COVID-19 and the Future of US Fertility: What Can We Learn from Google?

Joshua Wilde (MPIDR)
Wei Chen (Fordham University)
Sophie Lohmann (MPIDR)
Introduction

The effect of the COVID-19 pandemic on births is – as of yet – unknown.

- Speculation of a positive effect ("Quaranteens"? "Coronials"?)
- Looking historically suggests a negative effect.
- Possible heterogeneous effects (increase in unintended or mistimed births, fall in planned births).
- Direction and magnitude unknown, since the births have not yet occurred.
Google has been used to predict population health, as well as demographic processes

- Google Flu, Dengue, Mental Health, Suicide, Well-being.
- Demographic outcomes (migration, sexual behavior, mortality).
In this paper, we develop a prediction model which uses Google search data to predict the effect of Covid-19 on future birth rates in the United States.

- Simple intuition: If searches for "Morning Sickness", "Missed Period", or "Am I pregnant?" suddenly increase, perhaps we would expect more births in 6-9 months.
Methods

Three steps:

1. Establish associations between search volumes and timing of births.
2. Cross-validate through out of sample tests, keep predictive words.
3. Use associations between selected words and births to forecast future births.
Two main sources of data:

1. Google Trends
   - Search Index, monthly, US state, Jan 2004 – present.
   - Topic (broad criteria) vs. Keyword (exact character string)

2. US Natality data from the National Center for Health Statistics (NCHS)
   - Total Births by subgroup (Education, age, parity, ethnicity, marital status, birthweight, etc.)
Data – Initial Keyword Set

40 keywords across 7 categories:

- **Control** (Composite)
- **Unplanned Pregnancy** (Emergency Contraception, Morning After Pill, Plan B, STI, Unprotected Sex, Online Dating (Composite)
- **Pregnancy Intention** (Clearblue, Conceive, hCG, IVF, Ovulation, Ovulation Test, Pregnancy Test)
- **Pregnancy Symptoms** (Missed Period, Morning Sickness, Pregnancy Symptoms)
- **Pregnancy Termination** (Baby Heartbeat, Cytotec, Dilation and Curettage, Medical Abortion, Miscarriage, Misoprostol, Pregnancy Bleeding)
- **Unemployment** (File Unemployment How, Furlough, Layoff, Unemployment, Unemployment Office)
- **Other** (Birth Control, Divorce, IUD, Porn, Pregnancy, Pregnant)
Establish associations between keyword search volumes and births.

\[ Y_{smy} = \alpha_{sm} + \sum_{w} \sum_{L=t_0}^{T} \beta_{w,t-L} I_{smy}^w + \sum_{L=t_0}^{T} \omega_{t-L} U_{smy}^{t-L} + \sum_{s=1}^{S} \gamma_s \ast t^S + \epsilon_{smy} \]

where \( s = \text{state}, m = \text{month}, y = \text{year} \), \( \alpha_{sm} \) is a state-month fixed effect, \( \sum_{s=1}^{S} \gamma_s \ast t^S \) is a state specific time polynomial, and \( I_{smy}^w \) is the natural log of the normalized search volume for different keywords \( w \) at a number of monthly time lags \( t_0 \) to \( T \). \( Y \) is the natural log of births, and \( U_{smy} \) is the unemployment rate. Here \( t_0 = 1 \), \( T = 12 \), and \( S = 2 \).
Word Effects – Pregnancy Deterrents

Birth Control

%Δ in Births

-2%  0%  2%

t-12  t-10  t-8  t-6  t-4  t-2
Word Effects – Pregnancy Termination

Misoprostol (Cytotec)

%Δ in Births

-1% - 0% - 1%

t-12 t-10 t-8 t-6 t-4 t-2
Word Effects – Unemployment

Unemployment

%Δ in Births

% 0% 2%

-2%

t-12 t-10 t-8 t-6 t-4 t-2

Month Lag
Step Two: Cross-Validation

Keyword Set "Significance":

1. Keep keywords which are both:
   1.1 Significant at the 5% level for at least one lag between 7 and 12
   1.2 Jointly significant at the 5% level for lags 7-12.

2. Repeat 1 until both criteria are met for all words.

Keyword Set "MSPE":

1. Use a machine learning technique called Forward Stepwise Selection to select keyword with highest predictive power (by MSPE reduction) across all iterations of a Leave-2 years-out (L2YO) cross validation technique.

2. Stop selecting words when the marginal reduction in MSPE is less than 1%.

Both methods yield a reduction in out-of-sample MSPE of 25%, with 14 and 6 words respectively.
Step Three: Predict

- Apply estimated coefficients from the model with lags 7 to 12 to current Google search volumes.
- Can predict 7 months into the future.
Search Volumes and Covid-19

How did the pandemic effect search volumes for keywords?

1. Internet usage rose.
2. No effect – or small decline – for words related to pregnancy intention.
3. Large drop in searches related to clinical services.
4. Huge increase in searches for unemployment related search terms.
Word Searches – Example

Pregnancy Test

Search Volume Index

Filtered Data

Raw Data

April Fools!
Word Searches – General Internet Use

Control

Search Volume Index

Word Searches – Pregnancy Intention

Ovulation

Search Volume Index

Jan 2018
Apr 2018
Jul 2018
Oct 2018
Jan 2019
Apr 2019
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Oct 2020
Word Searches – Pregnancy Intention

Clearblue

Search Volume Index

Word Searches – Unemployment

Unemployment

Search Volume Index

[Graph showing search volume for unemployment from Jan 2018 to Oct 2020 with a significant spike in Apr 2020, indicating a rise in search interest.]
Main Results

Relative Birth Index, Different Keywords Sets

Predicted Births (Oct 2020=100)
Heterogeneous Results – Education

**Graph 1:**
- Title: Education
- X-axis: Birth Index
- Y-axis: Time (Jan 2020 to Jan 2021)
- Categories: Second or Less, Some Tert., Tert. or More

**Graph 2:**
- Title: Predicted Fertility Decline
- X-axis: Bachelor's Degree or Higher
- Y-axis: Predicted Fertility Decline

States represented by their initials:
- AL
- AZ
- CA
- CO
- CT
- DE
- FL
- GA
- HI
- IA
- ID
- IL
- IN
- KS
- KY
- LA
- MA
- MD
- MI
- MN
- MO
- MS
- NC
- NE
- NH
- NJ
- NV
- NY
- OH
- OK
- OR
- PA
- SC
- TN
- TX
- VA
- WA
- WI
- WV
Heterogeneous Results – Ethnicity

- White (Non-Hisp.)
- Black
- Hispanic

Predicted Fertility Decline

Fraction Black or African-American
Discussion

Is it believeable?

- 15% reduction consistent with effect after Great Depression and Spanish Flu Pandemic
- 50% larger than the effect of the Financial Crisis.
- Plausible that higher SES individuals better able to mitigate the effect of the shocks.
- Pandemic has had larger effects in the US among low SES individuals, both in terms of incidence/caseload as well as economic effects.

Does it matter?

- Postponement or Permanent? Probably postponement (results not shown).
- Unexpected infertility?
• Using Google search data on pregnancy and unemployment keywords, we find that US births are likely to fall 15% between Nov 2020 and Feb 2021.

• Heterogeneous effects by education and ethnicity – higher SES individuals are predicted to have smaller declines.

• Unclear how much it matters to long run fertility.
Comments or Questions?
www.demogr.mpg.de