

## **Sammenhængen mellem smitterelevante adfærd og reproduktionstallet**

HOPE-projektet ([www.hope-project.dk](http://www.hope-project.dk))

Frederik Juhl Jørgensen & Michael Bang Petersen

21. marts 2021

*Analyserne i dette notat er ikke fagfællebedømt. Derudover tager de ikke højde for en række metodiske forhold og bør derfor i udtalt grad anses for foreløbige.*

## Indledning

Borgernes adfærd er central for smittespredning under corona-pandemien<sup>1</sup>, og en væsentlig faktor for at forstå smittespredningen er således tal og data for befolkningens adfærd. En udfordring i den forbindelse er, at borgernes smitteforebyggende adfærd i høj grad er privat adfærd. Afholder man sig fra at mødes med vennerne, og afspritter man sine hænder? Adfærdstracking ved hjælp af mobiltelefonernes gps-udstyr kan give nogle oplysninger. Men de data fortæller os mest om, hvordan folk bevæger sig rundt i samfundet. Og selvom man bevæger man sig meget rundt, så smitter man ikke, hvis det primært foregår i en bil med vaskede hænder.

En alternativ datakilde er befolkningens egne svar, indsamlet igennem spørgeskemaundersøgelser. Igennem simpel tilfældig udvælgelse kan man spørge et repræsentativt udsnit af befolkningen. Og igennem vel designede spørgsmål kan man måle privat adfærd såsom hyppighed af håndvask, antallet af kontakter og graden af fysisk afstand. Endelig har spørgeskemaer en anden fordel. Brugen af data fra Facebook, Google og telefoner kan gøre folk bekymrede. Opbevares data sikkert, og værnes der tilstrækkeligt om privatlivets fred?

Men spørgeskemaer kommer også med deres egne udfordringer. For kan man nu stole på, hvad folk siger, og kan de overhovedet huske, hvad de foretog sig i går? Det er væsentlige spørgsmål, men der er mindst to forhold, som indikerer, at vi kan trække væsentlige signaler ud af folks støjfyldte svar. For det første forsvinder en stor del af støjen ved aggregering. Tilfældig støj i de enkeltes svar udligner hinanden, når man tager et gennemsnit. For det andet er vi interesserede i ændringer. Er folk blevet mindre opmærksomme på sundhedsrådene, og holder de mindre afstand? Så længe støjen er konstant over tid, så vil sådanne ændringer i data give os et klart signal om ændringer i befolkningen.

I dette notat undersøger vi i hvilken grad selvrapporterede surveydata for befolkningens adfærd hænger sammen med smitteudviklingen målt ved hjælp af kontakt- eller reproduktionstallet, der fortæller, hvor mange personer en smittet person i gennemsnit smitter. De præsenterede analyser skal ses som foreløbige, forsimplede illustrationer, og der er en række metodiske udfordringer ved tidsserieanalyser, som nærværende analyser ikke tager højde for. Analyser indikerer således kun et sammenfald mellem data for adfærd og data for smittens udvikling og altså ikke en årsagsrelation, trods at der er stærke teoretiske grunde til at antage, at der i dette tilfælde bør være en sådan.

---

<sup>1</sup> Anderson, R. M., Heesterbeek, H., Klinkenberg, D., & Hollingsworth, T. D. (2020). How will country-based mitigation measures influence the course of the COVID-19 epidemic?. *The Lancet*, 395(10228), 931-934.

## Metode

Siden maj 2020 har HOPE-projektet fulgt danskernes adfærd ved hjælp af daglige surveys sendt ud via eBoks i samarbejde med Kantar Gallup. Data indsamles ved stratificeret udvælgelse på køn, alder og uddannelse via et udtræk fra CPR-registeret leveret af Danmarks Statistik. Det sikrer, at resultater kan anses som repræsentative for den danske befolkning generelt i forhold til ovennævnte karakteristika. Vi anvender post-stratificeringsvægte til at tage højde for tilbageværende mindre ubalancer i vores data.

I nærværende notat anvender vi konkret fire spørgsmål angående danskernes smitterelevante kontakter, der dækker over kontakt med familie, kontakt med kollegaer, kontakt med venner og bekendte samt kontakt med fremmede. Spørgsmålsformuleringerne fremgår nedenfor.

**Tabel 1. Kontaktspørgsmål**

Kontekst	”Vi er interesserede i at høre hvor mange mennesker du har været fysisk tæt på inden for de seneste 24 timer. Fysisk tæt skal her forstås som at du har været indenfor 1 meters afstand i mindst 15 minutter. Giv os venligst dit bedste bud.”  Angiv venligst 0, hvis du ikke var tæt på nogen
Familie	Hvor mange fra din familie som du ikke bor sammen med har du været fysisk tæt på?
Kollegaer	Hvor mange kollegaer har du været fysisk tæt på?
Venner	Hvor mange venner og bekendte (mennesker du kender navnet på) har du været fysisk tæt på?
Fremmede	Hvor mange, som du ikke allerede kender, har du været fysisk tæt på? (eksempelvis i offentlig transport, på legepladser, i supermarkedet)

På baggrund af ovenstående spørgsmål aggregerer vi hvert kontaktmål samt et samlet mål for total antal kontakter på dagsniveau for at danne et billede af hvordan danskernes kontaktmønstre har udviklet sig over tid. Vi topkoder de øverste 0,5 procent af besvarelserne fra for at undgå at ekstreme besvarelse ”trækker” for meget i de estimerede daglige gennemsnit. Derudover standardiserer vi hvert enkelt mål for sammenlignelighedens skyld. Dette datasæt fletter vi sammen med data for reproduktionstallet leveret af SSI.<sup>2</sup> Det giver os mulighed for at sammenholde udviklingen i danskernes kontaktmønstre med reproduktionstallet, og dermed studere hvorvidt kontaktmønstre –

---

<sup>2</sup> For metode bag beregning af kontakttallet, se Statens Serum Institut (2020). *Incident og fremskrivning af COVID-19 tilfælde*. Ekspertrapport, <https://covid19.ssi.dk/-/media/ssi-files/ekspertrapport-af-den-23-oktober-2020-incidens-og-fremskrivning-af-covid19-tilfælde.pdf?la=da>.

observeret via surveys – kan hjælpe til at prædiktere (den fremtidige) udviklingen i COVID-19 epidemien.

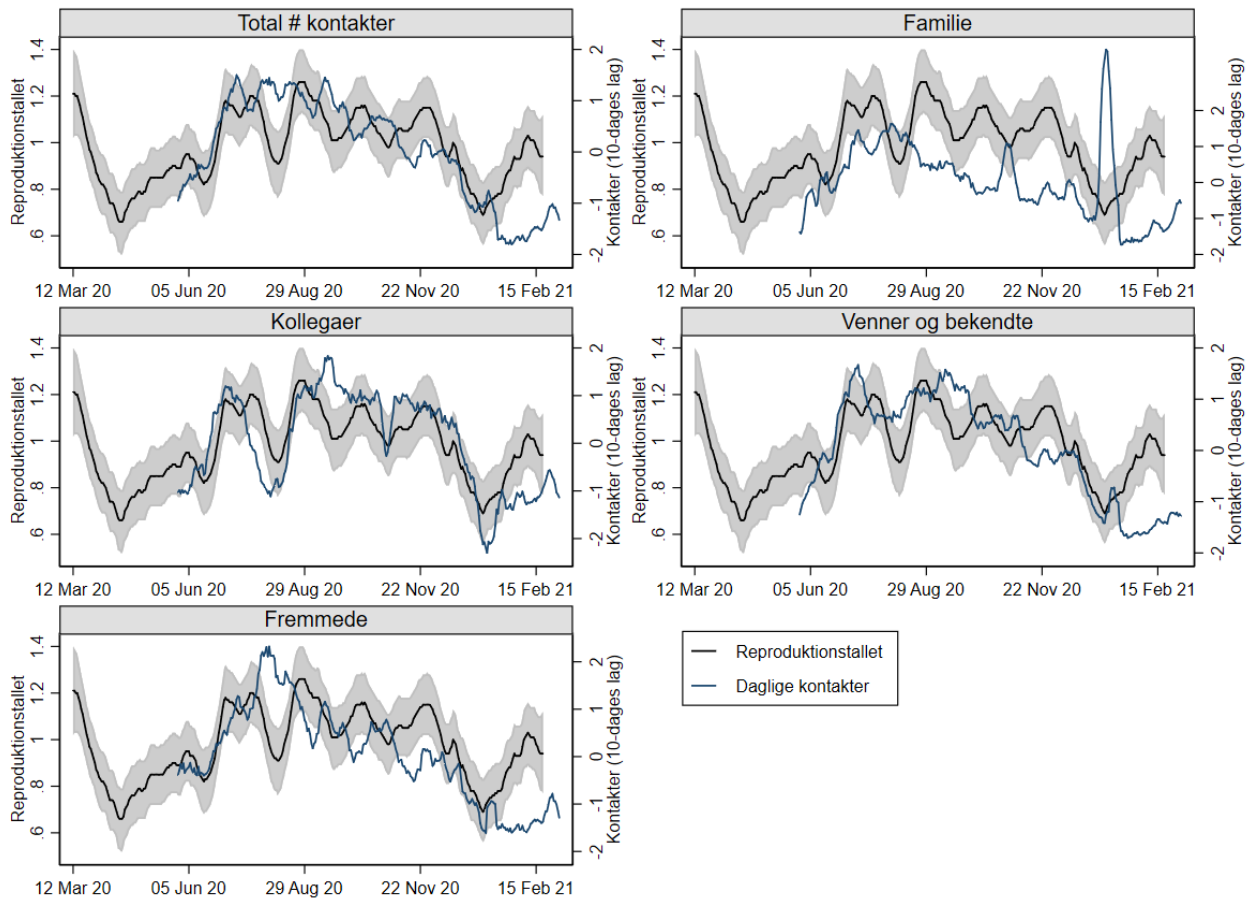
## Resultater

I figur 1 plotter vi tidsserierne for det totale antal kontakter og hvert af de underliggende kontaktmål sammen med tidsserien for reproduktionstallet. Når vi kigger henover figuren, ser vi, at særligt kontakt med antal kollegaer følger udviklingen i reproduktionstallet meget tæt. Bemærk at kontaktmålstidsserierne alle er laggede med 10 dage. Hvis vi kigger på den ”rå” korrelation mellem ”kontakt med kollegaer” og reproduktionstallet er denne helt oppe på 0.74, hvilket må betegnes som meget højt. Derimod er korrelation med ”kontakt med familie” helt nede på 0.14. Disse korrelationer skal dog fortolkes med stor varsomhed.<sup>3</sup>

---

<sup>3</sup> Johansen, S. (2007). Correlation, regression, and cointegration of nonstationary economic time series. *CREATES Research Paper*, (2007-35).

**Figur 1. Tidsserier: kontaktmål og reproduktionstallet**



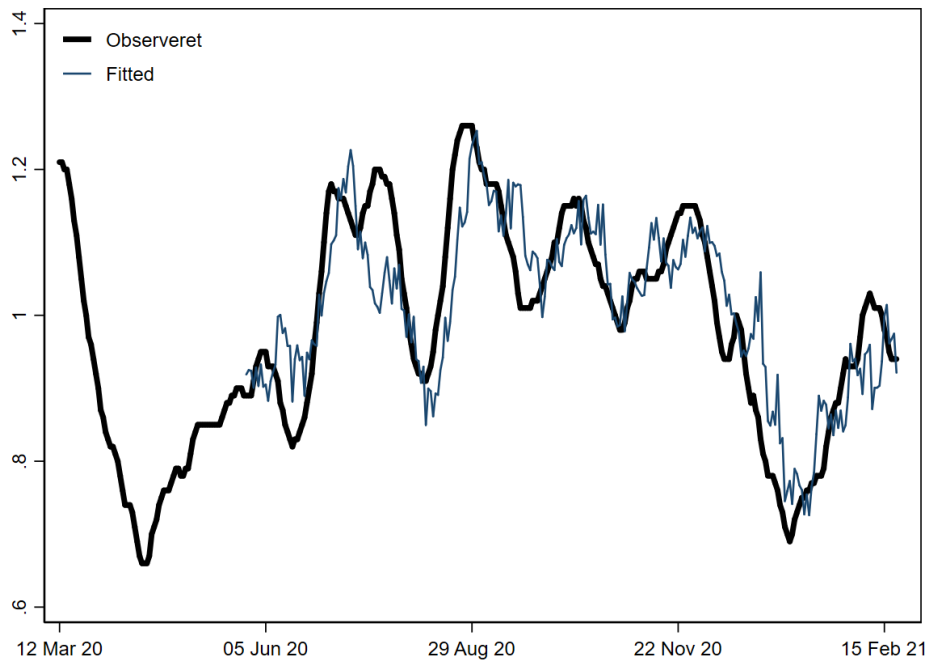
Ovenstående illustrerer altså, at der kan være et potentiale i at anvende i at udnytte surveydata over smitterelevante kontakter til at forstå, hvordan epidemien udvikler sig. For at illustrere dette potentiale yderligere, bevæger vi os i Tabel 2 videre til at anvende udviklingen i ”kontakt med kollegaer” til at prædiktere udviklingen i reproduktionstallet i forholdsvist simple regressionsmodeller. For at få et benchmark at evaluere anvendeligheden af kontaktmålet mod, anvender vi først en model, hvor vi bruger niveauet samt ændringen i reproduktionstallet med et lag på 10 dage til at forudsige udviklingen i reproduktionstallet over tid (Model 1). Det vil med andre ord sige, at niveauet i reproduktionstallet i dag samt ændringen fra i går til i dag bruges til at forudsige niveauet 10 dage frem i tid. Som det fremgår af Tabel 2 forklarer denne model 56 % af den samlede (”in-sample”) varians i reproduktionstallet. I Model 2 anvender vi udelukkende niveauet i ”kontakter med kollegaer” også med et 10-dages lag til at forudsige reproduktionstallet. Som det fremgår i Tabel 2 kan denne model forklare 55 % af den samlede (”in-sample”) varians i reproduktionstallet. Forskellen mellem Model 1 og Model 2 er ikke-signifikant. Med andre ord vil det sige, at niveauet af kontakter

med kollegaer forudsiger reproduktionstallet 10-dage frem i tid lige så godt, som reproduktionstallet i dag samt ændringen i tallet fra i går til i dag gør. Bevæget vi os videre til Model 3, hvor vi inkluderer både niveauet og ændringen i reproduktionstallet samt antal kontakter med kollegaer i samme model, ser vi, at modellen kan forklare 70 % af den samlede ("in-sample") varians i reproduktionstallet. Denne stigning er statistisk signifikant sammenlignet med både Model 1 og 2. I Figur 3 illustrerer vi forudsigelserne omkring udviklingen i reproduktionstallet baseret på Model 3. Den optrukne sorte linje markerer den observerede udvikling i reproduktionstallet (baseret på SSI estimer), mens den blå linje markerer den forudsagte udvikling baseret på Model 3.

**Tabel 2. Regressionsmodeller**

	Model 1	Model 2	Model 3
Reproduktionstallet	0.63		0.33
(10-dages lag)	(15.66)		(7.76)
$\Delta$ Reproductionstallet	3.30		2.79
(10-dages lag)	(9.35)		(9.42)
Kollega kontakt (10-dages lag)		0.10	0.07
		(17.93)	(11.14)
Konstant	0.38	1.01	0.67
	(9.30)	(181.03)	(15.72)
Adj. $R^2$	0.56	0.55	0.70
RMSE	0.09	0.09	0.07

**Figur 2. Modelforudsigelser**



## Konklusion

Ovenstående resultater indikerer et potentiale i at udnytte surveydata over smitterelevante kontakter til at forstå, hvordan epidemien udvikler sig. Det er imidlertid vigtigt at pointere, at man skal være forsigtig med fortolkningerne af ovenstående fund, da de baserer sig på basale modelleringer og udelukkende fokuserer på at forudsige udviklingen ”in-sample”. Dermed er der stor usikkerhed om, i hvor høj grad kontaktmålene kan prædiktere den *fremtidige* udvikling i epidemien. Vi vil derfor understrege, at ovenstående er ”work-in-progress”, og skal forstås som et første udkast, der deskriptivt viser, at der kan være potentiale i at udnytte surveydata over smitterelevante kontakter til at forstå, hvordan epidemien udvikler sig. Vi samarbejder i øjeblikket med Sune Lehmann, Andreas Koher, Søren Johansen og Katarina Juselius om at udnytte dette potentiale.