



# Indagini sui social media: esempi e sfide sull'utilizzo di Facebook


FRANCESCO RAMPAZZO  
*Career Development Fellow  
in Marketing and Consumer Demography*

*Saïd Business School and  
Leverhulme Centre for Demographic Science,  
and Nuffield College  
University of Oxford*


Presentazione al Corso di 'Teoria e Tecnica dell'Indagine  
Statistica e del Campionamento' presso l'Università degli  
Studi di Padova, Maggio 2021


# Chi sono?


Francesco Rampazzo

 Career Development Fellow presso la **University of Oxford**

 PhD presso la **University of Southampton** ed il **Max Planck Institute for Demographic Research**

 European Doctoral School of Demography presso il **Max Planck Institute for Demographic Research** e l'**Università di Roma la Sapienza**

 Master in Multidisciplinary Demography presso la **Stockholms Universitet**

 Laurea in Statistica, Economia e Finanza presso l'**Università degli Studi di Padova**

# Di cosa mi occupo...



Mi interessa **combinare fonti tradizionali con nuove fonti di dati** per tematiche legate alla **fecondità, migrazioni e transizione all'età adulta**.

# Collaboratori e co-autori



Leah Rosenzweig



Katherine Hoffman Pham



Parrish Bergquist



Matto Mildenberg



André Grow



Daniela Perrotta



Emanuele Del Fava



Jorge Cimentada



Sofia Gil-Clavel



René Flores



Ilana Ventura



Ingmar Weber



Emilio Zagheni

Ci servono le **indagini**,  
le **interviste** nell'era in cui  
viviamo?

Scrivilo in chat.

- 1 La Rivoluzione Digitale
- 2 Le Indagini nell'Era del Digitale
- 3 Facebook per fare indagini
- 4 Sfide

# Rivoluzioni nella Storia



Agricultural  
Revolution



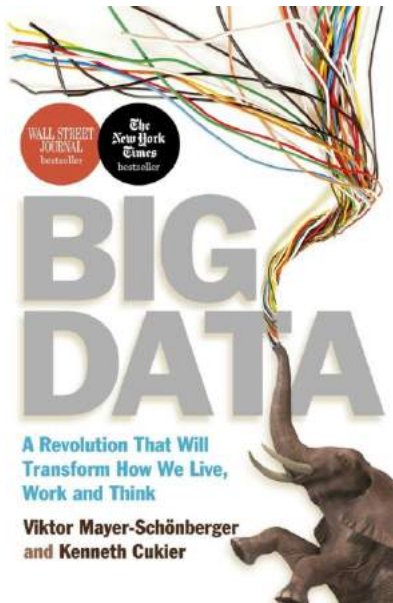
Industrial  
Revolution



Digital  
Revolution



# "Big Data"

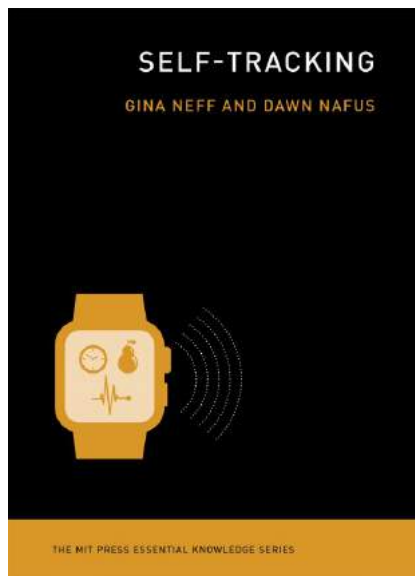


Rivoluzione in passaggi:

- 1 Digitalizzazione nelle nostre vite
- 2 Nuovi dati divengono accessibili

Dai nuovi dati nascono nuove ricerche e profitti per i mercati

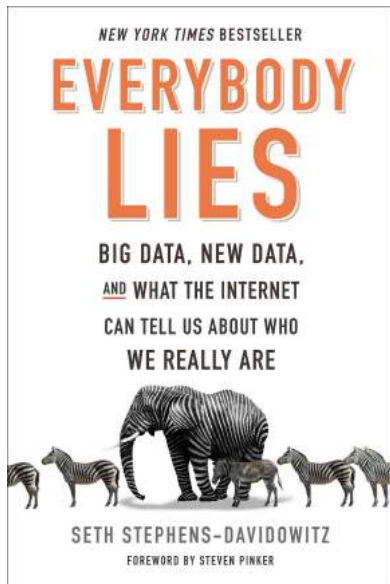




**Registriamo informazioni**  
come **ore dormite, passi**  
**fatti, calorie consumate,**  
**orari delle pastiglie...**

**Creiamo dati** che non erano  
disponibili in passato.

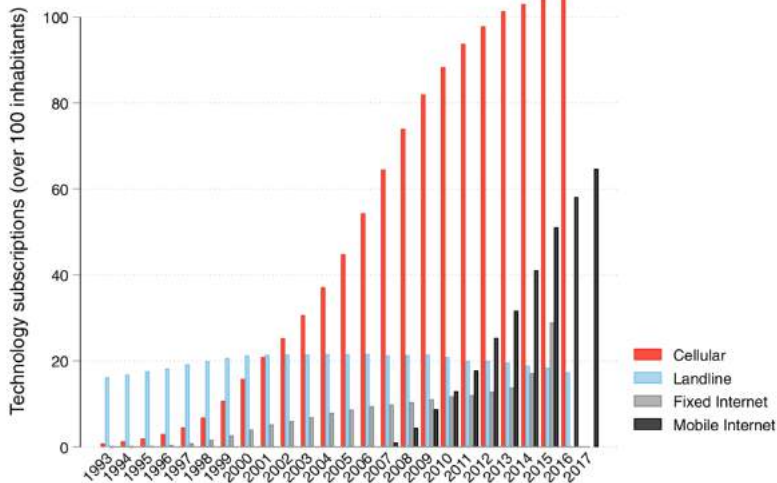
# Si può mentire a Internet?



Quando si risponde ad un **questionario** non si riporta sempre il vero in relazione a **comportamenti intimi o stigmatizzati**.

E' possibile che le digital traces mettano in risalto **comportamenti sommersi**.

# Technological Development



Rotondi et al. (2020) Figure 1 ICT penetration worldwide from 1993 to 2017.  
Source ITU data.

La **Rivoluzione Digitale** sta  
rendendo più **difficile** utilizzare  
i **vecchi metodi di**  
**campionamento**, ma allo  
stesso tempo ci sta dando  
**nuove opportunità.**

- Alto tasso di non risposta
- Costose
- Lente (in determinate situazioni)

# Caratteristiche delle Digital Traces



<https://www.bitbybitbook.com>

## Negative:

- incomplete
- inaccessible
- nonrepresentative
- drifting
- algorithmically confounded
- dirty
- sensitive

## Positive:

- big
- always-on
- non-reactive

Le digital traces sono **create e raccolte** da **compagnie private** e **governi** per scopi non di ricerca. Usare queste fonti di dati significa pensare a come **convertirle (repurposing)**.

## PNAS PNAS



David Garcia<sup>1,3</sup>, Yousuke Mitake Katsur<sup>1,2</sup>, Angel Corvaci<sup>1</sup>, Manuel Cobian<sup>1,2</sup>, Stefano Mero<sup>2,3</sup>, Iyad Rahme<sup>2,3,4</sup>, and Nelson Guevara<sup>4,5</sup>

David Garcia<sup>1,3</sup>, Yousuke Mitake Katsur<sup>1,2</sup>, Angel Corvaci<sup>1</sup>, Manuel Cobian<sup>1,2</sup>, Stefano Mero<sup>2,3</sup>, Iyad Rahme<sup>2,3,4</sup>, and Nelson Guevara<sup>4,5</sup>

[illegible]

Edited by Winston May, Cornell University, Ithaca, NY, and accepted by Editorial Board Members Mary E. Wilson May & C. 1988 (submitted to review October 14, 2007).

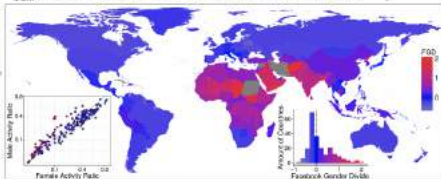


Fig. 1. The FGD across 237 countries. Countries are colored according their FGD from highly doored toward males (red) and doarded (blue) to highly doarded toward females (green) not doarded. Left hand shows the scatterplot of male and female activity values across all countries, revealing a spread around the diagonal. Right hand shows the histograms of FGD values in bins of width 0.2. While the mode of countries is slightly below zero, there is significant skewness toward high FGD values. An online interactive version of this figure can be found at <https://data-eq.github.io/AndersGGenderDavid/2020/eqgfgd.html>.

## Survey Recruitment

Article

Social Science Computer Review  
30(1), 141–154 (2012)  
© The Author(s) 2012  
Reprints and permission:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/0894412211416642  
journals.sagepub.com/home/sscr

Steffen Pötzschke<sup>1</sup> and Michael Braun<sup>1</sup>

Figure 1. Composition of a facebook advertisement with the objective 'lead people to your website'. Copyright of the picture: Igor Molnar fotolia.com.

## Gender, Migration, Male Fertility

Developed and developing countries



# Domande?

# Tre ere del campionamento

Groves, Robert M., 2011. **Three eras of survey research**. Public opinion quarterly, 75(5), pp.861-871.

	<b>Sampling</b>	<b>Interviews</b>
1 <sup>st</sup> era	Area probability	Face-to-face
2 <sup>nd</sup> era	Random digit dial probability	Telephone
3 <sup>rd</sup> era	Non-probability	Computer-administered

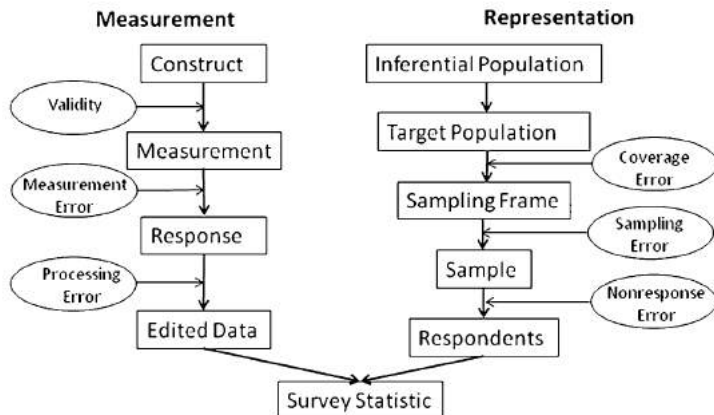
**Total Survey Error = Representation + Measurement**

**Rappresentazione** = problemi legati alla popolazione target

**Misura** = problemi legati a cosa si sta misurando

Errore può essere sistematico o casuale, osservabile e non osservabile.

# Total Survey Error Framework



Groves and Lyberg (2010) Figure 3. Total Survey Error Components Linked to Steps in the Measurement and Representational Inference Process (Groves et al. 2004).

# Total Error with Big Data

Amaya A., Biemer P.P., and Kinyon D. (2020). **Total Error in a Big Data World: Adapting TSE Framework to Big Data.** Journal of Survey Statistics and Methodology 8, 89-119.

Error components	Error causes
Coverage error	Undercoverage, overcoverage, duplicates
Sampling error	Large, but non probabilistic samples
Specification error	Not clear definitions
Nonresponse/missing data error	Missing data counfounded with undercoverage
Measurement/content error	Selection in what we see
Processing error	Download and cleaning phases
Modeling/estimation error	Unknown errors
Analytic error	Interpretation

# **Pubblicità!!!**

## **Due passaggi:**

- Quota Sampling
- Post-Stratification

# Quota Sampling

*Political Science Research and Methods* (2018), 8, 558–568  
doi:10.1017/pers.2018.40



## RESEARCH NOTE

### Quota sampling using Facebook advertisements

Baohao Zhang<sup>1\*</sup>, Matti Mildenberger<sup>1</sup>, Peter D. Howe<sup>2</sup>, Jennifer Macken<sup>3</sup>, Seth A. Rosenthal<sup>4</sup>  
and Anthony Leiserowitz<sup>5</sup>

<sup>1</sup>Department of Political Science, Yale University, New Haven, CT, USA; <sup>2</sup>Department of Political Science, University of California, Santa Barbara, Santa Barbara, CA, USA; <sup>3</sup>Department of Environment and Society, Utah State University, Logan, UT, USA and <sup>4</sup>School of Forestry and Environmental Studies, Yale University, New Haven, CT, USA

\*Corresponding author. Email: baohao.zhang@yale.edu

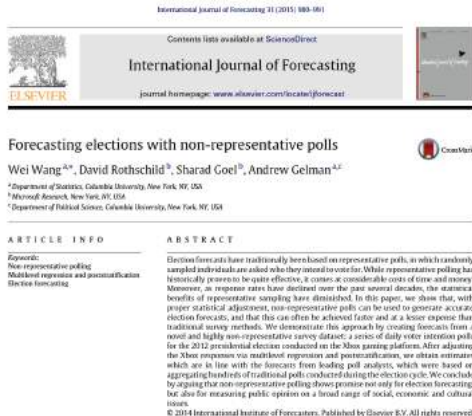
#### Abstract

Researchers in different social science disciplines have successfully used Facebook to recruit subjects for their studies. However, such convenience samples are not generally representative of the population. We developed and validated a new quota sampling method to recruit respondents using Facebook advertisements. Additionally, we published an R package to semi-automate this quota sampling process using the Facebook Marketing API. To test the method, we used Facebook advertisements to quota sample 2432 US respondents for a survey on climate change public opinion. We conducted a contemporaneous nationally representative survey asking identical questions using a high-quality online survey panel whose respondents were recruited using probability sampling. Many results from the Facebook-sampled survey are similar to those from the online panel survey; furthermore, results from the Facebook-sampled survey approximate results from the American Community Survey (ACS) for a set of validation questions. These findings suggest that using Facebook to recruit respondents is a viable option for survey researchers wishing to approximate population-level public opinion.

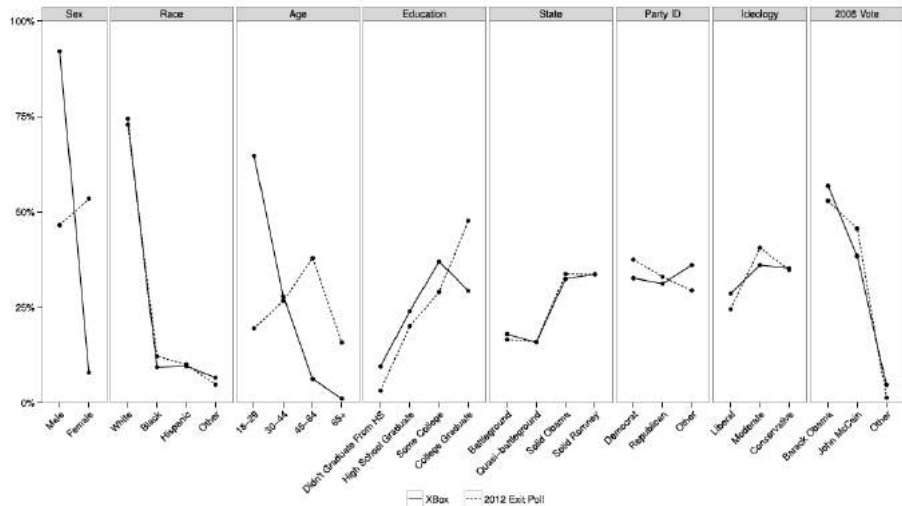
**Non** è possibile avere un **campione casuale** sulle piattaforme online (a meno che non si collabori con la piattaforma stessa).

**Quota Sampling** una versione del stratified random sampling tuttavia **non** è un **campione probabilistico**.

Modelli per  
correggere le  
stime attraverso  
pesi che  
comparano i gruppi  
del campione con  
la popolazione di  
riferimento.

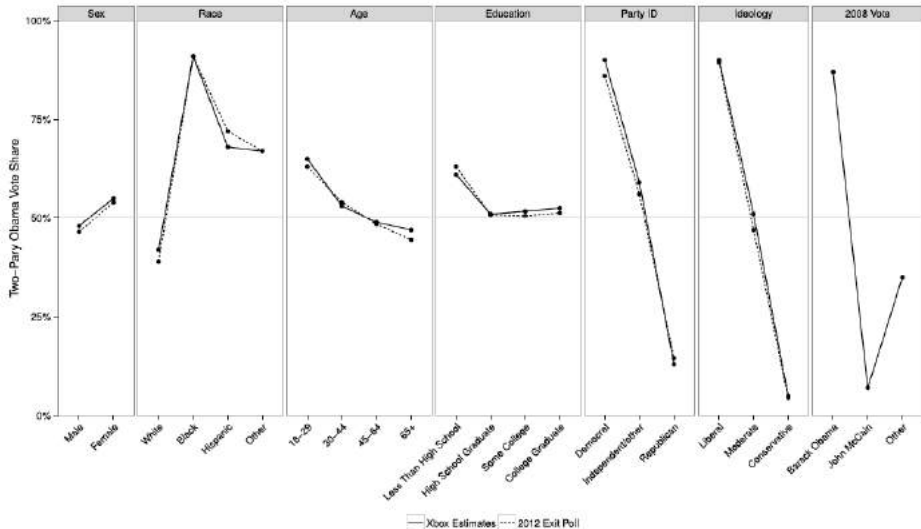






A comparison of the demographic, partisan, and 2008 vote distributions in the Xbox dataset and the 2012 electorate (as measured by adjusted exit polls). As one might expect, the sex and age distributions exhibit considerable differences.

# Xbox – Post-Stratified



Comparison of the two-party Obama vote share for various demographic subgroups, as estimated from the 2012 national exit poll and from the Xbox data on the day before the election.

Con le **giuste accortezze**  
**campioni non-casuali** possono  
essere usati per **stimare quantità**  
**generalizzabili all'intera**  
**popolazione.**

Domande?

# **Possiamo usare Facebook per condurre indagini?**

# Facebook Advertising Platform

<https://www.facebook.com/adsmanager>



Add locations in bulk

## Age

18 - 65+

## Gender

All genders

## Detailed targeting

Include people who match ⓘ

[Behaviours > Ex-pats](#)

Lived in Poland (formerly Ex-pats – Poland)

## Audience definition



Your audience selection is fairly broad.

Potential reach: 380,000 people ⓘ

## Estimated daily results

Reach ⓘ

3.9K-11K



The accuracy of estimates is based on factors such as past campaign data, the budget you entered, market data, targeting criteria and ad placements. Numbers are provided to give you an idea of performance for your budget, but are only estimates and don't guarantee results.

[Were these estimates helpful?](#)

# Esempi di Pubblicità

Three WiFi Call 14:24

 **Because we are family**  
Sponsored · 

Your MAMA needs this Mug ❤️❤️❤️  
Love this design? Gift it for Your ❤️ MAMA ❤️  
Order here: <https://goo.gl/mU3ogU>



 402    178 comments    182 shares    

 Like     Comment     Share



WINDTRE 11:52 52%

 **UK Expat VPN**  
Sponsored · 

UK Expat VPN is the only provider that  
GUARANTEES you access to key UK se... See more

**MISSING UK TV?**





[WWW.UKEXPATVPN.COM](http://WWW.UKEXPATVPN.COM)    [Learn More](#)

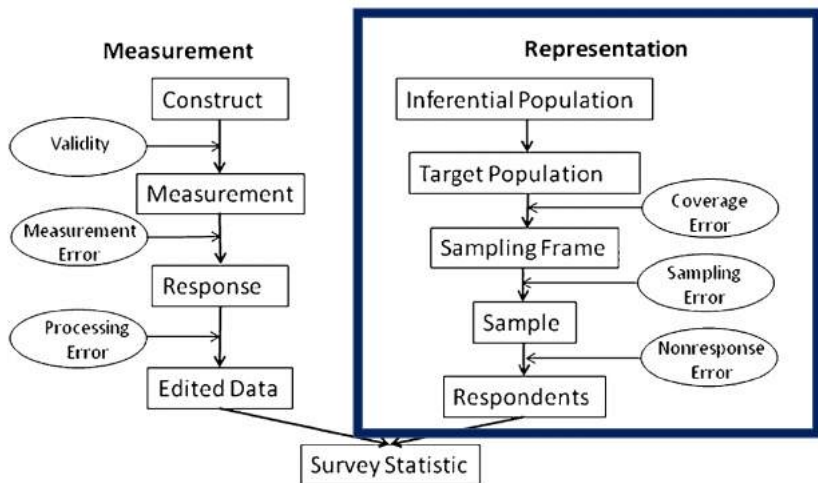
Get access to BBC iPlayer, ITV, More4  
etc abroad with UK Expat VPN. Our...

 842    365 comments    121 shares    



# Focus on Representation

Per rispondere analizzeremo:





# **Inferential Population**

=

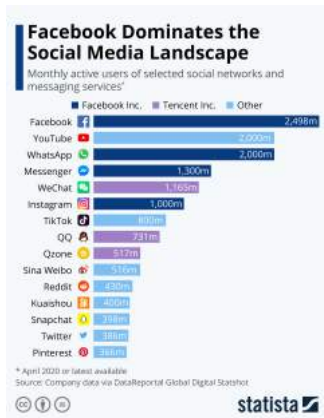
# **Global Population**

Obiettivo di usare Facebook  
per stimare la Popolazione  
Globale

# Target Population

Analisi della distribuzione  
degli utenti Facebook nelle  
variabili chiave

# Uso dei Social Media



WORLD MAP OF SOCIAL NETWORKS  
January 2020

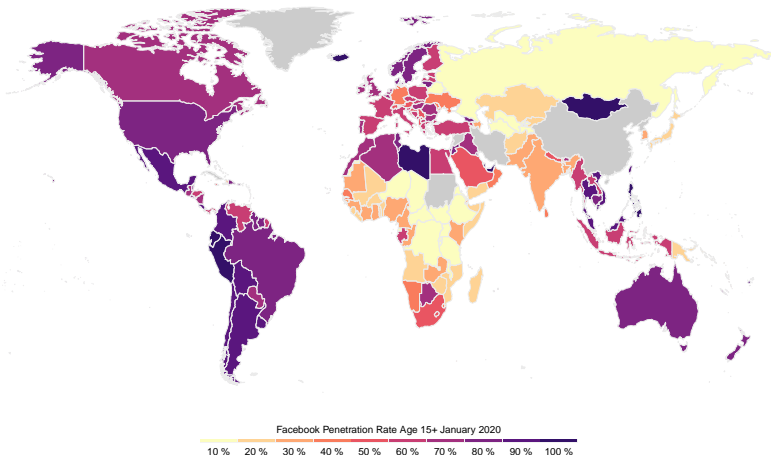


WORLD MAP OF SOCIAL NETWORKS  
December 2019



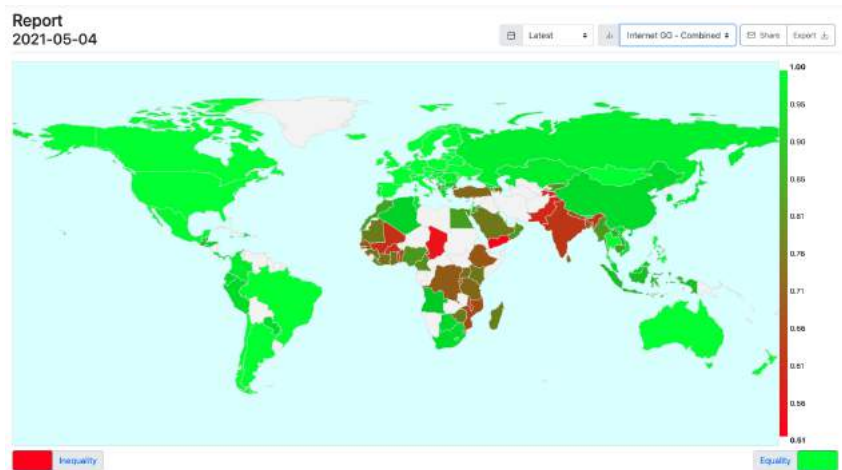
<https://www.statista.com/chart/5194/active-users-of-social-networks-and-messaging-services/>  
<https://vincos.it/world-map-of-social-networks/>

# Uso di Facebook nel Mondo



Facebook penetration rates among users age 15 years and older around the world. The rate is computed as of January 2020 with Facebook and United Nations data (UN 2019, Ševčíková 2020). The Facebook data is courtesy of Palotti et al. 2020.

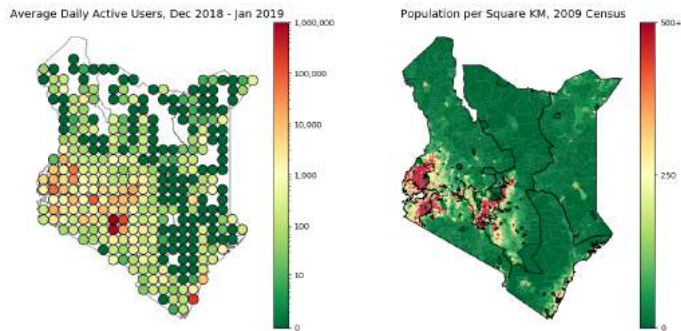
# Gender Index



<https://www.digitalgendergaps.org/>

# Anche nel Sud del Mondo

Esempio di dati di Facebook in Kenya e confronto con i dati del censimento del 2009.



Rosenzweig, R., Bergquist, P., Hoffmann Pham, K., Rampazzo, F., and Mildenerberger, M. (2020). *Survey sampling in the Global South using Facebook advertisements*. SocArXiv.

Copertura di Facebook affetta da:

- Età
- Sesso
- Istruzione
- Paese

# Sampling Frame

Quote degli utenti di Facebook  
distribuiti come inferential  
population.



# Struttura Campagna Pubblicitaria

## Stratum:

- Sex (Uomo e Donna),
- Age (18-24, 25-44, 44-64, 65+),
- Regione di residenza (NUTS1/US Census regions).



Grow, A., Perrotta, D., Del Fava, E., Cimentada, J., Rampazzo, F., Gil, S. and Zagheni, E. (2020). *Addressing Public Health Emergencies via Facebook Surveys: Advantages, Challenges, and Practical Considerations*. Journal of Medical Internet Research, 22(12):e20653.

# Pubblicità delle Indagini

**Yale Survey Research Lab**  
Sponsored · 

Queremos conocer tu opinión. Ayúdanos a saber qué piensas contestando esta encuesta:



YALESURVEY.CA1.QUALTRIC...  
**Queremos conocer tu opinión**

[LEARN MORE](#)

**Kenya SoMe Research Study**  
Written by Leah Rosenzweig · 17 · September 23, 2019 · 

Share your opinion in English, Kiswahili, Gikuyu, Soomaali, or Luo!



**Chukua dakika 20 kwa utafiti  
Upata KES 50 muda wa maongezi kwa simu yako!**



MIT.C01.QUALTRICS.COM  
**Fun Survey!**  
20 minute survey

[Learn More](#)

# Immagini

Images Used in the FB ads campaign.



1 – Male athlete  
©Adobe Stock/gnici



2 – Group of athletes  
©Adobe Stock/nd3000



3 – Woman blowing nose  
©iStockphoto/Goodboy  
Picture Company



4 – Couple blowing noses  
©iStockphoto/Goodboy  
Picture Company

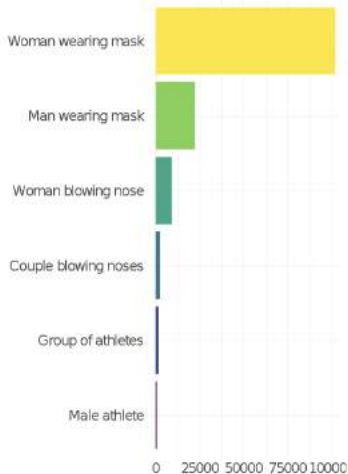


5 – Woman wearing mask  
©Adobe Stock/shintartanya



6 – Man wearing mask  
©iStockphoto/Mixam  
Pydenko

Number of respondents in each image.



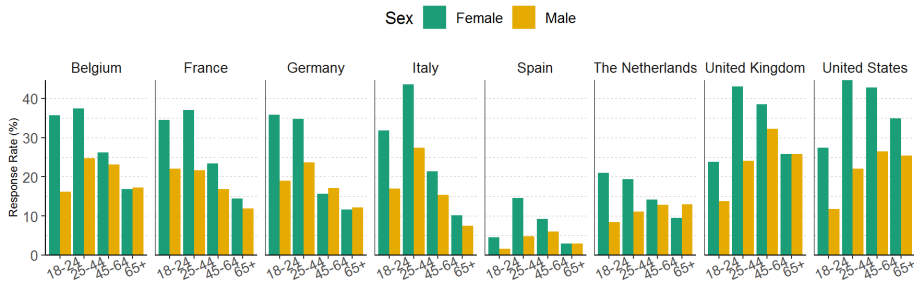
Errore campionario affetto da:

- Auto-Selezione dei rispondenti (self-selection)
- Algoritmo di Facebook
- Trust
- Incentive

# Sample

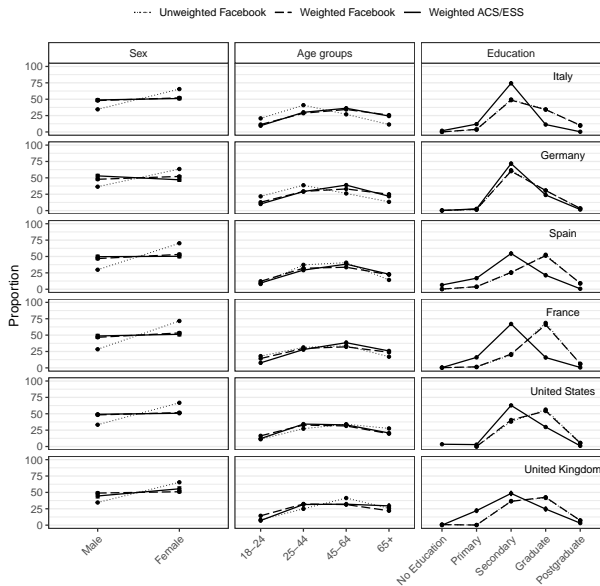
# Tasso di Risposta per Sesso

- **Tasso di Risposta** (% Facebook users who completed the questionnaire after clicking on the ad) generally reasonable with some variability
- **Osservazione 1:** Spagna
- **Osservazione 2:** Donne > Uomini

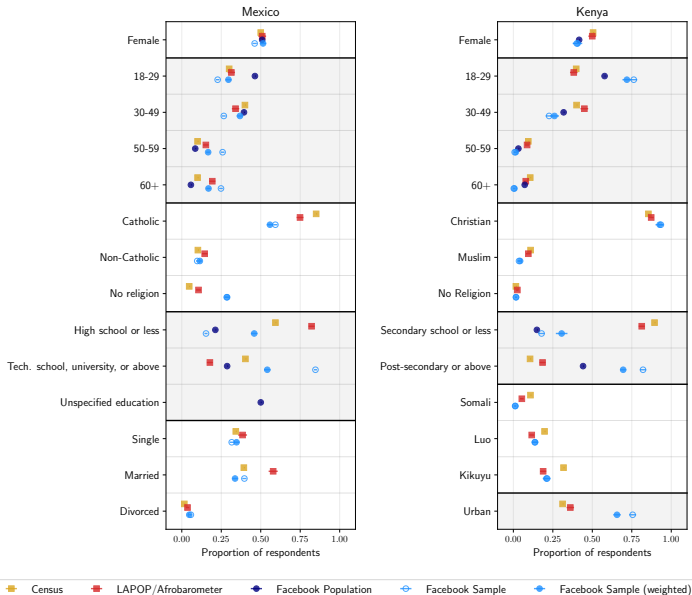


Perrotta, D., Grow, A., Rampazzo, F. et al. *Behaviours and attitudes in response to the COVID-19 pandemic: insights from a cross-national Facebook survey*. EPJ Data Sci. 10, 17 (2021).  
<https://doi.org/10.1140/epjds/s13688-021-00270-1>

# Confronto fra Facebook e European Social Survey (ESS)



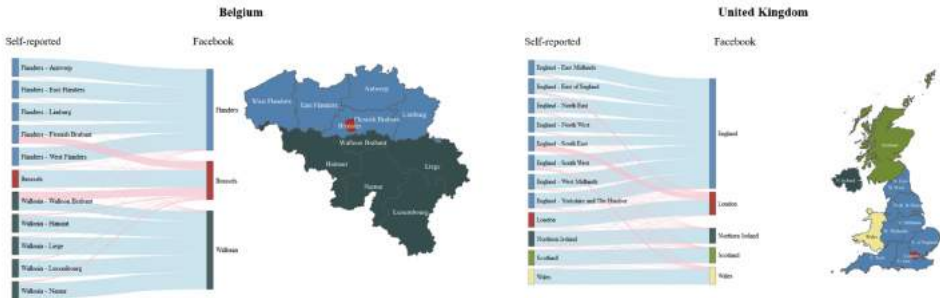
# Mexico e Kenya





Su **137,224** questionari circa il **90%**  
è stato identificato **correttamente**  
per **età, sesso, e luogo di**  
**residenza** (confronto fra pubblicità  
target e dati questionario).

## Problemi principali con luogo di residenza in **Regno Unito** e **Belgio**.



Errore		
Copertura	Campionario	Specifica
Età Sesso Istruzione Paese	Auto-Selezione Algoritmo Facebook Fiducia Incentivo	Non Risposta per Strati Area di Residenza

Domande?

Ci servono le **indagini**,  
le **interviste** nell'era in cui  
viviamo?

Ci servono le **indagini**,  
le **interviste** nell'era in cui  
viviamo?

Sì!

- Campione quasi globale
- Targeting per caratteristiche demografiche
- Veloce e non troppo costoso

## Trasparenza

**Definizioni**

**Algoritmi**

**Etica**



## Trasparenza

**Definizioni**

**Algoritmi**

**Etica**

Non abbiamo molto controllo su campione.

**Integrare le  
digital traces in  
demografia e  
statistica sociale**

**Imparare dalle  
altre discipline  
tecniche per  
analizzarle**

**Creare  
collaborazioni  
con compagnie  
private**

# Alcuni Articoli su Facebook e Survey

- Zhang, B., Mildenerberger, M., Howe, P. D., Marlon, J., Rosenthal, S. A., Leiserowitz, A. (2020). Quota sampling using Facebook advertisements. *Political Science Research and Methods*, 8(3), 558-564.
- Perrotta, D., Grow, A., Rampazzo, F. et al. *Behaviours and attitudes in response to the COVID-19 pandemic: insights from a cross-national Facebook survey*. *EPJ Data Sci.* 10, 17 (2021). <https://doi.org/10.1140/epjds/s13688-021-00270-1>
- Grow, A., Perrotta, D., Del Fava, E., Cimentada, J., Rampazzo, F., Gil, S. and Zagheni, E. (2020). *Addressing Public Health Emergencies via Facebook Surveys: Advantages, Challenges, and Practical Considerations*. *Journal of Medical Internet Research*, 22(12):e20653.
- Rosenzweig, R., Bergquist, P., Hoffmann Pham, K., Rampazzo, F., and Mildenerberger, M. (2020). *Survey sampling in the Global South using Facebook advertisements*. *SocArXiv*.
- Grow, A., Perrotta, D., Del Fava, E., Cimentada, J., Rampazzo, F., Gil-Clavel, S., ... Weber, I. (2021). *How reliable is Facebook's advertising data for use in social science research? Insights from a cross-national online survey (No. WP-2021-006)*. Max Planck Institute for Demographic Research, Rostock, Germany.

# Grazie!

Francesco Rampazzo

🐦 @chiccorampazzo

✉ francesco.rampazzo@sbs.ox.ac.uk

Funded by the

LEVERHULME  
TRUST