

Betere overheidsdata voor sterk beleidsonderzoek en goed bestuur

Vlaams Parlement

Commissie voor Algemeen Beleid,
Financiën, Begroting en Justitie

22 februari 2022



structureel overleg tussen overheid, onderzoekers en
andere gebruikers opzetten



Betere verzameling en deling van overheidsdata nodig

Inleiding

Rik Van de Walle – voorzitter VLIR

Thematische toelichting

Case Arbeidsmarktbeleid

Sarah Vansteenkiste (KU Leuven)

Case Gezondheidsbeleid

Niel Hens (UHasselt, UAntwerpen)

Toelichting oproep en advies SERV-VLIR

Koen Verlaeckt (VLIR)

Stijn Wouters (SERV)

Conclusie

Hans Maertens – voorzitter SERV

Thema arbeidsmarktbeleid

Prof. dr. Sarah Vansteenkiste
KU Leuven

KU LEUVEN

**STEUNPUNT
WERK**



Vlaanderen
is werk



Datalacunes en –gebreken: **survey data**

Analyse



Oplossingen

- Datalacunes arbeidsreserve, bv.
 - Geen zicht op uitkeringsstelsels groepen van arbeidsreserve
 - Geen info over doorstroom naar arbeidsmarkt
 - Geen structurele survey-infrastructuur/panels die inspelen op actuele arbeidsmarktbeleid-vraagstukken
 - VRAAGZIJDE: geen representatieve info. Bv. welke knelpunten bij aanwervingen? Welke competenties nodig, welke tekorten? Kwetsbare werkgevers?
 - AANBODZIJDE: welke drempels om zich aan te bieden op arbeidsmarkt? Welke competenties zijn er (niet)? Bereik opleidingsaanbod en loopbaanbegeleiding?
- Maak datakoppelingen tussen survey en admin. data makkelijker
 - Investeer in survey- en panelinfrastructuur, met mogelijkheid om actuele beleids- of onderzoeksvragen op te nemen in het design via in te schuiven modules
 - Voordeel van meer coördinatie in bevestigingen, en tegengaan survey-moeheid

Datalacunes en –gebreken: admin. data

Analyse



Oplossingen

- Datalacunes arbeidsreserve in DWH AM&SB:
 - Geen zicht op groepen van arbeidsreserve zoals herintreders (huisvrouwen,...), NEET-jongeren
 - Geen beroepeninformatie i.f.v. competentieprognoses
 - Data loopt sterk achter, waardoor je actuele beleidsvragen moeilijker kan beantwoorden
 - Te weinig personeelscapaciteit leidt tot:
 - Lange duurtijden in verwerken van de vele data-aanvragen en aanleveren data
 - Fouten in data-aanlevering
 - Onvoldoende incentives bij dataleveranciers om data goed te documenteren
 - Risico op verkeerde interpretaties en resultaten
- Betrek kennis expertcentra beleidsondersteuning en studiediensten om datahiaten te duiden
 - Wetgevend kader kan helpen
 - Maak data actueler en relevanter door dataleveranciers aan te sturen vanuit gezamenlijk comité (creatie bovenstructuur)
 - Verlicht aanvraagprocedures en voorzie meer capaciteit. Zet thematische datasets open voor onderzoeksinstellingen en studiediensten

Gebrek aan of te tijdrovende data-uitwisseling en -gebreken: admin. data

Analyse



Oplossingen

- Vaak jaar op jaar dezelfde data-aanvragen herhalen en procedures doorlopen
 - Tijdrovende datagoedkeuringen, vb. arbeidsreserve aanvraag DWH AM&SB
 - 1,5 jaar (!) geduurd
 - Interessante variabelen moeten laten vallen omdat data anders niet verkregen zou worden zoals info werkverleden, regionale insteek,...
 - Data-uitwisselingen lopen steeds moeilijker, vb.
 - Monitoring arbeidsmarktposities tot op lokaal niveau: Uitwisseling RSZ data met FOD WASO
 - Monitoring arbeidsmarktposities naar herkomst: Uitwisseling herkomstgegevens DWH AM&SB
- Laat verlengingen van aanvragen makkelijker toe + werken op vaste dataset
 - Capaciteit uitbreiden DWH AM&SB
 - Werk naar eenzelfde interpretatie van privacy/GDPR-regels door alle overheidsdiensten en toegepast op onderzoek
 - Heb aandacht voor/vertrouwen in procedures die onderzoekers dienen te hanteren in omgang met gedetailleerde data

Te gebrekkige monitoring en – evaluatiecultuur in het arbeidsmarktbeleid

Analyse



Oplossingen

- Heel veel beleidsinitiatieven om 80% werkzaamheidsgraad te bereiken en krapte te helpen beantwoorden
- Maar: geen structurele (ex ante en ex post) monitoring aan verbonden. Wat werkt? Wat leidt tot werk?
 - Activering:
 - Tijdelijke Werkervaring, wijkwerken, gemeenschapsdienst, ESF programma's,...
 - Recommender systemen VDAB bij matching werkzoekenden en vacatures.
 - Opleiding en LLL
 - Opleiding voor TW, ESF programma's, 5 opleidingsdagen,...
 - Interregionale mobiliteit
- Versterk capaciteit voor monitoring, ook in functie van Vlaamse Brede Heroverweging
- Cultuur van evidence-based beleid omarmen

Open data creëert kansen, maar zet ook (participatie aan) arbeidsmarktonderzoek onder druk

Analyse



Oplossingen

- Open data informeert beleid, onderzoekers en burgers
 - Maar ook keerzijde cf. nieuwe voorgestelde dataverwerkersovereenkomst DWSE en Steunpunt Werk. Is het Steunpunt dataverwerker of onderzoeksinstelling?
 - Willen mensen nog participeren aan onderzoek als alle data met overheid gedeeld moet worden en herkenbaar kan worden?
 - Zelfs pseudonomisering kan bij samenleggen data leiden tot achterhalen identiteit
- Beschouw expertisecentra (steunpunten) als onderzoeksinstellingen, die geaggregeerd kunnen communiceren met overheid
 - Stel verwerkersovereenkomsten en open data strategieën op in samenwerking met onderzoekers

Thema gezondheidsbeleid

Prof. dr. Niel Hens

UHasselt & UAntwerpen



Universiteit
Antwerpen

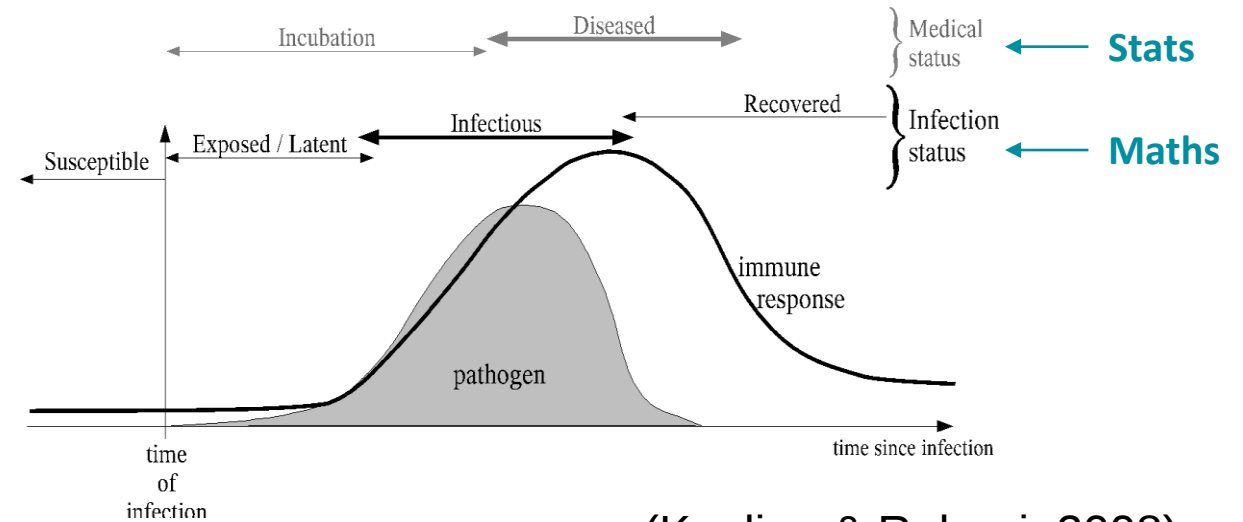




Focus: huidige SARS-CoV-2 pandemie



Daniel Bernoulli (1700 – 1782)
wiskundig model voor pokken
(1760-1766)

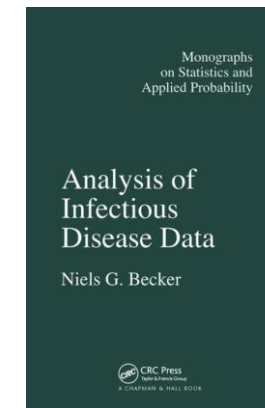


(Keeling & Rohani, 2008)

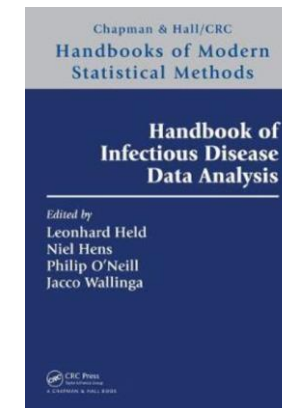
Wiskundige en statistische epidemiologie

- beter begrip verspreiding
- forecasting scenario's
- balans accuraatheid, transparantie, flexibiliteit

data driven
←→
validation



Becker (1989)

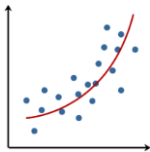
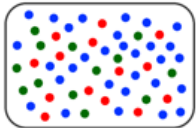
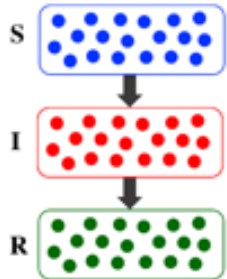


Held et al. (2019)

Moeilijkheden

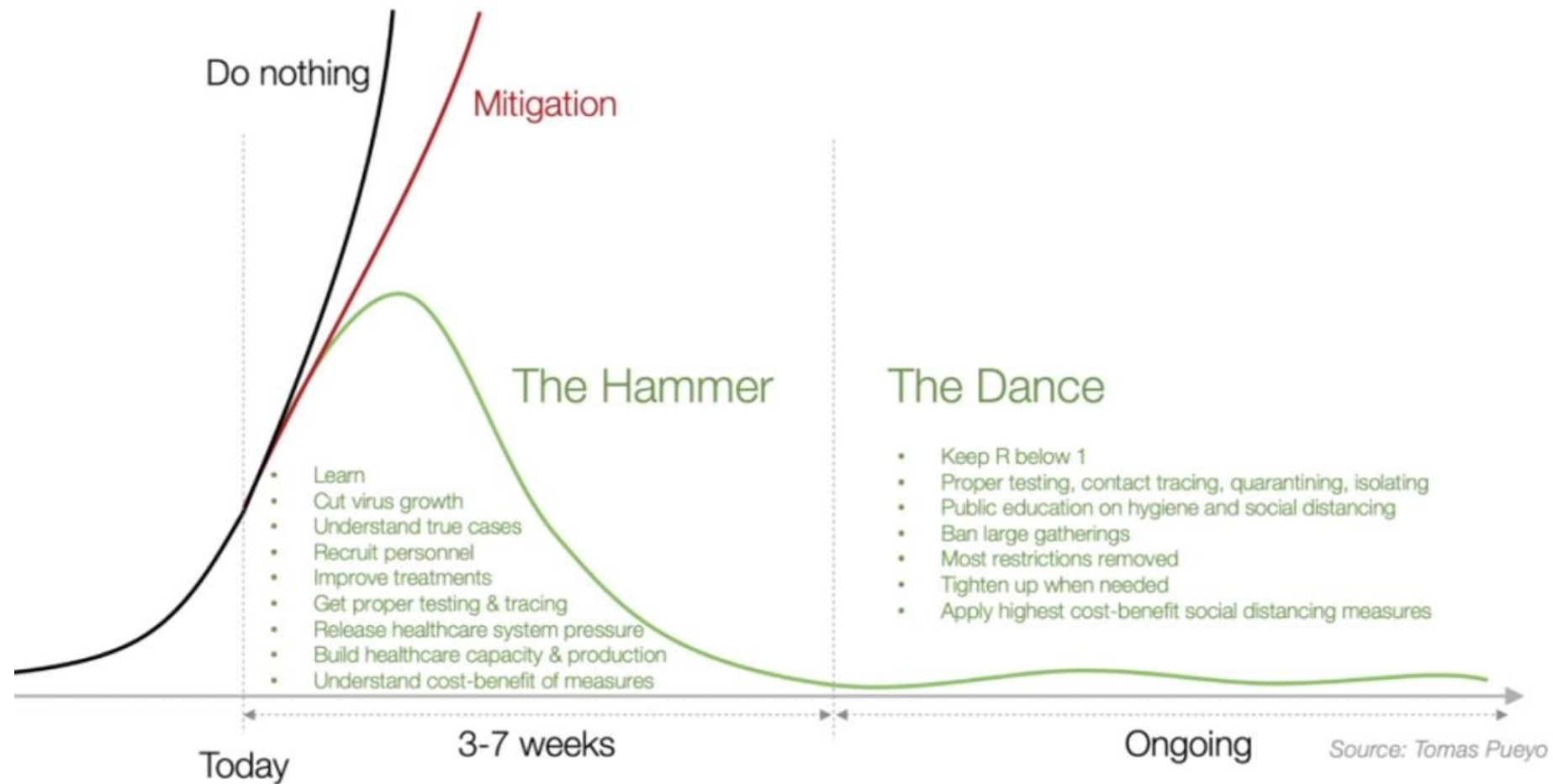
- correlaties
- onvolledige data
- observationeel karakter
- ...

Modellen en data

	Input		Output	Voornaamste veronderstellingen en beperkingen	Team
Zeer korte termijn: Regressiemodel	Indirecte datapunten van laatste 2 weken gelinkt aan hospitalisaties 		Inschatting aantal nieuwe hospitalisaties in komende 2 weken	1. Indirecte data die een mogelijk verband hebben met hospitalisatie (telecom, absenteïsme, test positiviteitsratio...) als proxy 2. Enkel nuttig als <i>early warning</i> voor ziekenhuiscapaciteit	 <i>Simulation Model for Infectious Diseases (SIMID)</i> groep: 30-tal onderzoekers van UHasselt en UAntwerpen o.l.v. Niel Hens en Philippe Beutels
Korte termijn: Individueel Gebaseerd Model (IBM)	11 miljoen individuen die onderling interageren, gekalibreerd op hospitalisaties van voorbije maanden 		Inschatting relatieve impact van NPIs op hospitalisaties en R_e voor de komende 2 maanden	1. Sociale contacten als proxy voor transmissiemomenten, dus geen differentiatie in verschillende NPIs 2. Gebaseerd op aantal hospitalisaties, niet het aantal besmettingen, en loopt dus ca. 2w "achter in de tijd" 3. Geen rekening met reizen 4. Geen rekening met weersomstandigheden 5. Complexiteit van 11 miljoen individuen die interageren met elkaar; supercomputer nodig voor berekening	
Lange termijn: Stochastisch compartimenteel SEIR/PDE model (Referentiemodel)	Populatie van 11 miljoen opgedeeld in compartimenten (<i>susceptible, exposed, infectious, recovered</i>), gekalibreerd op hospitalisaties en doden van voorbije maanden 		Inschatting relatieve impact opening sectoren op hospitalisaties en doden voor de komende 6-7 maanden; belangrijkste model voor politiek beleid en aankoop FAGG bv.	1. Sociale contacten als proxy (idem IBM) 2. Zowel pre-/asymptotische als symptomatische (licht én ernstig) hebben een verschillend niveau van besmettelijkheid en proportie asymptomatische gevallen is constant 3. Immuniteit na doormaken infectie 4. Geen externe invoer van de ziekte in de populatie vs. realiteit dat er bv. 1,3 miljoen PLFs <u>ingevuld</u> werden in aug '20(!) 5. Erg beperkte informatie m.b.t. ziekenhuisontslag, waardoor enkel indirecte voorspelling van ziekenhuiscapaciteit mogelijk is	

Disclaimer: hét juiste of 'beste' ziektevoorspellingsmodel bestaat niet en elk model is onderhevig aan ad hoc veronderstellingen; vandaar ook sensitiviteitsanalyses.

Epidemische fasen en datanoden



Essentielle data

	Invasion (outbreaks)	Exponential growth	The dance	Control	Endemicity
Models	Branching processes	Renewal equation	Mathematical models (SIR, etc)	Next generation approaches (modified) herd immunity	Heterogeneity in susceptibility
Data	Well documented data on transmission chains	Incidence data a/symptomatic incidence	Surveillance data including hospital data, mortality data, genomic data Serological data Mobility & social contact data	Risk-group specific data Vaccination data (coverage, VEx) Serological data	Surveillance data as well as cohort-specific studies monitoring immunity
Management & policy	Characterisation of the disease	Burden	Data on perception, adherence, mental health	Early-warning indicators	Multi-pathogen data

Essentiële data

Data omtrent clusters
contact tracing data

Surveillance data
eerste lijn
ziekenhuizen
mortaliteitsdata
genomische data
serologische data

Risico groepen
leeftijd

Vaccinatie data

Mobiliteitsdata

Contactdata

Mentaal welzijn

Individu-specifieke data

Real time data

Kosten-baten analyse is duidelijk

Data én expertise

Initieel verzamelde data

Beschikbaar na kritiek in de media

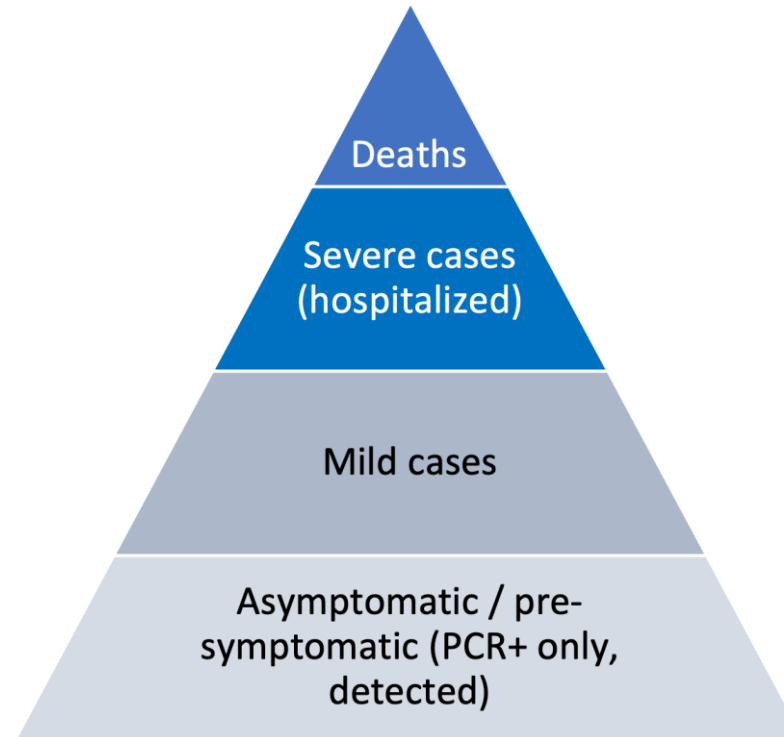
Bevestigde gevallen
Ziekenhuisgegevens
Sterfte data

...

Data Against Corona

Telecom data
Klimatologische data
Absenteïsme data

...



Data én expertise

- Data omtrent transmissie clusters
 - transmissie data: Singapore, Tianjin (China), ...
- Presymptomatische transmissie

- Omikron
 - belangrijke inzichten

RESEARCH

Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020

Tapiwa Ganyani¹, Cécile Kremer¹, Dongxuan Chen^{2,3}, Andrea Torneri^{1,4}, Christel Faes¹, Jacco Wallinga^{2,3}, Niel Hens^{1,4}

1. I-BioStat, Data Science Institute, Hasselt University, Hasselt, Belgium

2. Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, the Netherlands

3. Leiden University Medical Center, Leiden, the Netherlands

4. Centre for Health Economics Research and Modelling Infectious Diseases (CHERMID), Vaccine and Infectious Disease Institute, University of Antwerp, Antwerp, Belgium

Correspondence: Tapiwa Ganyani (tapiwa.ganyani@uhasselt.be)

Citation style for this article:

Ganyani Tapiwa, Kremer Cécile, Chen Dongxuan, Torneri Andrea, Faes Christel, Wallinga Jacco, Hens Niel. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. Euro Surveill. 2020;25(17):pii=2000257. <https://doi.org/10.2807/1560-7917.ES.2020.25.17.2000257>

Article submitted on 06 Mar 2020 / accepted on 07 Apr 2020 / published on 30 Apr 2020

medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES



BMJ Yale

HOME | ABOUT

Search

[Comment on this paper](#)

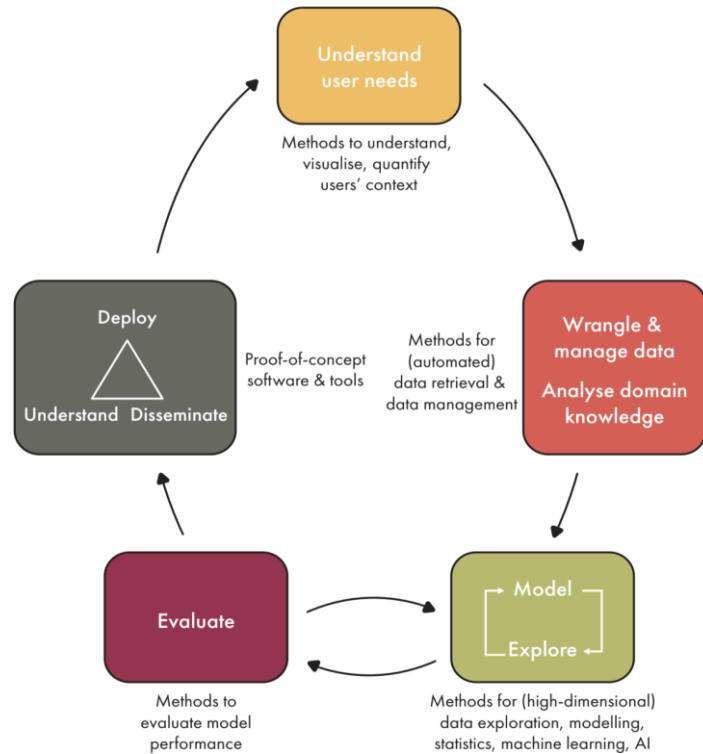
Observed serial intervals of SARS-CoV-2 for the Omicron and Delta variants in Belgium based on contact tracing data, 19 November to 31 December 2021

Cécile Kremer, Toon Braeye, Kristiaan Proesmans, Emmanuel André, Andrea Torneri, Niel Hens

doi: <https://doi.org/10.1101/2022.01.28.22269756>

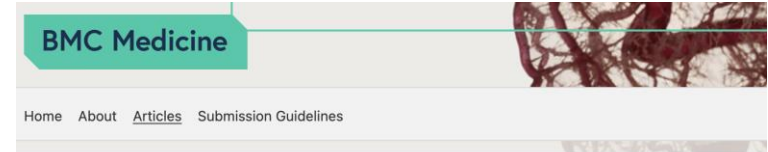
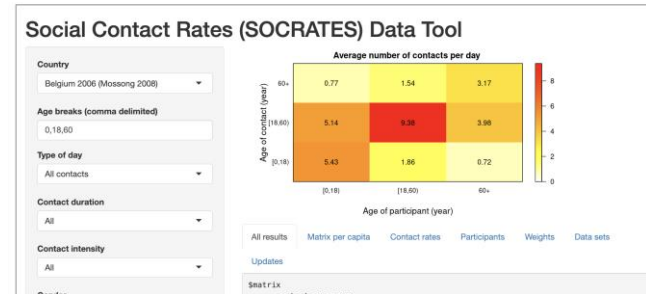
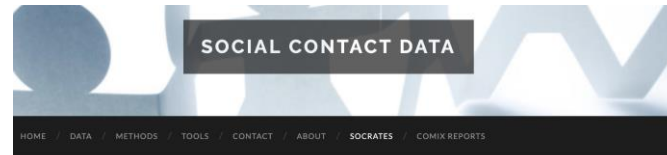
This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should not be used to guide clinical practice.

Data én expertise



Data Science Cycle

(Bron: Data Science Institute, UHasselt)



Database | [Open Access](#) | Published: 29 September 2021

SOCRATES-CoMix: a platform for timely and open-source contact mixing data during and in between COVID-19 surges and interventions in over 20 European countries

[Frederik Verelst](#), [Lisa Hermans](#), [Sarah Vercruysse](#), [Amy Gimma](#), [Pietro Coletti](#), [Jantien A. Backer](#), [Kerry L. M. Wong](#), [James Wambua](#), [Kevin van Zandvoort](#), [Lander Willem](#), [Laurens Bogaardt](#), [Christel Faes](#), [Christopher I. Jarvis](#), [Jacco Wallinga](#), [W. John Edmunds](#), [Philippe Beutels](#) & [Niel Hens](#)

BMC Medicine 19, Article number: 254 (2021) | [Cite this article](#)

1255 Accesses | 25 Altmetric | [Metrics](#)

Open vragen:

- Long Covid
- Mentaal welzijn (GCS, barometer, enquêtes Sciensano)

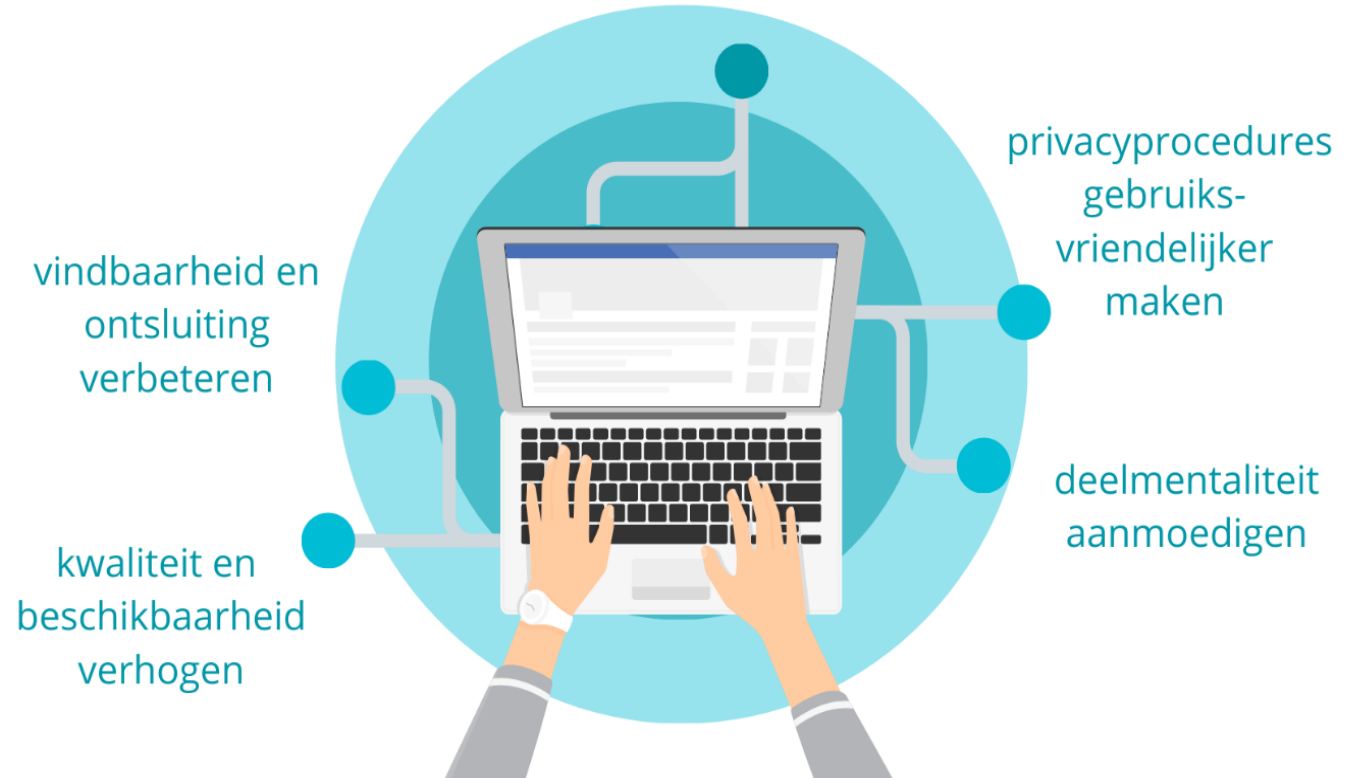
→ Nood aan opzetten van **online probability panel & pandemische cohorte**

Nood aan algehele visie: **FAIR** principes

Advies

**Betere overheidsdata
voor sterk
beleidsonderzoek en
goed bestuur**

structureel overleg tussen overheid, onderzoekers en
andere gebruikers opzetten



Betere verzameling en deling van overheidsdata nodig

Beschikbaarheid en kwaliteit

Analyse

Er zijn nog belangrijke lacunes;
data voldoen niet altijd aan de
FAIR-principes

Er zit vaak te veel vertraging op de
data

Data bevatten vaak nog veel
fouten; er zijn te weinig incentives
voor kwaliteitsvolle data en
metadata

Beschikbaarheid en kwaliteit

Aanbevelingen

Zorg voor mechanismen om datanoden te signaleren, dataleemtes te identificeren en datafouten te remediëren

Maak werk van datakwaliteitsmanagement

Voorzie in een algemene toegang tot confidentiële overheidsdata voor expertcentra in beleidsondersteuning

Vindbaarheid en ontsluiting

Analyse

Het is vaak onduidelijk welke data beschikbaar zijn; datareeksen zitten erg verspreid

Interbestuurlijke uitwisseling van data botst op regio- en landsgrenzen

Het aantal open-datasets is beperkt en te beperkt bruikbaar

Vindbaarheid en ontsluiting

Aanbevelingen

Versterk de inspanningen voor een datavindplaats

Blijf inzetten op thematische dataportalen

Investeer in data-infrastructuren uit met gestandaardiseerde producten (vaste datasets)

Privacy en ethische omgang

Analyse

Het kost veel tijd en moeite om procedures voor data-aanvragen met persoonsgegevens te doorlopen

De procedures zijn weinig gebruikersgericht

Regels voor goed databeheer vinden onvoldoende ingang

Privacy en ethische omgang

Aanbevelingen

Evalueer de privacyregelgeving en de hele keten van het bekomen en gebruiken van data

Zorg voor duidelijkheid en uniformiteit bij toepassing van privacyregels

Vervang individuele machtigingen en protocollen zoveel mogelijk door eenvormige afsprakenkaders

Cultuur en organisatie

Analyse

Er is soms een gebrek aan datagedreven en 'evidence based' beleidscultuur

Entiteiten zijn nog te veel gericht op hun eigen data; delen van data wordt gezien als een gunst (ipv data als publiek goed)

Er is te weinig interne en interbestuurlijke samenwerking

Cultuur en organisatie

Aanbevelingen

Wijzig de attitude van 'data is power' naar '*data sharing is even more power*'

Versterk de data governance:

- Maturiteitsbeoordelingen
- Een VO-brede datastrategie + in elk beleidsdomein
- Chief Data Officers/Policy Scientists

Investeer in datageletterdheid, op alle niveaus



Betere overheidsdata voor sterk beleidsonderzoek
Sterk beleidsonderzoek voor solide evidence
Solide evidence voor goed bestuur



Betere overheidsdata voor sterk beleidsonderzoek en goed bestuur

Hoorzitting

Commissie voor Algemeen
Beleid, Financiën,
Begroting en Justitie

22 februari 2022