

A. SELCE, F. NGUYEN

Dipartimento Fusione Nucleare e Tecnologie
per la Sicurezza Nucleare
Divisione Fisica della Fusione
Laboratorio Teoria, Simulazione e Modellistica
Centro Ricerche Frascati, Roma

INTERPRETAZIONE ANALITICA E QUANTITATIVA DELLA DIFFERENZA DI DECESSI TRA LOMBARDIA E HUBEI A CAUSA DEL VIRUS COVID-19

RT/2020/8/ENEA



AGENZIA NAZIONALE PER LE NUOVE TECNOLOGIE,
L'ENERGIA E LO SVILUPPO ECONOMICO SOSTENIBILE

A. SELCE, F. NGUYEN

Dipartimento Fusione Nucleare e Tecnologie
per la Sicurezza Nucleare
Divisione Fisica della Fusione
Laboratorio Teoria, Simulazione e Modellistica
Centro Ricerche Frascati, Roma

INTERPRETAZIONE ANALITICA E QUANTITATIVA DELLA DIFFERENZA DI DECESSI TRA LOMBARDIA E HUBEI A CAUSA DEL VIRUS COVID-19

RT/2020/8/ENEA



AGENZIA NAZIONALE PER LE NUOVE TECNOLOGIE,
L'ENERGIA E LO SVILUPPO ECONOMICO SOSTENIBILE

I rapporti tecnici sono scaricabili in formato pdf dal sito web ENEA alla pagina www.enea.it

I contenuti tecnico-scientifici dei rapporti tecnici dell'ENEA rispecchiano l'opinione degli autori e non necessariamente quella dell'Agenzia

The technical and scientific contents of these reports express the opinion of the authors but not necessarily the opinion of ENEA.

INTERPRETAZIONE ANALITICA E QUANTITATIVA DELLA DIFFERENZA DI DECESSI TRA LOMBARDIA E HUBEI A CAUSA DEL VIRUS COVID-19

A. Selce, F. Nguyen

Riassunto

Obiettivo di questo studio è confrontare i dati in Lombardia con quelli della provincia di Hubei relativi ai decessi dovuti a Covid-19, tenendo conto della differenza di età tra i campioni analizzati: è così possibile riscalarne i dati cinesi per essere confrontati con quelli italiani.

Parole chiave: Covid-19, Analisi Statistica.

Abstract

Main scope of this study is a comparison of data between the Lombardy region and the Hubei province on the deaths due to the Covid-19 pandemic, accounting for age differences between the analysed samples: thus it is possible to rescale the data from China in order to be properly compared with those from Italy..

Keywords: Covid-19, Statistical Analysis.

INDICE

1. Introduzione	7
2. Scelta del campione	7
3. Analisi dei dati	8
4. Scaling dei dati e relativo confronto	9
a. Scaling conservativo (C-rescale)	10
b. Scaling mediato (M-rescale)	11
c. Scaling non-conservativo (NC-rescale)	13
5. Valutazione delle sistematiche	14
6. Campioni di validazione	16
7. Conclusioni	17
8. Bibliografia	18

1. INTRODUZIONE

L'epidemia da virus Covid-19 rappresenta un unicum nella storia recente e non solo: è la prima volta che una patologia altamente infettiva si diffonde in un ambiente altamente globalizzato. Non avendo precedenti su cui basarsi, l'unica possibilità di confronto ad oggi è utilizzare i dati a nostra disposizione, in particolare quelli provenienti dalla Cina, dove l'epidemia è esplosa con circa un mese di anticipo rispetto a quella italiana.

Assumendo la correttezza dei dati ufficiali a nostra disposizione, sorprende il numero di decessi italiani rispetto a quelli in Cina, in particolare confrontando Lombardia e Hubei. Nonostante alcune differenze, i due focolai sono confrontabili, e ci si aspetterebbe una maggiore diffusione e conseguentemente mortalità nella provincia cinese. Questo a maggior ragione nel primo periodo, prima dell'emergere di eventuali differenze negli effetti del 'lockdown', o nella reazione di un diverso sistema sanitario, tra Cina e Italia.

Scopo di questo studio è dimostrare come le differenze del numero cumulativo di decessi tra Lombardia ed Hubei possano essere spiegate considerando principalmente la differenza di età tra i due campioni in oggetto.

2. SCELTA DEL CAMPIONE

Il confronto è stato effettuato tra i singoli focolai della regione Lombardia e della provincia di Hubei. L'estensione a livello nazionale implicherebbe la considerazione di una serie di quantità non facilmente controllabili.

Come riportato in Tab.I, nonostante la diversa estensione e popolazione delle due regioni, la densità abitativa è confrontabile.

TABLE I: Caratteristiche regioni [1]

	Lombardia	Hubei
Superficie (km^2)	23 863,65	185 900
Abitanti (ab.)	10 088 484	57 237 740
Densità (ab./ km^2)	423	308

L'analisi può essere effettuata su tre variabili principali:

- numero di contagi

- numero di ricoverati in terapia intensiva
- numero di decessi

Il primo è direttamente correlato non solo al numero di tamponi effettuati, ma anche alle modalità con cui vengono effettuati; inoltre dal 12 Febbraio nella provincia di Hubei il numero di contagi è stato determinato senza utilizzo del tampone (tramite sintomatologia e lastra toracica).

Il secondo ha una dipendenza cruciale dalla disponibilità di posti in terapia intensiva, che in particolare in Lombardia sono in saturazione.

Su questi due punti quindi i campioni cinese ed italiano sono difficilmente confrontabili. Si è scelto quindi di basare lo studio sul numero di decessi, che sembra essere il dato più attendibile a disposizione.

3. ANALISI DEI DATI

In Fig. 1 è presente il confronto diretto tra il numero cumulativo di decessi in Lombardia [2] e ad Hubei [3], mentre in Fig.2 analogo confronto tra i decessi giornalieri. Il primo giorno considerato è per convenzione quello nel quale si superano i 100 decessi cumulativi, rispettivamente il 28 Gennaio per la Cina ed il 6 Marzo per l'Italia, 37 giorni dopo. Emerge chiaramente come in Lombardia il numero di decessi si allontani velocemente dall'andamento in Hubei.

Varie ipotesi sono state prese in considerazione per cercare di spiegare questo effetto, dalle misure tardive della quarantena e dell'istituzione della zona rossa, all'impreparazione della sanità italiana ad un'emergenza di questa portata, passando per la differente età media della popolazione italiana rispetto a quella cinese.

Scopo di questo studio è analizzare se la differenza di andamento dei decessi tra Cina ed Italia sia spiegabile, almeno in prima approssimazione, con la differente età media delle persone decedute, correlata alla diversa età media dei due stati, 45 anni per l'Italia [4] (2018) e 37 per la Cina [5] (2019).

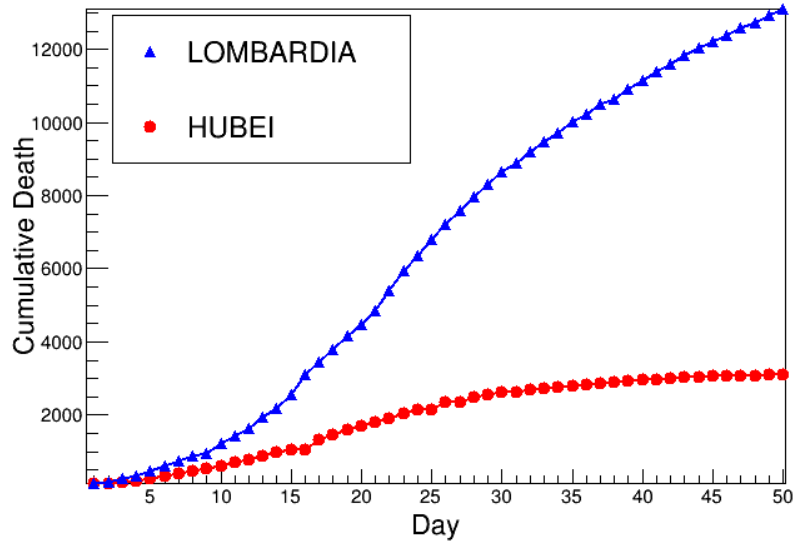


FIG. 1: Confronto tra il numero cumulativo di decessi in Lombardia (blu) [2] ed Hubei (rosso) [3].

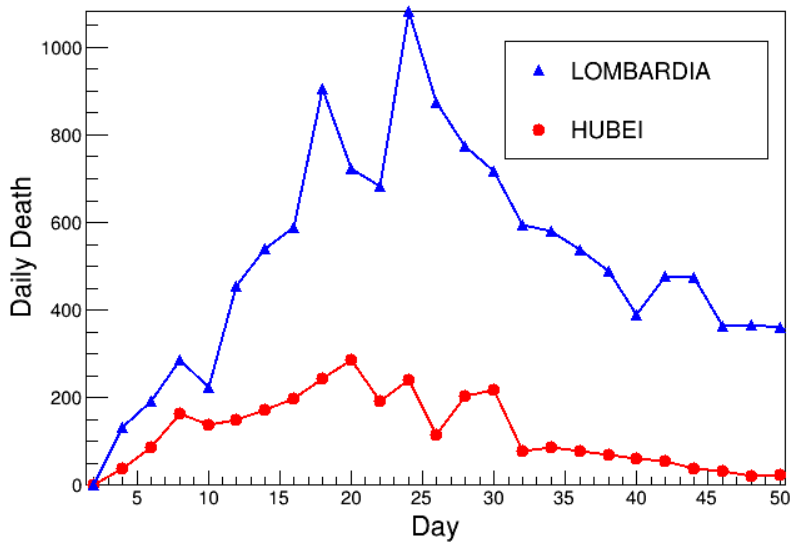


FIG. 2: Confronto tra il numero di decessi bi-giornaliero in Lombardia (blu) [2] e in Hubei (rosso) [3].

4. SCALING DEI DATI E RELATIVO CONFRONTO

L'età media dei decessi in Italia si aggira intorno ad 80 anni [6][7] mentre per la Cina, a partire dai dati disponibili divisi per fascia di età [8], è possibile calcolare un'età media di 66 anni. Simile differenza in età si trova guardando l'età media dei contagi, 63 anni per l'Italia [9] e circa 50 per la Cina (calcolata a partire da [8]). In Tab.II sono riassunti i dati sulle età medie.

TABLE II: Età medie [6][7][8][9]

	Italia	Cina
Età media popolazione	45	37
Età media contagi	63	55
Età media decessi	80	66

Per lo scaling si è scelto di utilizzare l'età media dei decessi. Essendo i dati disponibili divisi per fasce di età di 10 anni, si è inquadrata la Cina nella fascia 60-70. Per l'Italia sono state seguite due ipotesi: una conservativa, collocando l'Italia nella fascia 70-80 (Sez.4 A), e una non conservativa Sez.4 B, utilizzando la media delle fasce 70-80 e 80-90.

Possibili altre scelte sono discusse in Sez. 5.

A. Scaling conservativo (C-rescale)

Come detto, si è scelto di inquadrare la Cina nella fascia 60-70 e l'Italia conservativamente nella fascia 70-80. E' possibile così riscaldare il numero di decessi avvenuti nella provincia di Hubei se l'età media cinese fosse più alta di 10 anni, utilizzando il rapporto tra le mortalità cinesi per le due fasce di età:

$$N_{Death}^{rescaled} = N_{Death}^{real} \times \frac{M_{(70-80)}}{M_{(60-70)}} = N_{Death}^{real} \times S_{\frac{70-80}{60-70}} \quad (1)$$

dove $S_{\frac{70-80}{60-70}}$ è il fattore di scalo, M_i è la mortalità delle singole fasce di età, che può essere calcolata utilizzando vari criteri (si rimanda a [8] per approfondimento); inizialmente si è utilizzato il rapporto diretto tra numero di decessi e numero di contagi (*Crude estimate*); l'impatto della scelta di un criterio è discusso in Sez.5.

I valori delle mortalità in Cina sono rispettivamente (8.5 ± 0.8) % per la fascia compresa 70 -80 anni e (3.8 ± 0.4) % per 60-70 anni; il fattore di scaling risulta quindi essere $S_{\frac{70-80}{60-70}} = \frac{M_{(70-80)}}{M_{(60-70)}} = 2.2 \pm 0.3$. Gli errori sulle singole mortalità, e conseguentemente sul fattore di scala, sono derivati da [8] in approssimazione gaussiana.

Di seguito ci riferiremo a $N^{rescaled}$ come $N^{C-rescaled}$, per intendere l'uso del fattore di scala conservativo $S_{\frac{70-80}{60-70}}$. In Fig.3 è possibile vedere il confronto tra l'andamento dei decessi in Italiae in Hubei riscaldato, mentre in Fig. 6 l'analogo confronto per i decessi giornalieri.

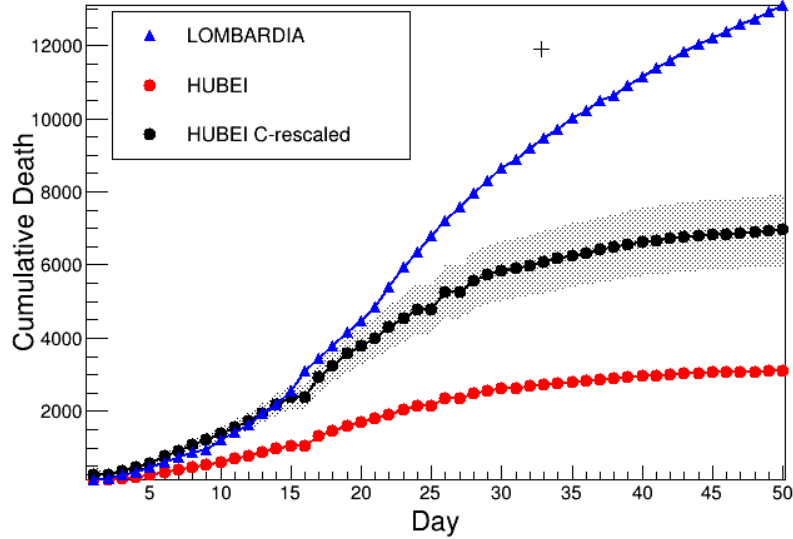


FIG. 3: Confronto tra il numero cumulativo di decessi in Lombardia (blu) [2], Hubei (rosso) [3] ed Hubei riscalatato conservativamente (C-rescaled, nero); La banda di errore è dovuta all'errore sul fattore di scalo.

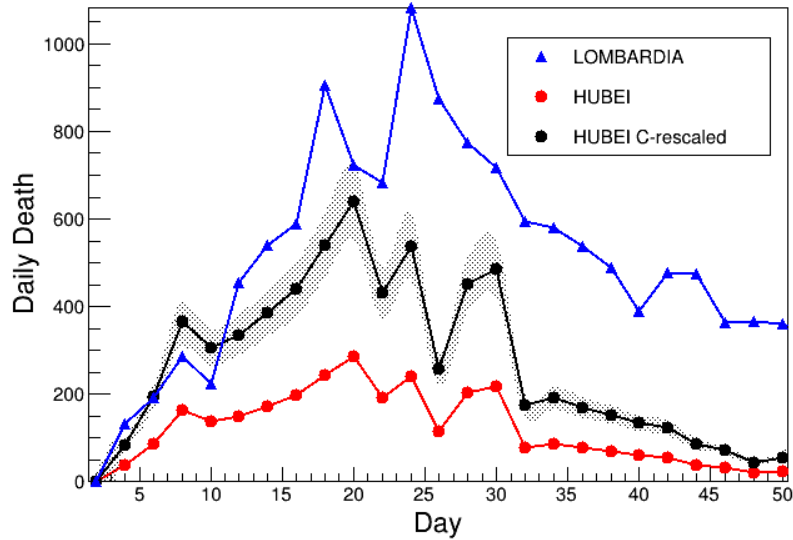


FIG. 4: Confronto tra il numero di decessi bi-giornaliero in Lombardia (blu) [2] e in Hubei (rosso) ed Hubei riscalato conservativamente (C-rescaled, nero) Le bande di errori sono dovute all'errore sul fattore di scala.

B. Scaling mediato (M-rescale)

Visto che l'età media dei decessi in Italia si attesta intorno agli 80-81 anni, l'assegnazione della fascia 70-80, e la relativa mortalità utilizzata al numeratore dei Eq.1 è, come detto, una scelta conservativa. E' possibile ad esempio utilizzare una media delle

mortalità della fascia 70-80 ed 80-100, che chiameremo $M_{70-100} = (11.7 \pm 0.2)\%$, da cui si ottiene $S_{\frac{70-100}{60-70}} = \frac{M_{(70-100)}}{M_{(60-70)}} = 3.1 \pm 0.6$. Di seguito ci riferiremo al numero di decessi di Hubei riscalato usando il fattore di scala non conservativo $S_{\frac{70-100}{60-70}}$ come $N^{M-rescaled}$. In Fig.5 è possibile confrontare l'andamento di questa stima con le precedenti; mentre in Fig.6 l'analogo confronto per i decessi giornalieri.

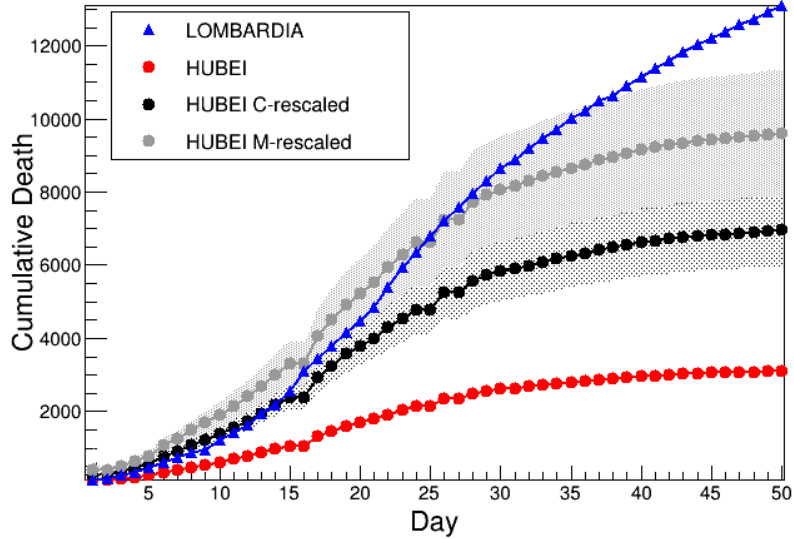


FIG. 5: Confronto tra il numero cumulativo di decessi in Lombardia (blu) [2], Hubei (rosso) [3], Hubei riscalato conservativamente (C-rescaled, nero) ed Hubei riscalato mediato (M-rescaled, grigio). Le bande di errori sono dovute all'errore sul fattore di scala.

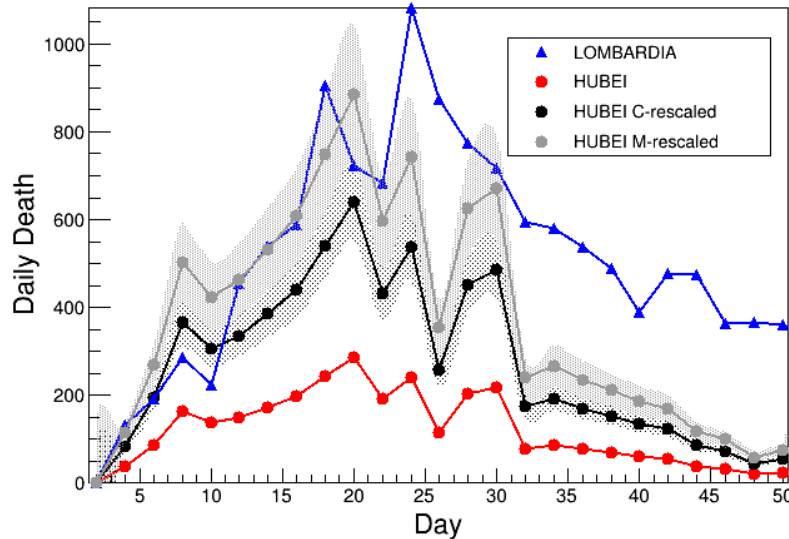


FIG. 6: Confronto tra il numero di decessi bi-giornaliero in Lombardia (blu) [2] e in Hubei (rosso) [3], Hubei riscalato conservativamente (C-rescaled, nero) ed Hubei riscalato mediato (M-rescaled, grigio). Le bande di errori sono dovute all'errore sul fattore di scala; gli errori statistici dei singoli punti sono inferiori alle dimensioni dei marker.

C. Scaling non conservativo (NC-rescale)

E' anche possibile formulare una ulteriore ipotesi, inquadrando la Lombardia nella fascia di età 80-100, che chiameremo $M_{80-100} = (15.0 \pm 1.5)\%$, da cui si ottiene $S_{\frac{80-100}{60-70}} = \frac{M_{(80-100)}}{M_{(60-70)}} = 4.9 \pm 0.6$. Di seguito ci riferiremo al numero di decessi di Hubei riscaldato usando il fattore di scala $S_{\frac{80-100}{60-70}}$ come $N^{NC-rescaled}$. In Fig.7 è possibile confrontare l'andamento di questa stima con le precedenti.

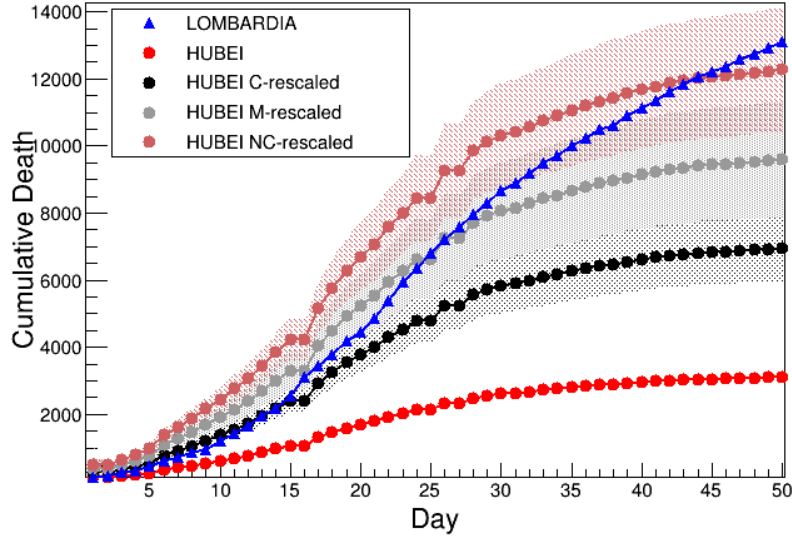


FIG. 7: Confronto tra il numero cumulativo di decessi in Lombardia (blu) [2], Hubei (rosso) [3], Hubei riscaldato conservativamente (C-rescaled, nero), Hubei riscaldato mediato (M-rescaled, grigio chiaro), ed Hubei riscaldato non conservativamente (NC-rescaled, grigio scuro). Le bande di errori sono dovute all'errore sul fattore di scala.

Visto che la convenzionalità sulla scelta del primo giorno da considerare per ciascun sample, è possibile traslare ciascuna distribuzione rispetto all'asse x. Ad esempio utilizzare come riferimento la soglia dei 1500 decessi (invece dei 100), e conseguentemente traslare la distribuzione della Lombardia di 5 giorni, cominciando dall'11 marzo. In questo modo si ottiene un ottimo accordo tra la Lombardia e Hubei non conservativo, come riportato in Fig. 8.

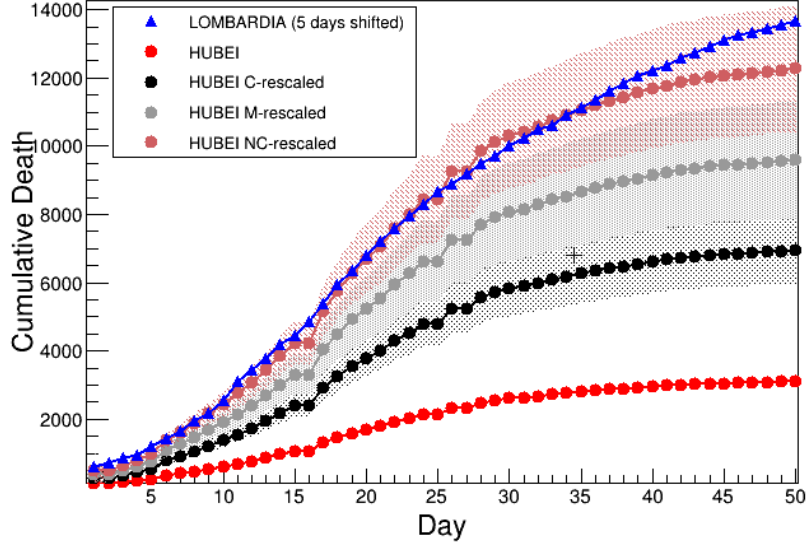


FIG. 8: Confronto tra il numero cumulativo di decessi in Lombardia (blu) [2] traslato di 5 giorni, Hubei (rosso) [3], Hubei riscaloato conservativamente (C-rescaled, nero), Hubei riscaloato mediato (M-rescaled, grigio chiaro), ed Hubei riscaloato non conservativamente (NC-rescaled, grigio scuro). Le bande di errori sono dovute all'errore sul fattore di scala; gli errori statistici dei singoli punti sono inferiori alle dimensioni dei marker.

5. VALUTAZIONE DELLE SISTEMATICHE

La principale sistematica di questo studio riguarda il calcolo del fattore di scaling, che dipende sia dalla scelta delle fasce di età da considerare, che dal metodo di stima della mortalità utilizzato. Riguardo le fasce di età, sono state considerate le età medie dei decessi, inquadrandoli nella fascia 60-70 per la Cina e conservativamente nella fascia 70-80 per l'Italia. In questo modo è stato ottenuto il fattore di scaling S il valore utilizzato in Sez.4 A:

$$S_{\frac{70-80}{60-70}} = \frac{M_{(70-80)}}{M_{(60-70)}} = 2.2 \pm 0.3$$

dove le singole mortalità sono $M_{(70-80)} = (8.5 \pm 0.8)\%$ e $M_{(60-70)} = (3.8 \pm 0.4)\%$ [8].

E' possibile quindi variare il criterio di inquadramento in fascia di età da utilizzare per lo scaling, e il criterio di stima della mortalità.

Nel primo caso, si può privilegiare come criterio l'età media dei contagi (63 anni per l'Italia [9] e circa 50 per la Cina [8]) invece dell'età media dei decessi (Italia 80-81 anni [6][7], Cina 66 anni [8]), mantenendo fisso il criterio di stima della mortalità. In questo

modo si ottiene:

$$S_{\frac{60-70}{50-60}} = \frac{M_{(60-70)}}{M_{(50-60)}} = 2.7 \pm 0.5$$

in accordo con il risultato precedente entro una deviazione standard. Utilizzando questo fattore di scala si otterrebbe un andamento che andrebbe a collocarsi circa a metà tra gli andamento di Hubei riscaldato conservativamente e non conservativamente.

Riguardo il criterio di calcolo della mortalità, in [8] sono presentati quattro differenti criteri:

- Crude estimation (*CrE*), usato nell'analisi presentata in Sez.4
- Adjusted for delayed mortality (*ADJ1*)
- Adjusted for unidentified symptomatic cases (*ADJ2*)
- Adjusted for both (*ADJ3*)

per la descrizione dei differenti criteri si rimanda all'articolo [8].

Utilizzando i diversi criteri si ottiene (per il fattore di scaling $S_{\frac{70-80}{60-70}}$ per fasce standard 70-80 e 60-70):

$$S_{CrE} = 2.24 \pm 0.32$$

$$S_{ADJ1} = 2.23 \pm 0.29$$

$$S_{ADJ2} = 2.16 \pm 0.30$$

$$S_{ADJ3} = 2.13 \pm 0.26$$

in ottimo accordo tra di loro.

Nessuna sistematica dovuta ai criteri di scelta è stata quindi assegnata.

6. CAMPIONE DI VALIDAZIONE

Per validare il metodo di scaling usato, vi è la necessità di individuare un campione di controllo indipendente, con caratteristiche simili a quelli di Lombardia ed Hubei: densità di popolazione, sistema sanitario, singolo focolaio, età media della popolazione, statistica sufficiente. Vista la difficoltà di reperire i dati 'regionali', al momento si è identificato un solo campione utile, la regione spagnola della Catalogna, le cui caratteristiche sono riportate in Tab.III.

TABLE III: Caratteristiche Catalogna [1]

Superficie (km^2)	32 108
Abitanti (ab.)	7 548 825
Densità (ab./ km^2)	235
Età media	43
Età media contagi	60

Avendo caratteristiche simili a quelle italiane, ci si aspetta che l'andamento sia in accordo con quello della Lombardia, e con quello di Hubei riscaldato. In particolare, essendo sia l'età media della popolazione che l'età media dei contagi leggermente più bassa di quella italiana, ci si aspetta un numero di decessi inferiore in Catalogna rispetto alla Lombardia, più vicino agli andamaneti di Hubei riscaldato conservativamente.

Il risultato è presentato in Fig.9, dove il primo giorno considerato (più di cento decessi cumulativi) è il 19 Marzo, 13 giorno dopo l'Italia. Si riscontra un ottimo accordo nel periodo iniziale con l'andamento con quello di Hubei riscaldato conservativamente; in seguito l'andamento sembra mantenere livelli inferiori, probabilmente a causa dell'età media inferiore dei contagiati.

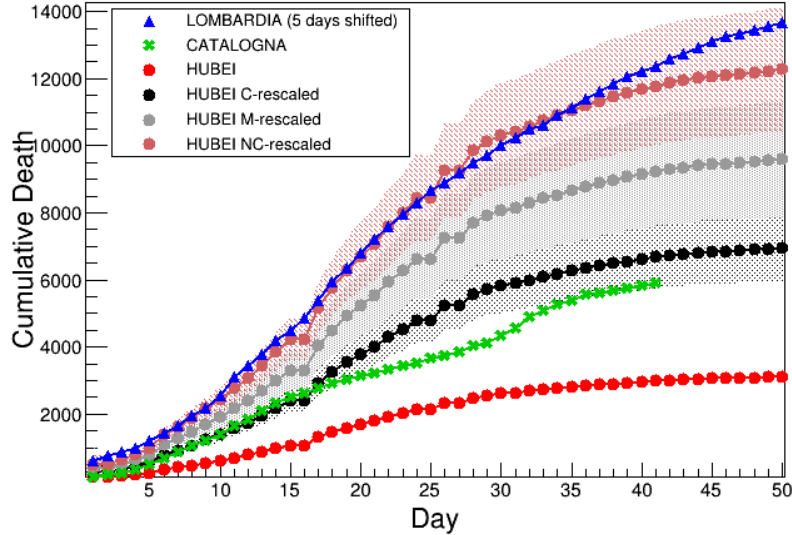


FIG. 9: Confronto tra il numero cumulativo di decessi in Lombardia (blu) [2] traslato di 5 giorni, Catalogna (verde), Hubei (rosso) [3], Hubei riscaldato conservativamente (C-rescaled, nero), Hubei riscaldato mediato (M-rescaled, grigio chiaro), ed Hubei riscaldato non conservativamente (NC-rescaled, grigio scuro). Le bande di errori sono dovute all'errore sul fattore di scala; gli errori statistici dei singoli punti sono inferiori alle dimensioni dei marker.

7. CONCLUSIONI

Questo studio dimostra come in prima approssimazione le differenze del numero cumulativo di decessi tra Lombardia ed Hubei possano essere spiegate considerando unicamente la differenza di età tra i due campioni in oggetto; questo è stato verificato usando sia le età medie dei decessi, che quelle dei contagi. Si è anche verificato che i risultati sono indipendenti dalla modalità con cui la mortalità viene calcolata.

Sono state valutate due diverse possibilità di scaling, dipendenti dalla modalità di inquadramento in fascia di età della mortalità italiana, una conservativa, una mediata e una non conservativa.

Alla luce degli ultimi dati considerati (29 Aprile), utilizzando la soglia dei 100 giorni come giorni di partenza per ciascun campione, il comportamento della Lombardia segue inizialmente Hubei riscaldato conservativamente, per poi superare quello mediato, attestandosi a quello non conservativo.

Traslando la distribuzione della Lombardia di 5 giorni, utilizzando la soglia dei 1500 decessi, si ottiene un ottimo accordo con Hubei riscaldato non conservativamente.

E' necessario tenere in considerazione che l'intervento di variabili non considerate in

questo studio, come l'inefficienza delle modalità di 'lockdown' o la saturazione dei posti in terapia intensiva, potrebbe portare ad un numero di decessi differente da quello di Hubei, nonostante il rescaling, in particolare sul lungo periodo.

Lo studio è stato validato su di un campione analogo ma indipendente, la regione spagnola della Catalogna, trovando ottimo accordo nel periodo con l'andamento conservativo, per poi attestarsi a valori più bassi, probabilmente dovuti anche ad una inferiore età media dei contagi.

-
- [1] <http://www.wikipedia.it/>
 - [2] <http://www.protezionecivile.gov.it/>
 - [3] <https://www.who.int/>
 - [4] Eurostat by <https://www.infodata.ilsole24ore.com/2019/11/10/vecchio-continente-invecchia-leta-media-aumentata-43-anni/>
 - [5] <https://www.worldometers.info/world-population/china-population/>
 - [6] Comunicato Stampa Istituto Superiore di Sanità N.15/2020 "Studio ISS Su 105 deceduti con Covid-2019, età media 81 anni e patologie preesistenti in due terzi dei casi"
 - [7] Comunicato Stampa Istituto Superiore di Sanità CS N. 21/2020 "Coronavirus, letalità fino a questo momento è del 5,8%, più alta negli uomini"
 - [8] <https://www.medrxiv.org/content/10.1101/2020.03.04.20031104v1.full.pdf>
 - [9] https://www.epicentro.iss.it/coronavirus/bollettino/Infografica_23marzo%20ITA.pdf
 - [10] https://public.flourish.studio/visualisation/1538522/?utm_source=embed&utm_campaign=visualisation/1538522

ENEA
Servizio Promozione e Comunicazione
www.enea.it

Stampa: Laboratorio Tecnografico ENEA - C.R. Frascati
maggio 2020