

Verso l'integrazione dei dati per la misurazione della violenza

Giancarlo Ragozini
Università di Napoli Federico II

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. Dora has received funding from the European Commission under the European Union's CERF programme (Grant Agreement N. 101095366)

Gruppo di Lavoro

Marcella
Corduas

- Professoressa Ordinaria di Statistica presso il Dipartimento di Scienze Politiche

Giancarlo
Ragozini

- Professore Ordinario di Statistica Sociale presso il Dipartimento di Scienze Politiche

Stefania
Capecchi

- Professoressa Associata di Statistica Sociale presso il Dipartimento di Scienze Politiche

Vincenzo
Fotticchia

- Assegnista di ricerca presso il Dipartimento di Medicina e Chirurgia

Serena
Pulcini

- Assegnista di ricerca presso il Consiglio Nazionale delle Ricerche (CNR), IAC

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. Dora has received funding from the European Commission under the European Union's CERF programme (Grant Agreement N. 101095366)

Obiettivi Generali



Co-funded by the
European Union

Tra gli obiettivi del progetto:

- ✓ Identificazione delle lacune informative sul tema VAC
- ✓ Sviluppo di soluzioni statistiche in grado di combinare, integrare e produrre dati immediatamente utilizzabili

Strategia operativa:

- Pervenire a una stima *accurata* dei fenomeni di interesse
- Definire nuove misure costruendo indicatori opportuni e utilizzando approcci modellistici

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. Dora has received funding from the European Commission under the European Union's CERF programme (Grant Agreement N. 101095368)



UNIVERSITÀ
DEGLI STUDI
DI BRESCIA



OUTLINE



Co-funded by the
European Union

- Perché integrare i dati sulla VAC
 - lo stato delle fonti
- La Data Integration (DI)
 - Principali approcci statistici
 - Problemi e limitazioni
- Proposte di nuove misure e modelli
 - Indicatori
 - Calibrazione

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. Dora has received funding from the European Commission under the European Union's CERF programme (Grant Agreement N. 101095366)



UNIVERSITA'
DEGLI STUDI
DI BRESCIA





Per le diverse aree tematiche abbiamo fonti che:

- si riferiscono a unità amministrativo-territoriali diverse
- si riferiscono a periodi diversi
- si modificano nel tempo
- sono frammentate

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. Dora has received funding from the European Commission under the European Union's CERF programme (Grant Agreement N. 101095368)

Stato delle fonti



Co-funded by the
European Union

Violenza Psicologica

FONTE DEI DATI → ISTAT e Health Behaviour in School-aged Children (HBSC) ITALIA

CONTENUTO → diverse indagini: «Autori e vittime di reati segnalati dalle forze dell'ordine all'autorità giudiziaria» da ISTAT (2013-2017), «Numero di reati» da ISTAT (2013-2022) e «Bullismo» da HBSC ITALIA (2014, 2018, 2022)

PERIODO → 2013-2022

VARIABILI INIZIALI COINVOLTE → 34 variabili: 32 variabili disponibili a livello regionale 2 variabili disponibili a livello nazionale

VARIABILI FINALI COINVOLTE → Rimosse le 2 variabili disponibili a livello nazionale, restano 32 variabili

LIVELLO DI AGGREGAZIONE → Dati disponibili a livello regionale e per macro-regioni ISTAT (5)

ALTRE FONTI UTILIZZATE → «Bullismo: Comportamenti offensivi e violenti tra i giovani» da ISTAT

Year	Starting situation				Situation Pre-Imputation (deleted variables)			
	Var	Tot	Pre	Na	Var	Tot	Pre	Na
2013	20	600	494	106	32	960	494	466
2014	26	780	628	152	32	960	628	332
2015	20	600	496	104	32	960	496	464
2016	20	600	496	104	32	960	496	464
2017	20	600	501	99	32	960	501	459
2018	20	600	488	112	32	960	488	472
2019	8	240	224	16	32	960	224	736
2020	8	240	224	16	32	960	224	736
2021	8	240	224	16	32	960	224	736
2022	10	300	268	32	32	960	200	760
TOT	160	4.800	4.043	757	320	9.600	3.975	5.625

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. Dora has received funding from the European Commission under the European Union's CERF programme (Grant Agreement N. 101095366)



Stato delle fonti



Co-funded by the
European Union

Violenza Sessuale

Fonte Dati → ISTAT

Contenuto → Fusione di due diverse indagini:

“Casi e reati per tipologia di comune” (ISTAT, 2013-2017)

“Imputati per tipologia di reato” (ISTAT, 2013-2022)

Periodo → 2013-2022

Variabili Iniziali Coinvolte → 81

variabili:

6 variabili disponibili a livello regional

75 variabili disponibili a livello nazionale

Variabili Finali Coinvolte → Rimosse 6

variabili disponibili a livello nazionale, rimanendo 75 variabili

Livello di Aggregazione → Dati disponibili a livello regionale e per macroregioni ISTAT (5)

Year	Starting situation				Situation Pre-Imputation (deleted variables)			
	Var	Tot	Pre	Na	Var	Tot	Pre	Na
2013	80	2.400	1.859	541	75	2.250	1.824	426
2014	81	2.430	1.947	483	75	2.250	1.905	345
2015	81	2.430	1.967	463	75	2.250	1.925	325
2016	81	2.430	1.996	434	75	2.250	1.954	296
2017	81	2.430	1.952	478	75	2.250	1.910	340
2018	20	600	560	40	75	2.250	560	1.690
2019	20	600	560	40	75	2.250	560	1.690
2020	20	600	560	40	75	2.250	560	1.690
2021	20	600	560	40	75	2.250	560	1.690
2022	20	600	560	40	75	2.250	560	1.690
TOT	504	15.120	12.521	2.599	750	22.500	12.318	10.182

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. Dora has received funding from the European Commission under the European Union's CERF programme (Grant Agreement N. 101095366)



UNIVERSITÀ
DEGLI STUDI
DI BRESCIA



Stato delle fonti



Co-funded by the
European Union

Inoltre abbiamo fonti «indirette»
campionarie e non campionarie:

- Indagine sulla sicurezza dei cittadini
- Indagine sulla violenza sulle donne
- Chiamate al 1522
- Dati dei CAV

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. Dora has received funding from the European Commission under the European Union's CERF programme (Grant Agreement N. 101095366)



La Data Integration

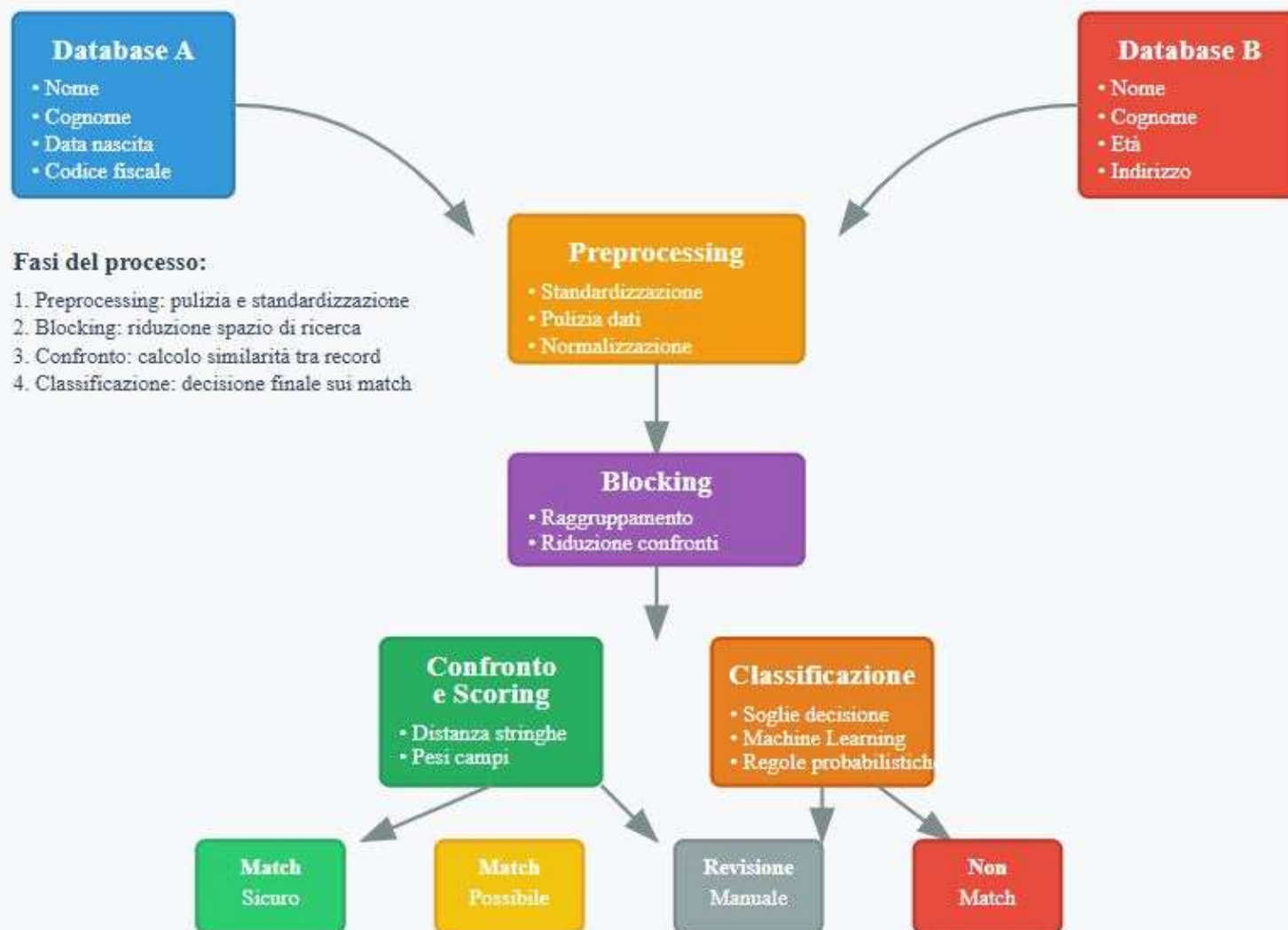
- Record linkage
- Privacy preserving record linkage
- Statistical Matching
- Probabilistic Statistical Matching
- Data fusion
- File Grafting

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. Dora has received funding from the European Commission under the European Union's CERF programme (Grant Agreement N. 101095366)

La Data Integration

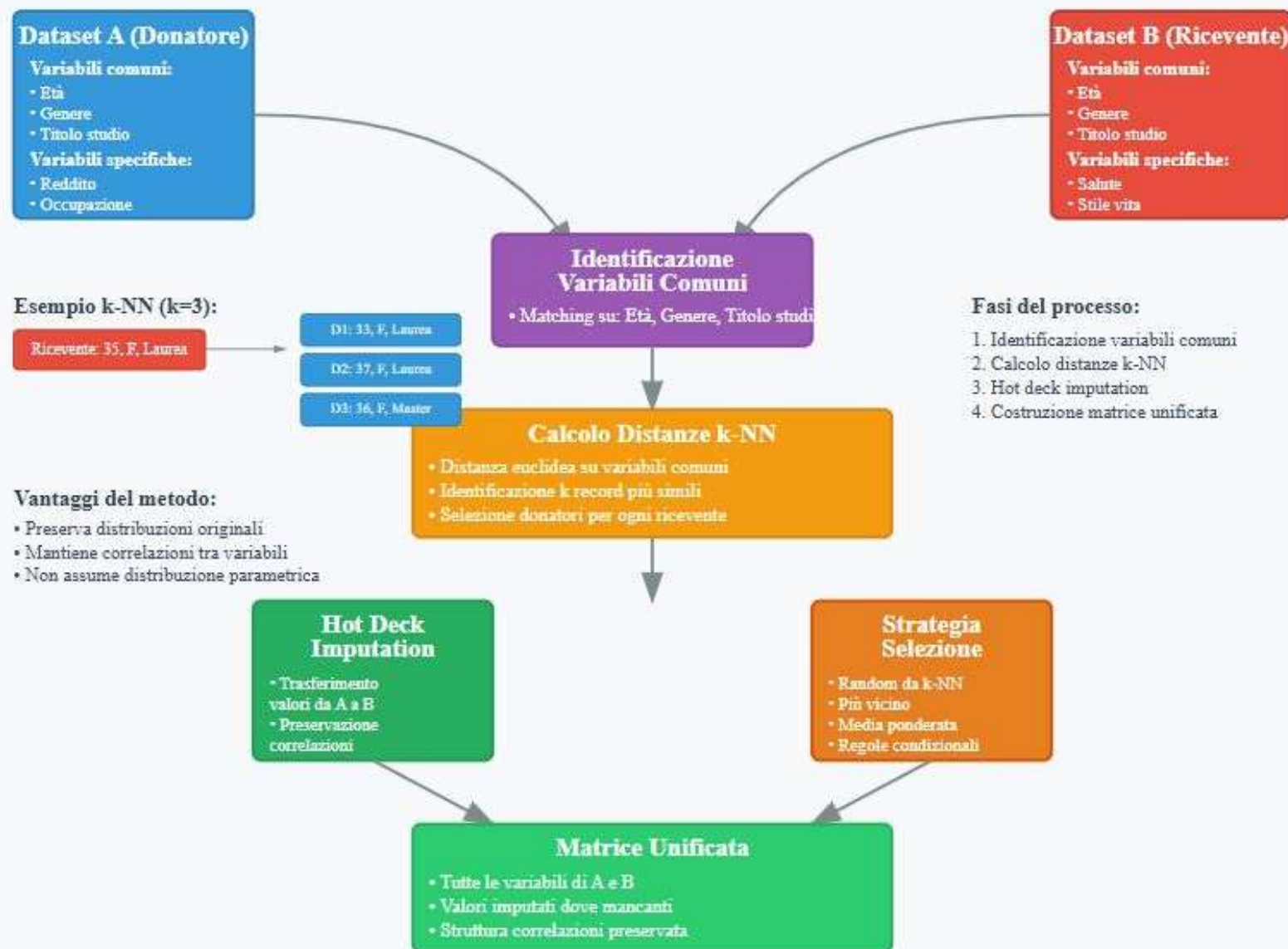
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union.

Record Linkage Statistico



La Data Integration

File Grafting con Hot Deck Imputation (k-NN)



La Data Integration



Co-funded by the
European Union



- Problemi di integrazione:

Le unità statistiche
non sono le stesse



Record linkage e
statistical matching non è
possibile



Uso di metodi di Data
Fusione e File
Drafting



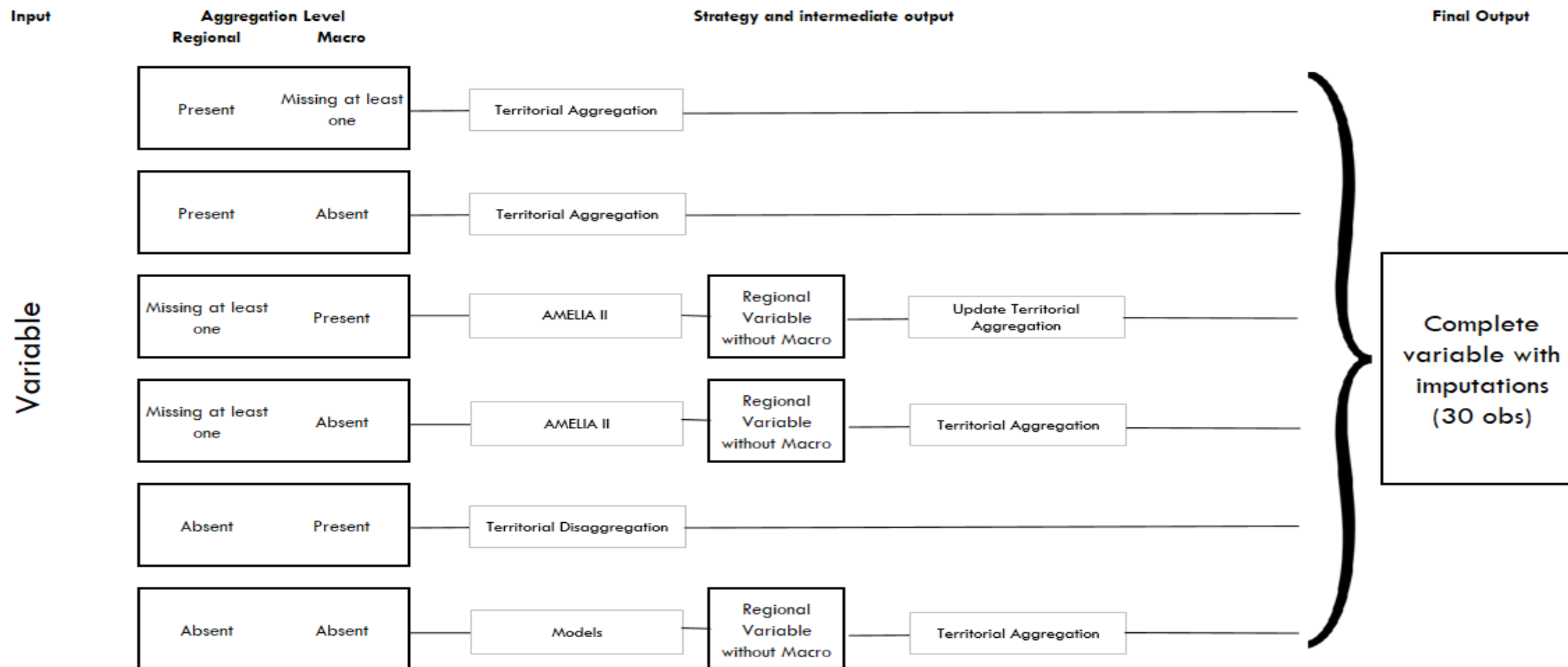
Troppi dati mancanti sulle
variabili comuni

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. Dora has received funding from the European Commission under the European Union's CERF programme (Grant Agreement N. 101095366)

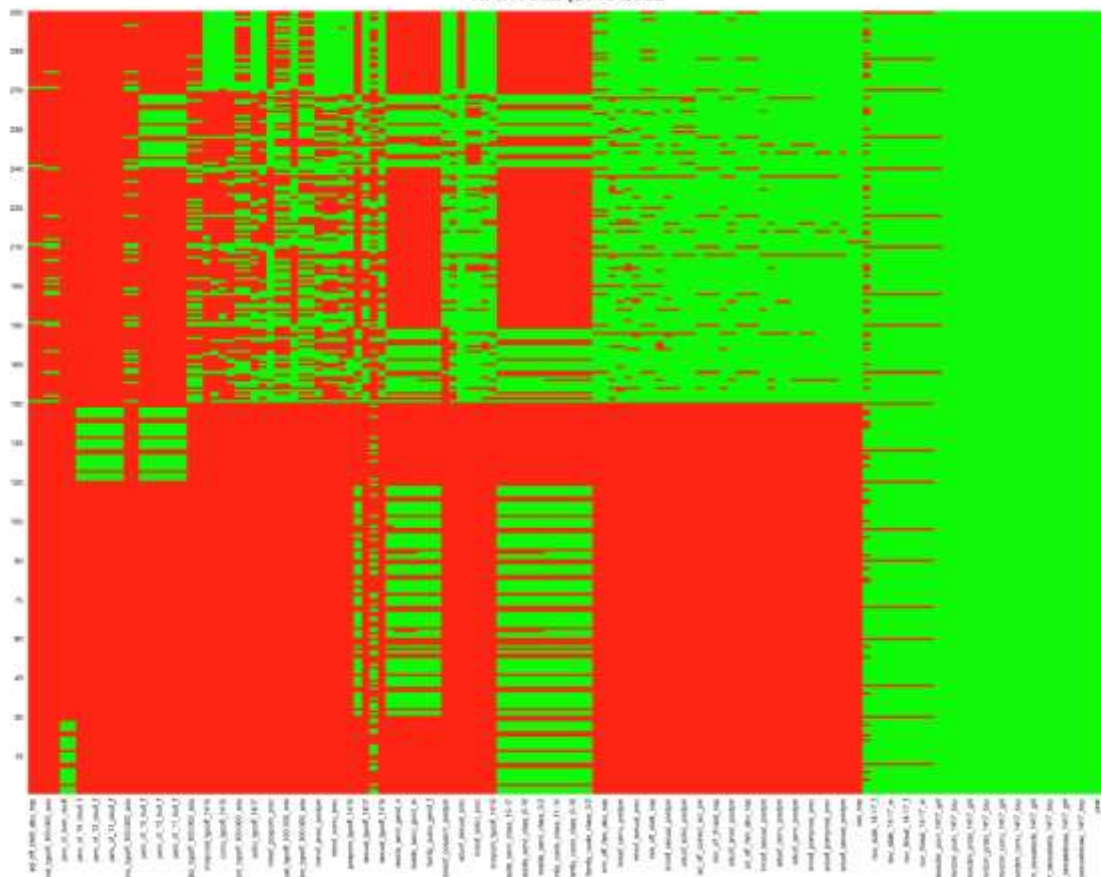
La nostra strategia



Co-funded by the
European Union



NAs Total (2013-2022)



Missing (53%)
Observed (47%)



134 VARIABILI



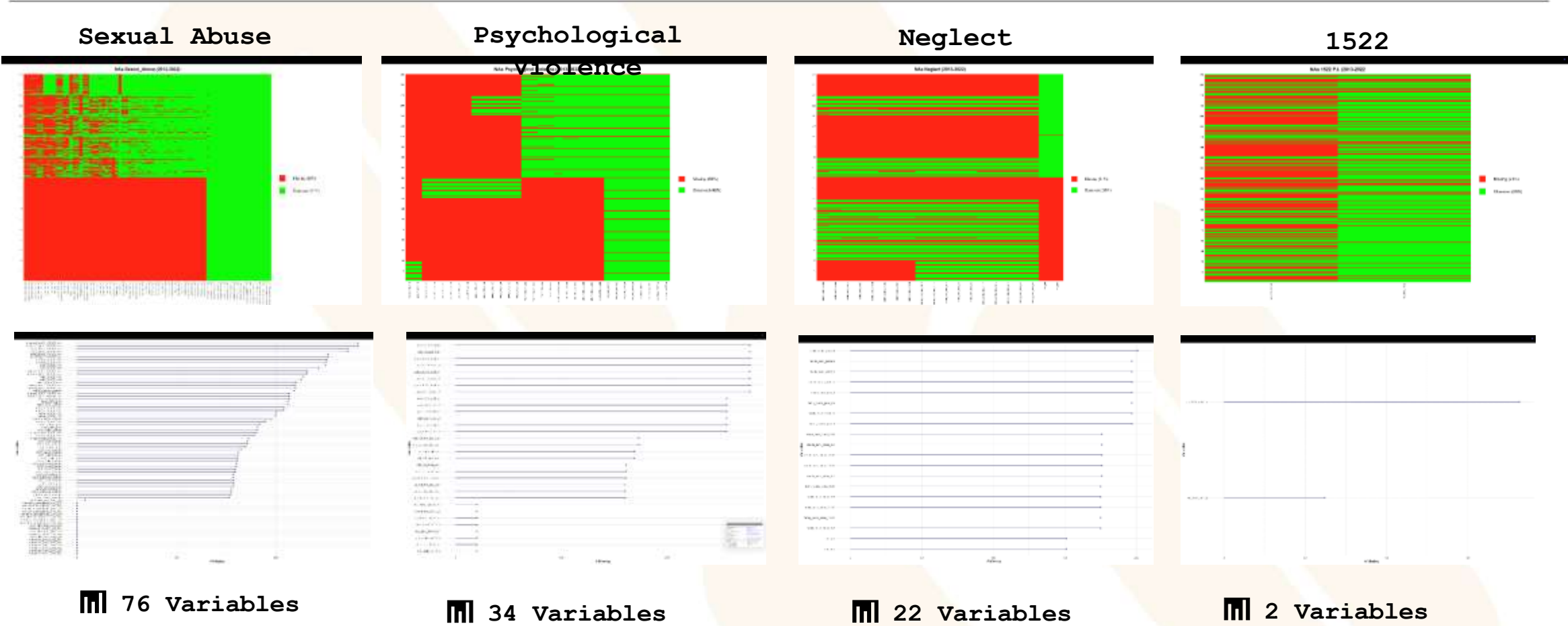
300 OSSERVAZIONI



40200 dati

53%
MISSING
VALUES

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. Dora has received funding from the European Commission under the European Union's CERF programme (Grant Agreement N. 101095366).



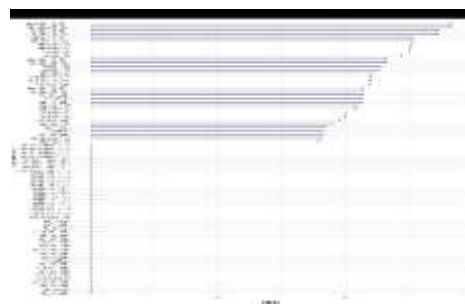
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. Dora has received funding from the European Commission under the European Union's CERF programme (Grant Agreement N. 101095366)

Alla fine del Processo

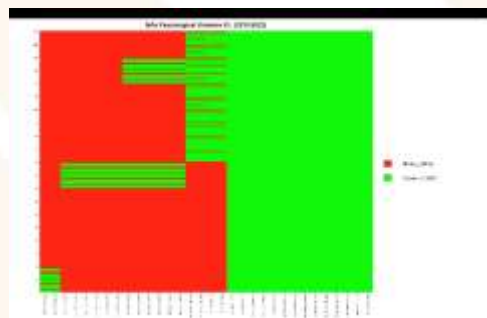


Co-funded by the
European Union

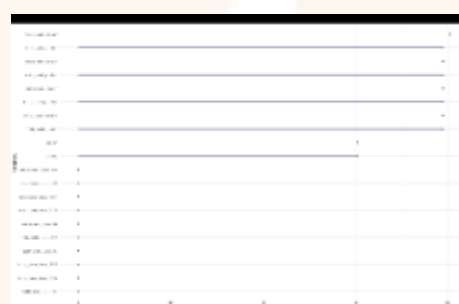
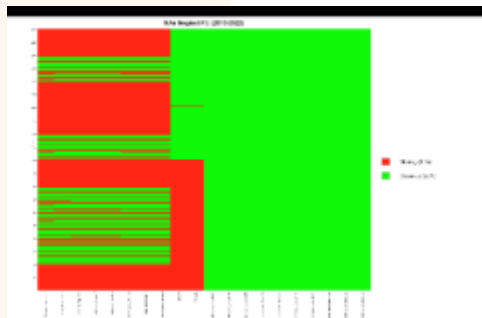
Sexual Abuse



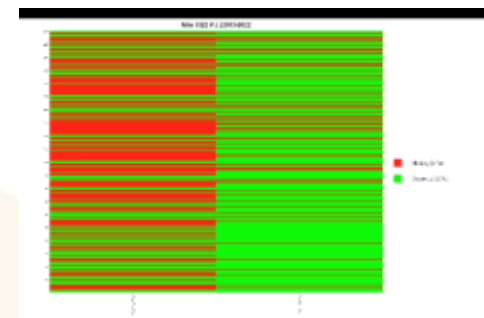
Psychological Violence



Neglect



1522



Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. Dora has received funding from the European Commission under the European Union's CERF programme (Grant Agreement N. 101095366).

Thematic Area	Observations	Variables_PRE	Complete Variables_PRE	%NA_PRE	Variables_POST	Complete Variables_POST	%NA_POST
Sexual	300	76	19	49%	76	42	32%
Psico	300	34	0	60%	32	14	46%
Neglect	300	22	0	61%	20	10	31%
1522	300	2	0	48%	2	0	41%



66 VARIABILI



300 OSSERVAZIONI



19800 DATI

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. Dora has received funding from the European Commission under the European Union's CERF programme (Grant Agreement N. 101095366)

Proposte di nuove misure

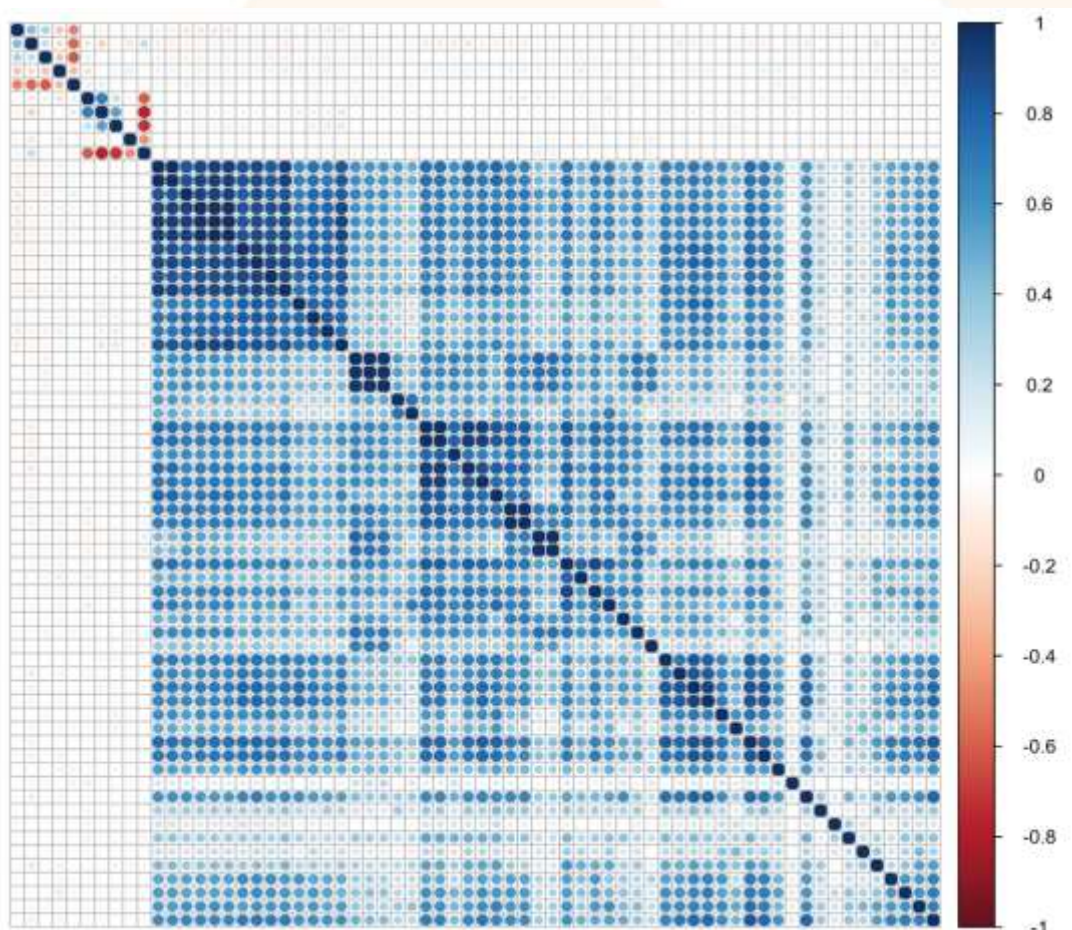
- Indicatori su base regionale (sul database ricostruito)
- Uso di fonti indirette per valutare il sommerso
- Uso di fonti indirette per misurare il sommerso

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. Dora has received funding from the European Commission under the European Union's CERF programme (Grant Agreement N. 101095366)

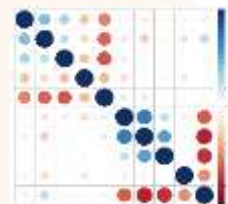
INDICATORI SU BASE REGIONALE



Co-funded by the
European Union



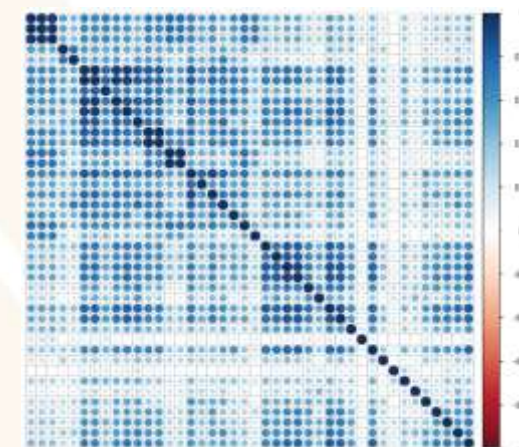
Neglect



Psychological
Violence



Sexual Abuse



Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. Dora has received funding from the European Commission under the European Union's CERF programme (Grant Agreement N. 101095368)

Combinazione delle aree
tematiche tramite modello
fattoriale



Indicatore composito che
unisce tre diversi aspetti

Sexual Abuse

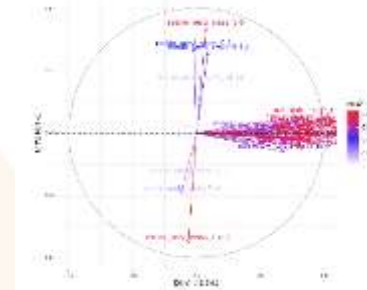
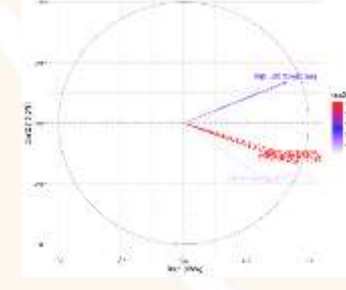
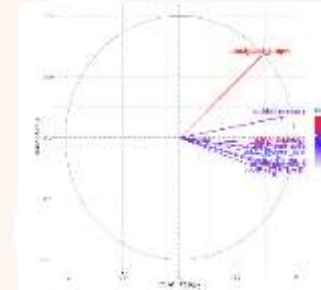
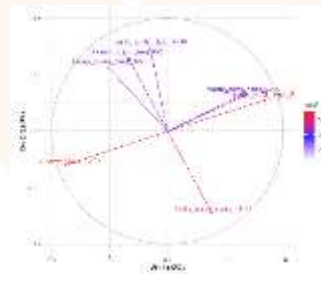
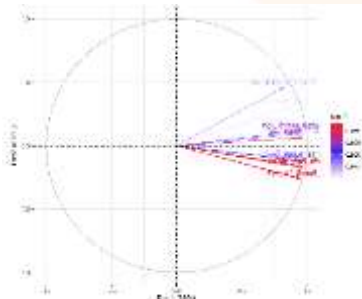
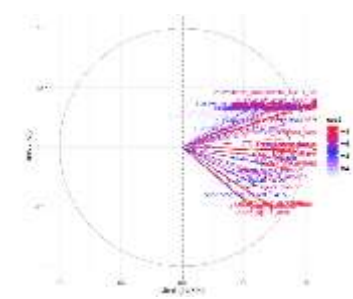
**Psychological
Violence**

Neglect

Adult Perpetrator

**Minor
perpetrator**

Child victim



Modelli fattoriali testati

Psycho =~

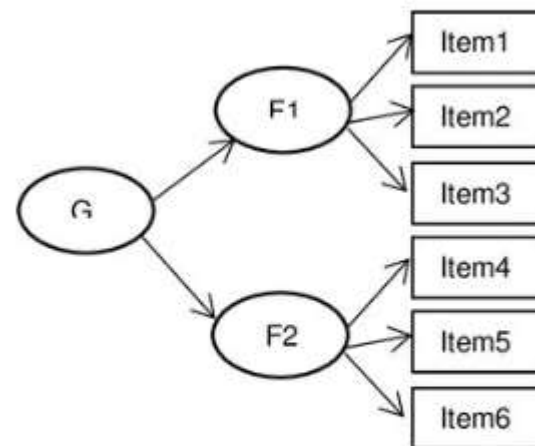
min_ff_thrt_pm	1.000				0.396	0.397
ad_ff_chld_b_p	1.191	0.088	13.586	0.000	0.472	0.473
ad_off_fm_b_pm	1.130	0.071	15.835	0.000	0.448	0.449
nv_thrt_ntl13_	1.235	0.140	8.853	0.000	0.489	0.491
nv_thrt_ntl13_	1.386	0.151	9.197	0.000	0.549	0.550
nv_stlk_ntl13_	0.972	0.128	7.603	0.000	0.385	0.391
nv_stlk_ntl13_	1.305	0.152	8.572	0.000	0.517	0.518

Sexual =~

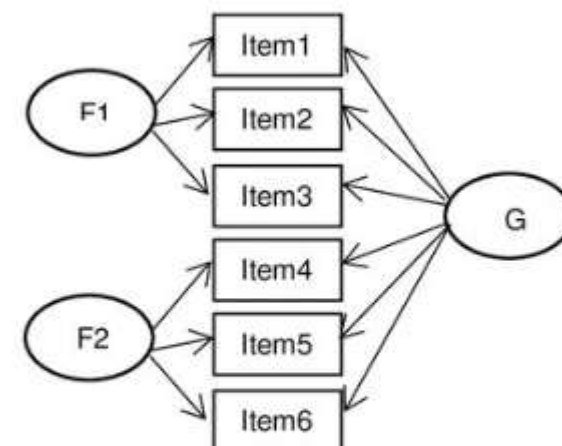
mnf_prnprst_ps	1.000				0.445	0.446
minf_sxl_pstpm	0.928	0.184	5.040	0.000	0.413	0.414
adf_prnprst_ps	0.851	0.132	6.457	0.000	0.378	0.379
aduf_sxl_pstpm	0.506	0.110	4.619	0.000	0.225	0.226
aduf_crr_pstpm	1.288	0.190	6.787	0.000	0.572	0.574
aduf_slc_pstpm	1.207	0.141	8.560	0.000	0.537	0.538
nmvctm_sxl_13_	-0.372	0.121	-3.062	0.002	-0.165	-0.166

Neglect =~

fmly_cst_cl_02	1.000				0.683	0.684
fmly_cst_cl_35	0.915	0.221	4.147	0.000	0.625	0.627
fmly_cst_c_610	0.617	0.155	3.979	0.000	0.421	0.423
rsd_srv_cls_02	-0.089	0.127	-0.702	0.482	-0.061	-0.061
rsd_srv_cls_35	-0.306	0.133	-2.305	0.021	-0.209	-0.210

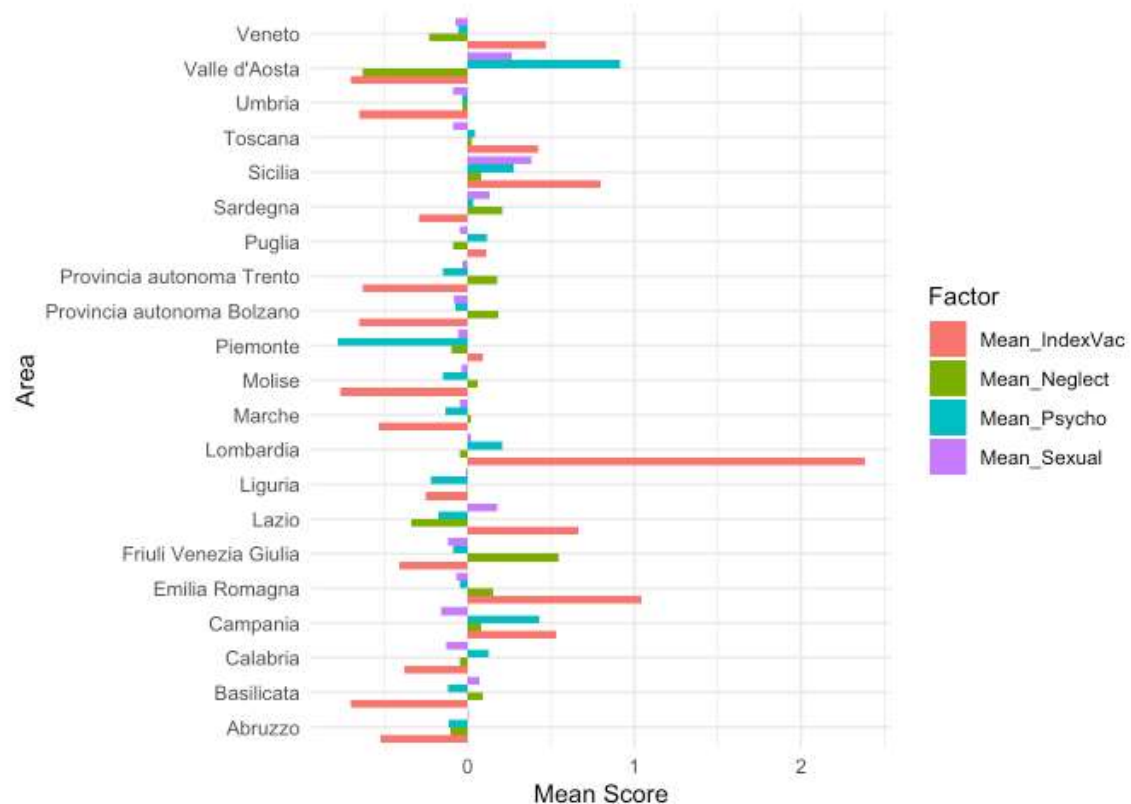


2nd order CFA

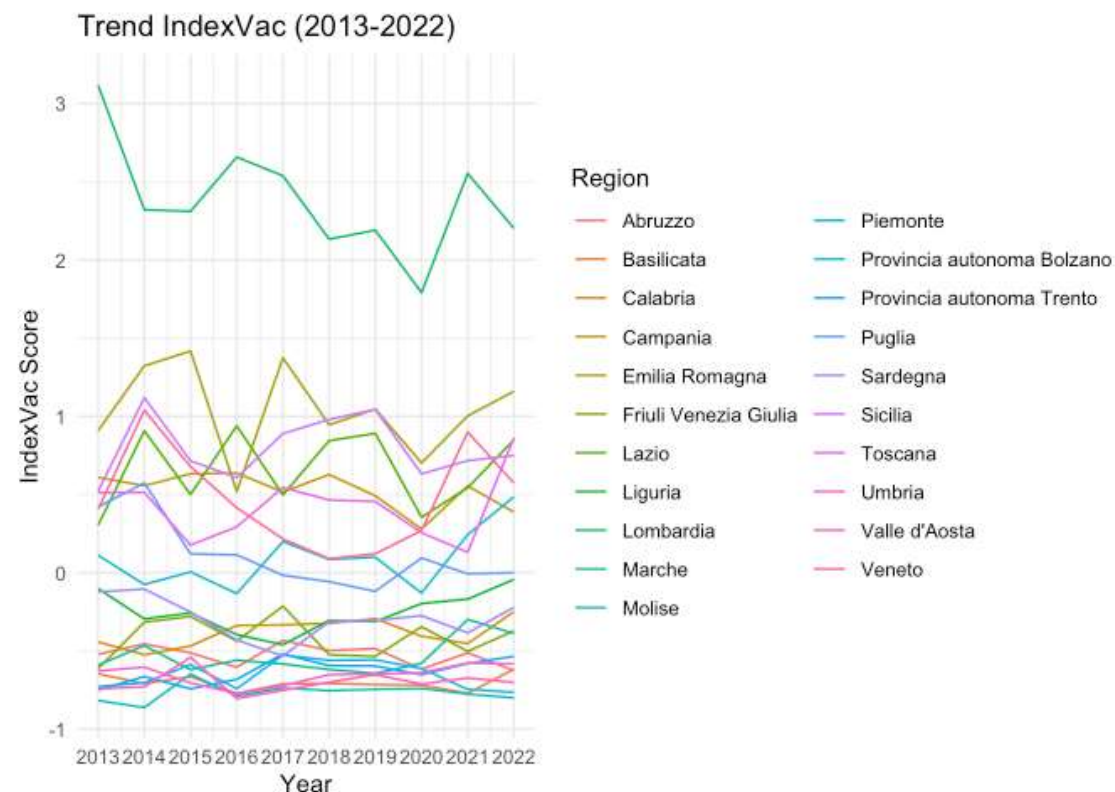


Bi-factor CFA

3 Grafico su medie degli anni



3.2 Grafico su tutti i fattori anno per anno



Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. Dora has received funding from the European Commission under the European Union's CERF programme (Grant Agreement N. 101095366).

Uso di fonti indirette per valutare il sommerso

- Integriamo due fonti:
- Chiamate al 1522
- Indagine sulla sicurezza delle donne

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. Dora has received funding from the European Commission under the European Union's CERF programme (Grant Agreement N. 101095366)

Criterio di selezione 1522 NUMERO ANTI VIOLENZA E STALKING



Chiamate dirette dei minori



Chiamate da un adulto che riporta
anche la presenza di minori

♀ = 100 👦 boys 👨 men 👧 girls 👩 women

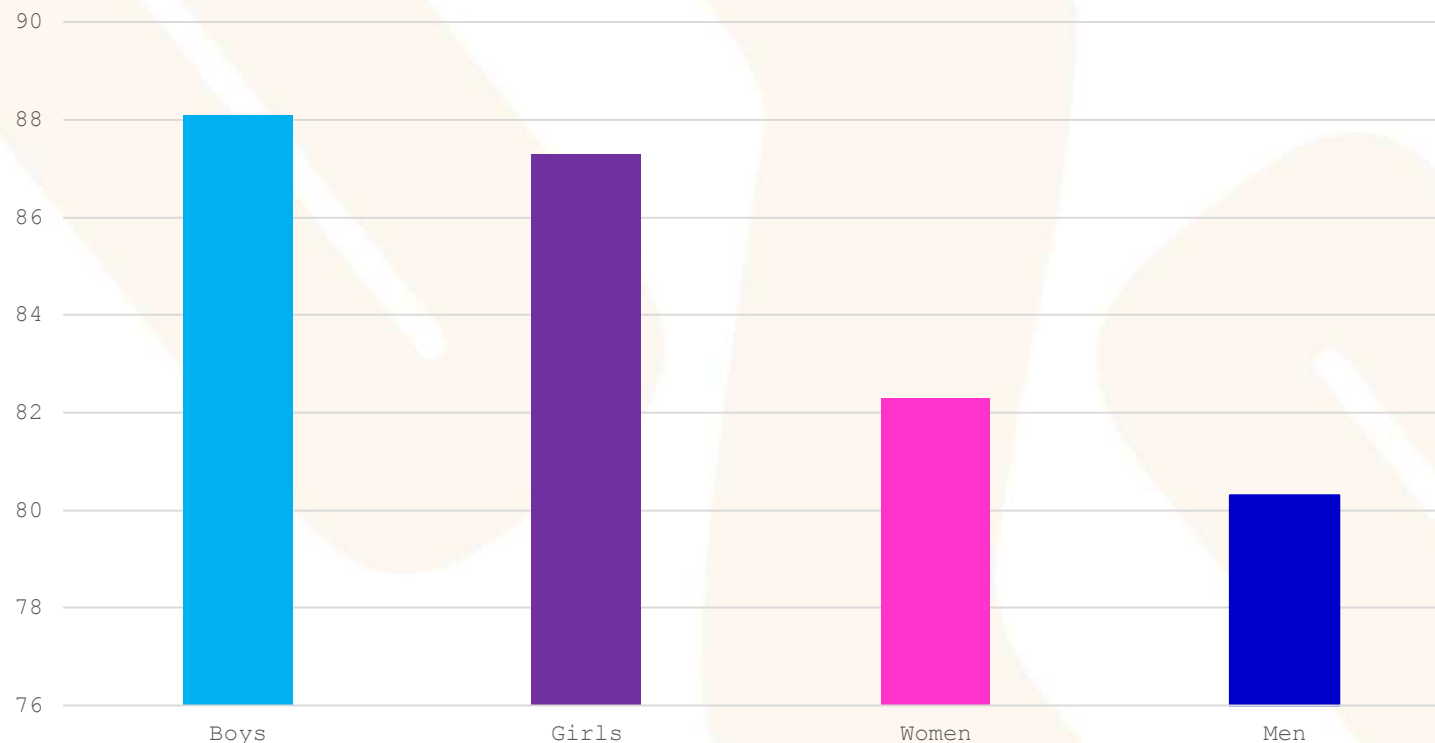
"Poiché le donne costituiscono il 94% delle chiamate, esse rappresentano la principale fonte indiretta per analizzare l'esposizione dei bambini alla violenza."



Sample composition

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. Dora has received funding from the European Commission under the European Union's CERF programme (Grant Agreement N. 101095366).

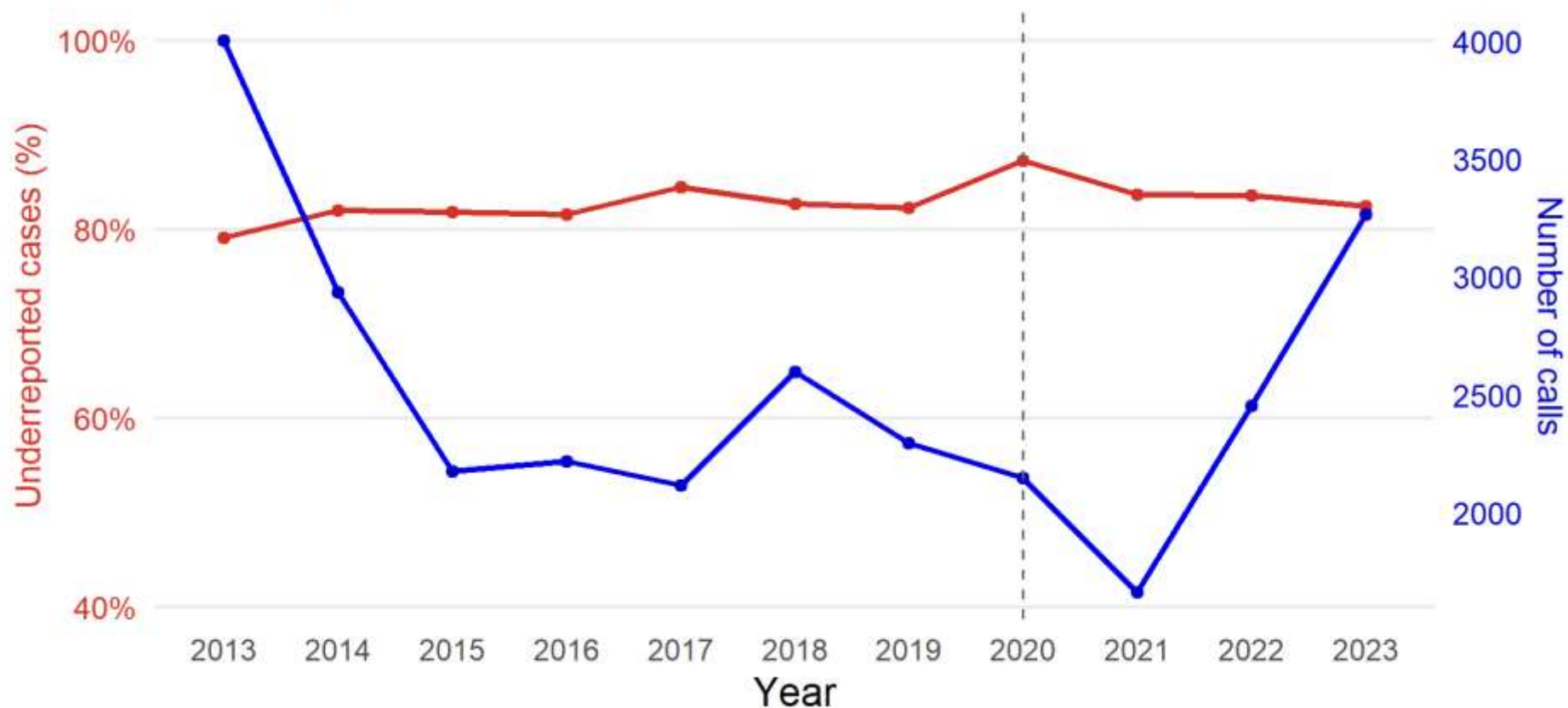
- **La non denuncia** ha una percentuale pari al 82.49%



Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. Dora has received funding from the European Commission under the European Union's CERF programme (Grant Agreement N. 101095366).



Underreporting vs Calls to 1522 (2013–2023)



Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. Dora has received funding from the European Commission under the European Union's CERF programme (Grant Agreement N. 101095366).

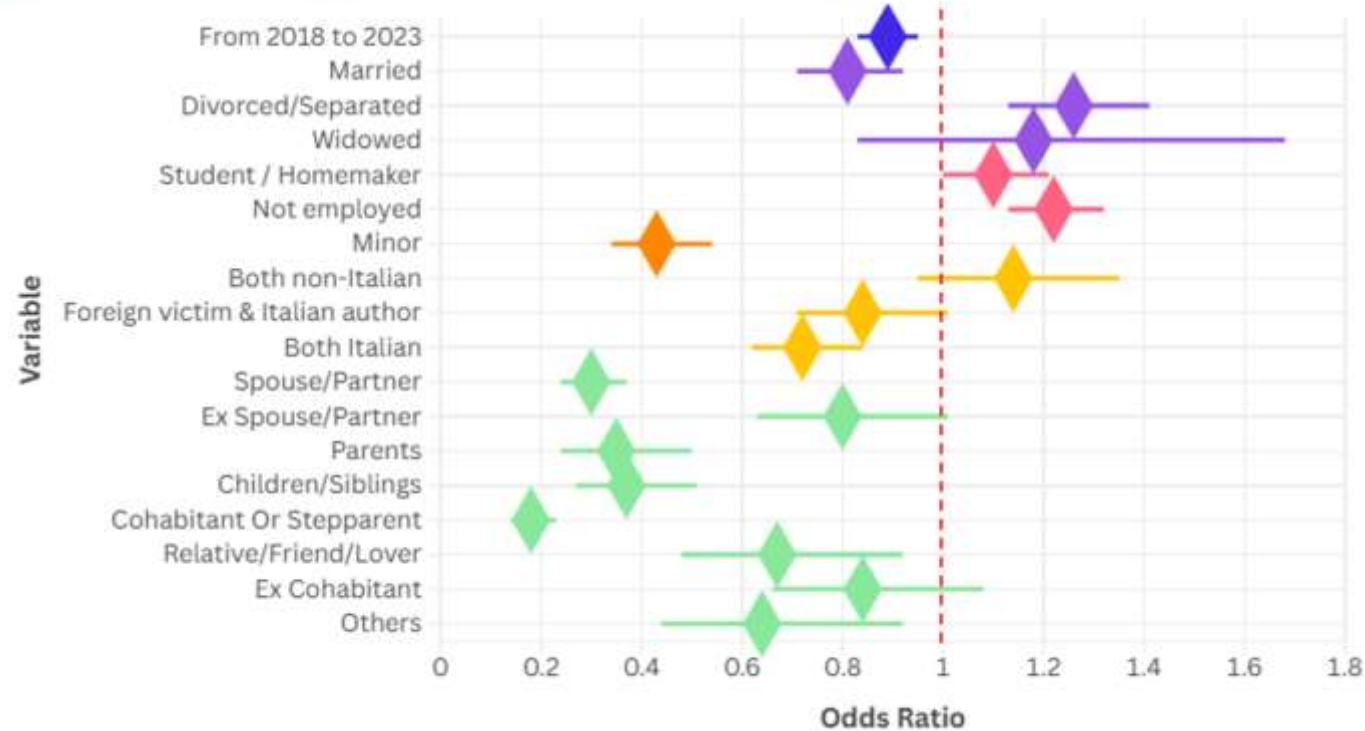
Modello di regressione logistica per la non denuncia

- La categoria di riferimento è stata impostata su 'No' per modellare la probabilità di denuncia.
- Tutte le variabili categoriche incluse:
variabile relativa al periodo, relazione tra autore e vittima, interazione tra nazionalità (italiana/non italiana), occupazione, fascia di età (minorenni vs adulti) e stato civile."

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. This has received funding from the European Union's Horizon programme (Grant Agreement N. 101095366)



Reference ● Period: From 2013 to 2017 ● Marital Status: Single ● Occupation: Employed
 ● Victim Age Group: Adult ● Origin: Italian victim & foreign author ● Perpetrator: Acquaintance/Neighbor



Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. Dora has received funding from the European Commission under the European Union's CERF programme (Grant Agreement N. 101095366).

Uso di fonti indirette per valutare il sommerso

- Integriamo due fonti:
- Chiamate al 1522
- Indagine sulla sicurezza delle donne
- Calibriamo le chiamate del 1522 utilizzando l'indagine sulla sicurezza delle donne del 2013

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. Dora has received funding from the European Commission under the European Union's CERF programme (Grant Agreement N. 101095368)

Calibrazione Indagine 1522 (2013-2015) con Totali Sicurezza delle Donne 2013

Indagine "Sicurezza delle Donne" 2013
(Indagine campionaria)

Totali di calibrazione:

- Donne vittime per età
- Ripartizione geografica
- Stato civile
- Condizione lavorativa

Chiamate al 1522 (2013-2015)
(Indagine totale)

Variabili osservate:

- Età vittima
- Area geografica
- Stato civile
- Lavoro (anche nero)

Esempio Totali di Calibrazione:

Variabile di Calibrazione
Età 16-24: 550.000 donne vittime
Età 25-44: 1.200.000 donne vittime
Nord: 1.100.000 donne vittime
Centro: 650.000 donne vittime
Sud-Isole: 900.000 donne vittime
Coniugate: 1.300.000 donne vittime
Occupate: 1.400.000 donne vittime

Definizione Totali di Calibrazione

- T₁: Donne vittime 16-24 anni
- T₂: Donne vittime 25-44 anni
- T₃: Nord, Centro, Sud, Isole
- T₄: Nubile, Coniugata, Separata...

Calcolo Pesi di Calibrazione

Metodo: Raking o Post-stratificazione

- w_{ij} = peso calibrato per unità i
- $\sum w_{ij} T_{ij} = T_j$ per ogni variabile j
- Minimizzazione distanza da pesi iniziali

Vantaggi della Calibrazione:

- Correzione bias di selezione
- Rappresentatività popolazione
- Stime coerenti con indagine ufficiale
- Riduzione varianza stime

Limiti del metodo:

- Ipotesi di omogeneità nei gruppi
- Disponibilità variabili comuni
- Qualità totali di riferimento

Timeline del processo:

2013 Indagine Sicurezza | 2013-2015 Chiamate 1522 | Post-2015 Stime calibrate

Applicazione Pesi alle Chiamate 1522

- Ogni chiamata riceve peso w_{ij}
- Rappresenta popolazione di donne vittime

Stima Variabili Target sui Bambini

- Bambini che assistono
- Bambini che subiscono
- Bambini che assistono E subiscono violenza

Stime Calibrate Finali

$\hat{Y} = \sum w_{ij} Y_{ij}$ per ogni variabile target

- Numero stimato bambini coinvolti nella violenza
- Rappresentativo della popolazione italiana

References



Co-funded by the
European Union

- Aluja-Banet, T., Daunis-i-Estadella, J., & Pellicer, D. (2007). GRAFT, a complete system for data fusion. *Computational statistics & Data analysis*, 52(2), 635-649.
- Aria, M., D'Ambrosio, A., & Siciliano, R. (2007). Robust Incremental Trees for Missing Data Imputation and data Fusion. In *Classification and Data Analysis 2007* (Vol. 1, pp. 287-290). eum edizioni università di Macerata.
- Dong, X.L. and D. Srivastava (2015). Big data integration. *Synth. Lect. Data Manag.*, vol. 7, no. 1, pp. 1-198.
- Donatiello, G., D'Orazio, M., Frattarola, D., Rizzi, A., Scanu, M., & Spaziani, M. (2016b). The role of the conditional independence assumption in statistically matching income and consumption. *Statistical Journal of the IAOS*, 32(4), 667-675.
- D'Orazio, M., Di Zio, M., & Scanu, M. (2006). *Statistical matching: Theory and practice*. John Wiley & Sons.
- Dunn, H. L. (1946). Record linkage. *American Journal of Public Health and the Nations Health*, 36(12), 1412-1416
- Gao, D., Sriukenthiran, S., Habib, K. N., & Miller, E. J. DATA FUSION: TECHNIQUES AND APPLICATIONS. Working Paper
- Rubin, D. B. 1972. A non-iterative algorithm for least squares estimation of missing values in any analysis of variance design. *Journal of the Royal Statistical Society, Series C* 21: 136-141. <https://doi.org/10.2307/2346485>.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. Dora has received funding from the European Commission under the European Union's CERF programme (Grant Agreement N. 101095366)

