions...

- There is a trade-off to make between complexity and robustness in the risk scoring model
- Living up to scientific standards is more than code on GitHub:
  - transparency is needed both at the high-level as well as at for detailed mathematical descriptions
  - evidence is continuously evolving and has to be updated
  - engagement with the user community is an opportunity for peer-review
  - scientific peer-review publication assures scientific standard
  - a plan is needed how to collect data about the use and impact of the app (data protection and privacy preservation is IMO not an excuse to do nothing)
- Details and an example of the GAENF risk calculation can be found in a recent blog post (Höhle, 2020)

References I

CWA Team (2020). Epidemiological Motivation of the Transmission Risk Level. Corona-Warn-App.

Ferretti, L., Wymant, C., Kendall, M., Zhao, L., Nurtay, A., Abeler-Dörner, L., Parker, M., Bonsall, D., and Fraser, C. (2020). Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing. Science, 368(6491):eabb6936.

He, X., Lau, E. H. Y., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y. C., Wong, J. Y., Guan, Y., Tan, X., Mo, X., Chen, Y., Liao, B., Chen, W., Hu, F., Zhang, Q., Zhong, M., Wu, Y., Zhao, L., Zhang, F., Cowling, B. J., Li, F., and Leung, G. M. (2020b). Author correction: Temporal dynamics in viral shedding and transmissibility of covid-19. Nature Medicine.

He, X., Lau, E. H. Y., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y. C., Wong, J. Y., Guan, Y., Tan, X., Mo, X., Chen, Y., Liao, B., Chen, W., Hu, F., Zhang, Q., Zhong, M., Wu, Y., Zhao, L., Zhang, F., Cowling, B. J., Li, F., and Leung, G. M. (2020a). Temporal dynamics in viral shedding and transmissibility of covid-19. Nature Medicine, 26(5):672–675.

Höhle, M. (2020). `https://staff.math.su.se/hoehle/blog/2020/09/17/gaen_riskscoring.html`. Blog post.

Appendix

Appendix

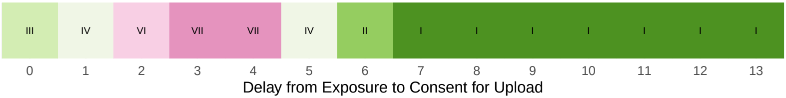
Case 2 (1)

- We know that A is symptomatic at time of upload, but do not know the exact DSO
- We can infer a probability distribution for the DSO T_S^A from knowledge about the duration between DSO and upload given by and knowing the date of upload t_0
- Marginalizing over the DSO

$$
\begin{aligned}
\lambda_A^{(*)}(d,\ T_S^A \leq t_0) &= \mathbb{E}_{T_S^A}\left[\lambda_A^{(*)}(d, T_S^A = t_S^A) \,\middle|\, T_S^A \leq t_0\right] \\
&= \sum_{t_S^A = t_0 - 13}^{t_0} \lambda_A^{(*)}(d, T_S^A = t_S^A) \cdot f_{T_U - T_S}(t_0 - t_S^A)
\end{aligned}
$$

Case 2 (2)

Resulting risk levels:



Case 3 and 4

See the details in CWA Team (2020).