

# Život během pandemie: jak pracovat s datovými soubory

Webinář Sociálněvědní data k tématu  
pandemie v Česku a na Slovensku



Michaela Kudrnáčová



29. 11. 2021

**PAQR**  
RESEARCH

# Obsah příspěvku

## Téma

## Strana

Metodologie a obsah výzkumu Život během pandemie

3

Informace k dostupným datovým souborům a doporučení k práci s nimi

11

1.

**O výzkumu**

**Život během pandemie**

# Život během pandemie

- \* **Online panelové** šetření dospělé populace ČR probíhající pravidelně **od 18. března 2020**
- \* Do konce listopadu 2021 realizováno **34 vln** dotazování
- \* Aktuálně připravovány **další minimálně 4 vlny**
- \* Obsah připravují **PAQ Research** (původní iniciátor) a **CERGE IDEA AntiCovid** (většinové financování)

## Cílem výzkumu je:

- \* zkoumat **epidemiologicky významné chování** (kontakty, sociální aktivity, testování, očkování) a **dopady pandemie** na pracovní život, ekonomiku domácností i zdraví

## Funkce:

- \* Monitorování trendů, aktuální informace pro veřejnost, média a tvůrce veřejných politik (PAQ Research)
- \* Akademický výzkum (CERGE IDEA)
- \* Modelování průběhu epidemie (BISOP)

# Kde najít výstupy a další informace o projektu?

## \* Datový dashboard:

[www.zivotbehempandemie.cz](http://www.zivotbehempandemie.cz)

## \* Tematické zprávy:

<https://www.paqresearch.cz/research-blog>

## \* z nedávných – odložené zdravotní dopady epidemie, bariéry ochoty nechat se očkovat

## \* připravované – ochota k přeočkování 3. dávkou a očkování dětí

### Život během pandemie

#### EKONOMICKÉ DOPADY

- Destabilizace práce
- Zasažení domácností
- Strategie domácností

#### CHOVÁNÍ V EPIDEMII

- Sociální aktivity
- Kontakt s lidmi
- Home office
- Testování
- Očkování

#### AKTIVITY CHRÁNÍCÍ PŘED NÁKAZOU

- Počet aktivit
- Jednotlivé aktivity
- Kontakty vs. aktivity
- Imunizace

#### OBAVY A DUŠEVNÍ ZDRAVÍ

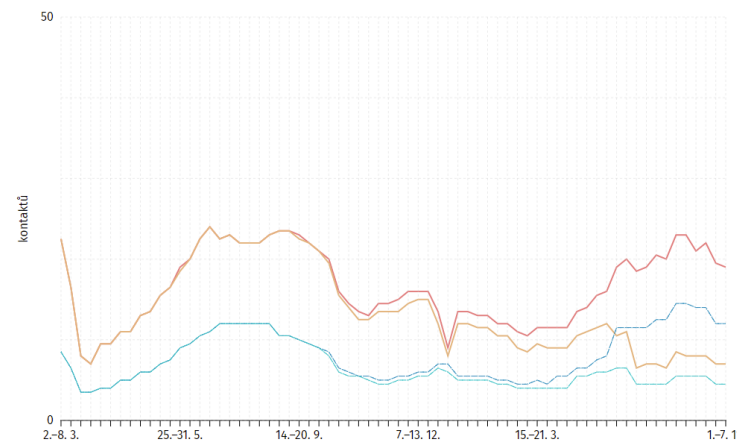
- Obavy z epidemie
- Obavy ze ztráty práce
- Duševní zdraví

## S kolika lidmi byli respondenti v bližším kontaktu?

Na jaře 2020 se podařilo snížit průměrné týdenní počty blízkých kontaktů z běžných 23 na 7, tedy zhruba na třetinu. Během dalších vln a lockdownů se jednalo o zhruba 11 až 13 kontaktů. Od poloviny letošního dubna začal postupný růst až na 23 přímých osobních kontaktů týdně během září. Listopadové hodnoty však opět poklesly na 19 přímých kontaktů týdně. Přesto v porovnání s ostatními vlnami epidemie kontakty klesají pomaleji. Část populace je ale díky očkování nebo v důsledku prodělání nemoci imunní. Průměrný odhadovaný počet kontaktů, ve kterých ani jedna strana nemá prokazatelnou plnou imunitu tak klesl na 7 v průběhu léta. Reálná míra nechráněných kontaktů tak byla na úrovni nejtěžšího lockdownu. Následoval nárůst na 8,5 kontaktů začátkem září a s rostoucím počtem osob, které prodělaly covid-19 v posledních třech měsících (a tudíž mají přirozené protilátky) došlo v listopadu k poklesu až na 7 kontaktů týdně, kde není ani jedna strana prokazatelně imunní.

Podívejte se na [interpretaci](#) a [metodické poznámky](#).

Souhrnné zobrazení Jak si vedou různé skupiny obyvatel?



pro zvýraznění jednotlivých kategorií, najděte na název kategorie

#### Průměrný počet osob v podstatném kontaktu

Průměrný odhadovaný počet lidí, se kterými byli respondenti v daném týdnu v přímém osobním kontaktu trvajícím alespoň pět minut.

#### Průměrný počet kontaktů, kde ani jeden není imunní

Průměrný odhadovaný počet osobních kontaktů, ke kterým dochází mezi osobami bez imunity vůči covid-19. Zohledňujeme, zda je respondent očkováný druhou dávkou v posledních šesti měsících či nemoc prodělal v posledních třech měsících, a u neimunních respondentů, že plnou imunitu mají jeho kontakty (METODIKA ilustrativního výpočtu viz níže).

#### Průměrný počet delších kontaktů bez roušky

Průměrný odhadovaný počet lidí, s nimiž měl respondent v daném týdnu osobní kontakt bez roušky/respirátoru trvajícím alespoň patnáct minut (v části zkoumaného období modelované hodnoty; viz METODIKA níže).

# Jak se sbírají data?

- \* Sběr dat zajišťuje agentura NMS Market Research
- \* Dotazování formou **CAWI** na stálé skupině respondentů
- \* Dotazování v rámci jedné vlny probíhá **přibližně jeden týden** (většina dotazníků během prvních dvou dnů)
- \* Vlny následují **dva až pět týdnů** po sobě (do června 2021 rozestup dva až tři týdny, poté delší intervaly)

## Kdo jsou respondenti?

- \* Členové **Českého národního panelu**
- \* Ve věku **18 let a starším**
- \* Český národní panel je společný projekt agentur Nielsen Admosphere, NMS Market Research a STEM/MARK, který představuje **online panel registrovaných respondentů** účastnících se výzkumů trhu a veřejného mínění

# Jak byl sestaven vzorek a jak se vyvíjí?

- \* **V první vlně kvótní výběr** respondentů Českého národního panelu (**online populace** => obezřetnost v závěrech o populaci 55+)
- \* Kvóty: pohlaví a věk, vzdělání, kraj, velikost místa bydliště
- \* **Nadvýběr** velkých měst (50 000 a více obyvatel) a Hlavního města Praha (200 %)
  - \* Statisticky korigováno post-stratifikačními vahami => **kvótně reprezentativní na dospělé populaci ČR**
- \* V první vlně v březnu 2020 dotázáno **3101** respondentů, poté dlouhodobě cca **2200 až 2400 respondentů**, v posledních vlnách 30 až 34 z podzimu 2021 1700 až 1900 respondentů
- \* Řada **strategií udržování retence**: avizování, pravidelné upomínky během sběru, poděkování, nadstandardní odměny
- \* V dalších vlnách oslovení jen **účastníci první vlny**, bez doplňování vzorku
- \* Respondenti první vlny se mohou zúčastnit kterékoli následné vlny, aniž by se museli zapojit do všech předchozích => množství kombinací vln, v nichž se respondenti zúčastnili

# Co je obsahem šetření?

## Dotazník obsahuje dva stálé moduly:

- \* Epidemiologický modul
- \* Ekonomický modul

⇒ **kontinuální část** modulů zobrazena na [www.zivotbehempandemie.cz](http://www.zivotbehempandemie.cz)

## Občasně jsou zařazovány také ad-hoc moduly, např.:

- \* Volební chování
- \* Energetická krize
- \* **Typicky nejsou součástí veřejných dat** s výjimkou:
  - \* Postoje k přistěhovalcům (SOÚ) v 27. vlně
  - \* Důvěra veřejným institucím (CVVM SOÚ) v 28. vlně

⇒ **Možno si zadat vlastní tematický modul**



# Epidemiologický modul

## Kontinuálně:

- \* **Počet kontaktů** v posledních 2 týdnech
  - \* Podstatné kontakty (po dobu alespoň 5 minut)
  - \* Delší kontakty alespoň 15 minut bez roušky (od 20. vlny)
- \* **Kontakt s nakaženým**
- \* **Sociální aktivity** v posledních 2 týdnech (nakupování, ježdění MHD, návštěva známých, návštěva restaurace, cestování atd.)
- \* **Preventivní opatření** (pečlivé mytí rukou, nošení ochrany dýchacích cest, vyhýbání se přelidněným místům...)
- \* **Symptomy** nemoci respondentů a členů rodiny
- \* Absolvování **testu** (čas, typ testu, výsledek)
- \* **Očkování** (absolvování, ochota)
- \* **Docházka na pracoviště**

## Nepravidelně:

- \* **Postoje k vládním opatřením**
- \* **Motivace a bariéry očkování**
- \* Narativy o očkování
- \* **Odkládání zdravotních prohlídek**

# Ekonomický modul

- \* Změny a omezení **pracovní aktivity** v souvislosti s pandemií
- \* Počet **odpracovaných hodin** v posledních 2 týdnech
- \* Obavy ze ztráty práce
- \* **Hospodaření domácnosti** (příjem, problémy s platbami, strategie řešení finančních potíží)

**2.**

# Práce s datovými soubory

# Jaké datové soubory jsou v ČSDA dostupné?

## Průřezové soubory

- \* soubory s proměnnými dotazovanými v dané vlně (aktuálně vlny 1 až 26, připravovány 27 až 34)

## Long spojený soubor

- \* respondent zahrnut tolikrát, kolika vln se zúčastnil (unikátní identifikací je kombinace ID respondenta a čísla vlny)
  - \* panelové regrese
  - \* časové řady po vlnách (křížení proměnná X vlna)

## Wide spojený soubor

- \* proměnná zahrnuta tolikrát, v kolika vlnách dotázána (unikátní identifikací je ID respondenta, vlnu proměnné odlišuje předpona vXX před kódem proměnné)
  - \* kontingenční tabulky proměnných v různých vlnách

**Dostupný je script pro převod mezi formáty (long => wide, wide => long, long => cross-sectional, wide => cross-sectional)**

# Na co si dát pozor?

## U některých proměnných se mezi vlnami měnily parametry

- \* Zejména v prvních vlnách úpravy ve znění otázek, kategoriích odpovědí, definicích filtrů => může vyžadovat rekódování, zohlednění při interpretaci výsledků, vypuštění určitých vln z analýz
  - \* **Doporučení:** zkontrolovat parametry v jednotlivých vlnách v **PŘEHLEDU OTÁZEK**
  - \* Pozor, i v průřezových datech labely odpovídající nejaktuálnější vlně

## Část proměnných dotazována jen v některých vlnách

- \* Některé proměnné z kontinuálních modulů nejsou zařazeny ve všech vlnách (některé otázky doplněny později, některé po čase vypuštěny – např. frekvence aktivit respondenta v týdnu 1 a aktivity členů domácnosti v týdnu 1)
  - \* **Doporučení:** zkontrolovat výskyt ve vlnách v **SEZNAMU PROMĚNNÝCH**

## Otázky mají různá referenční období, která se mohou mezi vlnami měnit

- \* U **testování** vždy celé období od posledního sběru (dva až pět týdnů)
- \* U **kontaktů, sociálních aktivit, home office** atd. vždy poslední dva týdny

# Jak byly připraveny spojené soubory?

- \* **Není běžná praxe** poskytovat spojené datové soubory z panelových šetření => Neexistuje jedno standardní „správné“ řešení, jak spojené soubory připravit.
- \* Zvoleno „**minimalistické řešení**“ tak, aby spojené soubory:
  - \* dávaly **celkový přehled** o obsahu šetření (není nutné prohlížet desítky souborů samostatně a propojovat vlny),
  - \* **nenarušila se původní struktura** průřezových dat (omezením spojeného souboru na příslušnou vlnu získáme původní průřezová data),
    - \* Hodnoty proměnných obsaženy jen ve vlnách, kde se dotazovaly – tj. žádné doplňování hodnot z některých vln do jiných.
    - \* Nutné vést v patrnosti **zvláště při práci s long souborem** (např. základní sociodemografika jako pohlaví, vzdělání, kraj, velikost místa bydliště byla zjišťována pouze v první vlně, tedy jejich hodnoty jsou uvedeny jen v datech z první vlny)
  - \* byly **ponechány původní hodnoty** z průřezových dat (tj. není provedeno žádné čištění odpovědí z jedné vlny na základě odpovědí z jiné vlny).
- \* **Váhy:** post-stratifikační váhy doplněny **z průřezových dat**, nejsou připraveny panelové váhy

# Kontakt



**Autoři**

**Michaela Kudrnáčová** / [michaela.kudrnacova@paqresearch.cz](mailto:michaela.kudrnacova@paqresearch.cz)

[/ www.paqresearch.cz](http://www.paqresearch.cz)

