



Official Data and Analytical Forecasts: Differences and Similarities Among Coronavirus Disease (COVID-19) Confirmed Cases and Deaths

Ottavia Eleonora Ferraro¹, Mariangela Valentina Puci¹, Cristina Montomoli¹, Sandro Rolesu², Stefano Cappai² and Federica Loi^{2*}

¹ Unit of Biostatistics and Clinical Epidemiology, Department of Public Health, Experimental and Forensic Medicine, University of Pavia, Pavia, Italy, ² Osservatorio Epidemiologico Veterinario Regionale della Sardegna, Istituto Zooprofilattico Sperimentale della Sardegna G. Pegreffi, Cagliari, Italy

Keywords: COVID-19, pandemic, Italy, public health, standardize mortality rate, official data, risk management

OPEN ACCESS

Edited by:

Zisis Kozlakidis,
International Agency For Research On
Cancer (IARC), France

Reviewed by:

Matteo Floris,
University of Sassari, Italy

*Correspondence:

Federica Loi
federica.loi@izs-sardegna.it

Specialty section:

This article was submitted to
Infectious Diseases - Surveillance,
Prevention and Treatment,
a section of the journal
Frontiers in Medicine

Received: 22 April 2020

Accepted: 06 May 2020

Published: 19 May 2020

Citation:

Ferraro OE, Puci MV, Montomoli C,
Rolesu S, Cappai S and Loi F (2020)
Official Data and Analytical Forecasts:
Differences and Similarities Among
Coronavirus Disease (COVID-19)
Confirmed Cases and Deaths.
Front. Med. 7:239.
doi: 10.3389/fmed.2020.00239

INTRODUCTION

The novel coronavirus disease (COVID-19) is spreading widely with an exponential growth infection rate in several countries worldwide: up to May 5th, 2020, about 3,517,345 cases and 243,401 deaths have been confirmed (1). In Europe overall, about 1.5 million official cases have been reported, and Spain, Italy, and the United Kingdom are the most affected countries. From the 31th of January, in order to better control the virus spread, the Italian government declared an “emergency state,” which is characterized by the implementation of massive containment measures (2, 3).

As now in China, the risk of COVID-19 spreading to other countries is a great concern, as well as the perspective of a secondary cases wave, and given that no vaccine is currently available, rapid and specific diagnostic procedures are an essential tool to allow accurate information of the disease (4, 5). Furthermore, reliable and timely data are fundamental tools to guide the right political and health interventions and to better understand the virus spread.

Since the first Italian spread of the disease from the highest risk area (Northern Italy) to the rest of the nation (the 2nd of March, 2020), the Italian Department of Civil Defense (DCD) have published official reports on COVID-19 distribution to all the Italian regions and provinces (6). The daily regional reports have provided data about number of tests executed (“Tamponi”) and the total COVID-19 cases (“Casi totali”), and details cover recoveries (“Dimessi/Guariti”) and the number of people who have died (“Deceduti”), who are hospitalized with symptoms (“Ricoverti con sintomi”), who are hospitalized in Intensive Care Units (ICUs) (“Terapia intensiva”), and who are in house isolation (“Isolamento domiciliare”). The smallest administrative units in which the data are aggregate are the Italian provinces. All these data are currently used by several scientists, stakeholders, and politicians to understand the daily disease evolution and forecast the possible disease spread in Italy (7–9).

Moreover, it should be noted that the estimates should be critically evaluated to identify their weaknesses and strengths. Data provided in the Italian general reports are based on the province and municipality of patient’s residence. This generates a major critical issue on the correct identification of the municipality where virus spread occurred, given that, when applied to COVID-19 cases detected in Assisted Health Residences (AHR), it may generate misleading information. Indeed, given that not all municipality territories includes AHR, several elderly people are hospitalized in commons different from their residence. In these cases, the formal residence

of patients has not necessarily changed, and no national regulation about this is available. As a consequence, most of COVID-19 cases that became infected inside the AHR are registered as if they have become infected in the municipality where they were a “resident.” The incorrect attribution of the viral contagion location causes the spread of misleading information on the geographical incidence of the disease as well as the infection, recovery, and lethality estimation rates.

In fact, one of the main important things in epidemiological studies is the ability to define as accurately as possible the reference population, i.e., the rate denominator (tested population). In this context, the number of COVID-19 laboratory tests executed, the sampling method, as well as the number of people tested, is a controversially issue. The choice of people to be tested is regulated by National Law¹, but each Region can apply most specific rules. Furthermore, in hospital, the local management can decide on the testing regime used for personnel.

Finally, all this leads to a situation where the number of tests executed does not reflect the number of people tested in specific areas, as the same person may be tested one, two, or three times. Only the declaration of recovery is bound by the performance of at least three tests (one to be declare positive and two to be declare recovered)¹. Based on this assumption, the number of people tested could be better estimated by subtracting the double number of recoveries from the number of tests executed. However, also this estimation is probably not fully reliable given that healthcare staff could be tested more than once to confirm the negative state of COVID-19, and this would invalidate the data of the reference population.

We must consider the time taken to provide a laboratory result: as reported by many sources, although the maximum time to communicate the swabs result should be 36 h, the results are frequently delayed due to the overload of laboratories. Therefore, the number of positive/recovered cases reported daily could be imprecise and include swabs results related to several days before. In any case, the daily data reported are not comparable, and recovery/lethality/infection rates cannot be properly estimated. Methods proposed by Ghiani et al. recently applied for the Sardinia region, could be more appropriate (10, 11). Furthermore, the declaration of death related to COVID-19 needs to be confirmed by usual and official laboratory tests based on swabs, and these create a well-known problem of the underestimation of deaths. Several estimations for the real numbers of COVID-19 deaths have been provided not only for Italy but also for China, South Korea, and the European Union (12).

STANDARDIZED MORTALITY RATES ANALYSIS

The Italian National Institute of Statistic (ISTAT) provides official data on the number of deaths for all causes at the municipality level, and they have done so for January, February, March, and first 2 weeks of April of 2015–2020, in a selected number of

municipalities (13). These data have been used to evaluate the hypothesis that an increment of the mortality rate could be related to COVID-19 causes.

Data regarding 6,866 municipalities (80% of the total) in 20 Italian regions, divided into 21 age classes (each lasting 5 years) and 3.5 months for a total of 914,622 records, have been recorded by the ISTAT website (13). A total of 247,978 observations have been excluded since 2020 data were not available, and the analysis has been performed based on 4,433 Italian municipalities.

As demonstrated in several previous studies, the standardized mortality rates (SMRs) are generally used to compare the observed event (i.e., mortality in 2020) in the cohort under study with the expected one, which is obtained using the rate of events in a reference population (i.e., mortality during the period between 2015 and 2019) (14).

Collecting data at a regional level, based on resident population at each year in the study (available at: <http://demo.istat.it/index.html>), mortality rates \times 10,000 people (observed mortality rate) have been calculated by age classes by taking into account the observed data from the 1st of January to the 15th of April of each year (2015–2020). By multiplying the median age-mortality rate of the 2015–2019 period by the resident population of 2020 for each age class, the expected age-mortality rate for the 2020 period has been estimated for each Italian region. The SMR, obtained by the ratio between observed and expected 2020 mortality rate, has been calculated for each age class at a regional level, and the same has been done for its 95% Confidence Intervals (95% CI). The exact confidence interval was calculated for each estimated SMR assuming a Poisson process (15).

An SMR $>$ 1 shows an excess of mortality, while an SMR $<$ 1 whos a “shortage” in mortality. If 95% CI includes the null value “1,” which cannot be considered statistically significant, the interpretation is that there is no significant excess/deficit in the mortality rate in the studied population compared to the general population.

The results, reported in **Figure 1**, underline an overall excess of deaths in Italy amongst elderly people over 75 years of age in 15 Italian regions in March 2020 compared to previous years, except for Basilicata, Calabria, Campania, Lazio, and Sicilia; January and February did not show statistically significant differences in mortality rates. A total of 10 regions (Calabria, Emilia-Romagna, Liguria, Lombardia, Marche, Piemonte, Toscana, Trentino-Alto Adige, Valle d’Aosta, and Veneto) have confirmed this increment in the first 2 weeks of April of 2020. The largest increase in mortality rate was detected in Lombardia (SMR = 2.929; 95% CI = 2.887–2.971) in March and in Valle d’Aosta (SMR = 2.647; % CI = 2.165–3.205) in April. The lowest excess in mortality rate was recorded in Puglia, which ahd an SMR value equal to 1.083 (95% CI = 1.041–1.127) in March and in Calabria in April (SMR = 1.162; 95% CI = 1.025–1.313).

Given the increasing attention to the consequences of COVID-19 amongst elderly people (10), the observed excess of SMR in this category at national level has been studied more thoroughly in the elderly Italian region (Sardinia). The Sardinian results, obtained by analyzing data from 327 municipalities (20,045 records), illustrate an excess in mortality rates during March 2020 with a statistically significant SMR equal to 1.134

¹<http://www.trovanorme.salute.gov.it/norme/renderNormsanPdf?anno=2020&codLeg=73799&parte=1%20&serie=null>

Italian Region	SMR (95% CI) March, 2020	SMR (95% CI) April, 2020	Elderly Median Mortality Rate × 10,000 people (I-III quartile), March 2015-2019	Elderly Median Mortality Rate × 10,000 people (I-III quartile), April 2015-2019	Elderly Observed Mortality Rate × 10,000 people March 2020	Elderly Observed Mortality Rate × 10,000 people April 2020
Abruzzo	1.121 (1.052-1.193)*	1.102 (0.979-1.235)	55.135 (54.843-56.525)	16.527 (16.352-18.232)	61.826	18.213
Basilicata	0.937 (0.840-1.043)	1.044 (0.767-1.389)	54.099 (48.036-56.418)	6.734 (6.486-7.344)	50.731	7.033
Calabria	1.009 (0.949-1.071)	1.162 (1.025-1.313)*	51.036 (50.061-55.539)	10.658 (10.610-10.850)	51.514	12.386
Campania	0.975 (0.939-1.013)	0.991 (0.901-1.088)	54.347 (53.791-54.647)	8.727 (8.230-9.607)	53.030	8.648
Emilia-Romagna	1.737 (1.693-1.783)*	1.655 (1.588-1.725)*	57.888 (57.613-61.996)	23.776 (22.932-23.804)	100.609	39.365
Friuli-Venezia Giulia	1.142 (1.065-1.223)*	0.877 (0.736-1.038)	43.088 (41.071-44.031)	9.280 (8.889-9.966)	49.214	8.142
Lazio	0.939 (0.906-0.973)	0.836 (0.748-0.932)	49.200 (47.173-49.700)	6.013 (5.705-6.094)	46.216	5.029
Liguria	1.599 (1.533-1.667)*	1.860 (1.747-1.980)*	56.817 (55.712-58.209)	21.941 (21.622-22.661)	90.852	40.824
Lombardia	2.929 (2.887-2.971)*	2.406 (2.345-2.468)*	55.087 (55.041-57.560)	20.925 (20.832-21.2628)	161.362	50.352
Marche	1.546 (1.472-1.622)*	1.688 (1.535-1.854)*	52.349 (51.961-52.426)	12.841 (11.442-13.466)	80.911	21.687
Molise	1.159 (1.013-1.322)*	0.790 (0.585-1.045)	49.345 (48.231-50.138)	15.770 (14.840-18.367)	57.230	12.463
Piemonte	1.545 (1.503-1.588)*	1.903 (1.821-1.988)*	56.454 (55.648-59.856)	17.682 (17.230-17.920)	87.257	33.645
Puglia	1.083 (1.041-1.127)*	1.042 (0.972-1.117)	52.949 (51.106-55.477)	17.810 (16.935-18.153)	57.362	18.569
Sardegna	1.134 (1.064-1.207)*	0.851 (0.746-0.967)	45.067 (38.295-45.472)	14.511 (13.148-14.666)	51.1237	12.349
Sicilia	0.995 (0.958-1.032)	0.932 (0.869-0.999)	54.597 (53.379-55.683)	16.441 (16.383-17.579)	54.326	15.331
Toscana	1.182 (1.142-1.225)*	1.224 (1.161-1.291)*	53.912 (53.019-56.196)	22.638 (21.276-22.671)	63.769	27.719
Trentino-Alto Adige	1.715 (1.613-1.821)*	2.292 (2.045-2.561)*	53.678 (51.355-53.946)	11.842 (11.089-12.082)	92.057	27.1417
Umbria	1.101 (1.024-1.180)*	1.018 (0.891-1.159)	58.628 (55.554-60.046)	18.624 (18.558-19.359)	64.501	18.971
Valle d'Aosta	1.846 (1.569-2.157)*	2.647 (2.165-3.205)*	55.420 (51.772-65.406)	25.682 (19.912-29.788)	102.318	67.996
Veneto	1.234 (1.195-1.273)*	1.375 (1.308-1.445)*	53.872 (50.585-53.988)	20.012 (19.683-20.777)	66.470	27.521

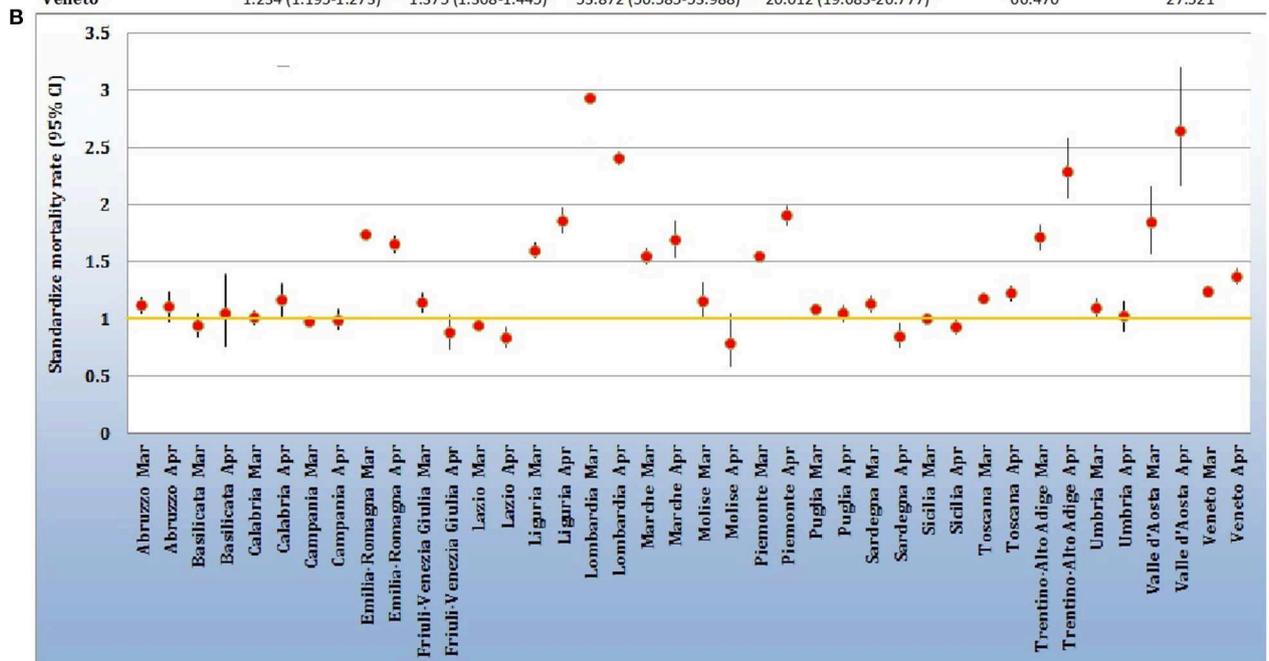


FIGURE 1 | (A) Table reporting regional data of Standardize Mortality Rates and 95% Confidence Interval (SMR, 95% CI); median value of mortality rate × 10,000 people in elderly category (75 years or more), related to March and April of 2015–2019; mortality rate × 10,000 people in elderly category (75 years or more) related to March and April of 2020. **(B)** Forest plot reporting the Standardize Mortality Rates values with 95% CI for the 20 Italian Regions for March and April, 2020. The yellow line represents the limit of one, defining the statistical significance (if not included in the 95% CI).

(95% CI 1.064–1.208). In a total of 20 municipalities, a statistically significant SMR > 1 was detected: Assolo, Bottidda, Burcei, Cabras, Calangianus, Galtelli, Gesturi, Monastir, Ossi, Pabillonis, Padru, Porto Torres, Riola Sardo, San Gavino Monreale, Sanluri, Sassari, Sant’Antioco, Siurgus Donigala, Sorgono, and Vallermosa.

Considering the absence of complete national data on COVID-19 related deaths at the municipal level, as well as the lack of information in Italian provincial reports, the issue of death underestimation has been evaluated by

comparing SMR results with official COVID-19 cases available for Sardinia (16). Among the 20 municipalities with statistically significant SMR, 11 municipalities (Assolo, Bottidda, Burcei, Galtelli, Gesturi, Monastir, Ossi, Pabillonis, Riola Sardo, Siurgus Donigala, and Sorgono) recorded zero COVID-19 cases during March 2020.

Since public health authorities often need to compare the mortality based on geographical areas, the present work provides a robust overview for those Italian regions with high difference in mortality rates caused by the spread of the pandemic in

Italy. The interesting example of excess in mortality, but not officially declared COVID-19 cases, deserves the attention of the health authorities. Furthermore, it must be underlined that the SMRs obtained are underestimated, given that the 2020 mortality rate estimate includes deaths related to incidents in or outside of the work place, or road accidents, which are drastically reduced due to the lockdown period. The inclusion of 80% of the total Italian municipalities has, at least partially, limited the possible bias, providing an important start point for the estimation of real COVID-19 pandemic consequences. However, taking into account the bias of the usage of the 2019 resident population to estimate the 2020 mortality rate, the overall excess of people who have died is about 16,000 deceased, which is in line with the estimates reported in the last ISTAT report (17).

DISCUSSION

As reported by the World Health Organization in their Pandemic Influenza Risk Management Guidelines (18), influenza pandemics are unpredictable but recurring events, and advance planning and preparedness are critical to help mitigate the impact of a pandemic. Furthermore, taking into account the lessons learned from previous pandemics is a fundamental part of ensuring the adequacy of health strategies in the field.

To date, the health organizations have tried to cope with the emergency; however, a better local health organization is now necessary and can be applied. These problems in official data generate an important issue related to the information provided by the authorities. This is likely not only an Italian issue, but Italy is merely an example for a general need for improved healthcare information collection systems.

REFERENCES

- WHO. *Coronavirus Disease (COVID-19) Outbreak situation*. WHO. (2020). Available online at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200408-sitrep-79-covid-19.pdf?sfvrsn=4796b143_6;https://covid19.who.int/ (accessed May 2, 2020).
- Italian President of the Council of Minister. *Decree of 23 February 2020. Italy: DECRETO-LEGGE 23 febbraio 2020*, n. 6, (2020). Available online at: <https://www.gazzettaufficiale.it/eli/gu/2020/02/23/45/sg/pdf> (accessed March 23, 2020).
- Italian President of the Council of Ministers. *Decree of 1 March 2020. DECRETO DEL PRESIDENTE DEL CONSIGLIO DEI MINISTRI 1 marzo 2020*. (2020). Available online at: <https://www.gazzettaufficiale.it/eli/gu/2020/03/01/52/sg/pdf> (accessed March 23, 2020).
- Leung K, Wu JT, Liu D, Leung GM. First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment. *Lancet*. (2020). 395:1382–93. doi: 10.1016/S0140-6736(20)30746-7
- Hong N, He J, Ma Y, Jiang H, Han L, Su L, et al. Evaluating the secondary transmission pattern and epidemic prediction of the COVID-19 in metropolitan areas of China. *Front Med*. (2020) 7:171–9. doi: 10.3389/fmed.2020.00171
- Dipartimento della Protezione Civile. *COVID-19 Italia - Monitoraggio della situazione*. (2020). Available at <https://github.com/pcm-dpc/COVID-19/tree/master/schede-riepilogative> (accessed April 14, 2020)
- Remuzzi A, Remuzzi G. COVID-19 and Italy: what next? *Lancet*. (2020). 2:10–3. doi: 10.1016/S0140-6736(20)30627-9
- Caccavo D. Chinese and Italian COVID-19 outbreaks can be correctly described by a modified SIRD model. *medRxiv*. (2020). doi: 10.1101/2020.03.19.20039388
- Jia W, Han K, Song Y, Cao W, Wang S, et al. Extended SIR prediction of the epidemics trend of COVID-19 in Italy and compared with Hunan, China. *Front Med*. (2020) 8:153–63. doi: 10.3389/fmed.2020.00169
- Puci MV, Loi F, Ferraro OE, Cappai S, Rolesu S, Montomoli C. COVID-19 trend estimation in the elderly Italian region of Sardinia. *Front Public Health*. (2020) 8:153. doi: 10.3389/fpubh.2020.00153
- Ghani AC, Donnelly CA, Cox DR, Griffin JT, Fraser C, Lam TH, et al. Methods for estimating the case fatality ratio for a novel, emerging infectious disease. *Am J Epidemiol*. (2005) 162:479–86. doi: 10.1093/aje/kwi230
- Rajgor DD, Lee MH, Archuleta S, Bagdasarian N, Quek SC. The many estimates of the COVID-19 case fatality rate. *Lancet*. (2020). doi: 10.1016/S1473-3099(20)30244-9. [Epub ahead of print].
- ISTAT – Istituto Nazionale di Statistica. (2020). Available online at: <https://www.istat.it/it/files/2020/03/dati-settimanali-comune-16aprile.zip> (accessed April 22, 2020).
- Kelsey JL, Gold EB. Observational epidemiology. In: *International Encyclopedia of Public Health*. Davis, CA: Elsevier Inc. (2016). p. 295–307. doi: 10.1016/B978-0-12-803678-5.000310-6
- Breslow NE and Day NE. *Statistical methods in cancer research: Vol. 2. The Design and Analysis of Cohort Studies*. Lyon: IARC. (1987).
- Regione Sardegna. *Coronavirus nell'isola, la situazione all'8 Aprile* (Coronavirus in Sardinia: update at 8th April, 2020). Available online at: <https://www.lanuovasardegna.it/foto/2020/04/08/fotogalleria/coronavirus->

The main focus that needs to be taken into account should be the localization of the virus and not only of the infected people. Thus, the correct identification of infected people and their localization is essential for a robust epidemiological analysis and mortality rate estimation. The hypothesized subsequent phase (Italian Phase 2) must necessarily be carried out on the basis of these assumptions. This will allow for, as much as possible, the understanding of the true prevalence of the disease compared to the official cases diagnosed.

The survey should be based on the most relevant characteristics of the population (i.e., sex, age, residence, comorbidities, and symptoms) in order to provide a valid risk analysis and predict the spread of infection. Future investigations could start based on this increase in mortality rate at municipal level and testing the personal contacts of the deceased. On the other hand, simple corrections in data collection and its transmission (i.e., time of sample, localization by residence or hospitalization, and number of swabs/person) could be fundamental tools with which to plan the next steps. This could be applied first in regions with a low population, where additional field sanitary measures will facilitate faster virus localization and promote a spatial epidemiological analysis. Collaboration amongst nations should be encouraged, as that the virus is not bounded by geographical limits.

AUTHOR CONTRIBUTIONS

FL performed data collection and *ad hoc* database and supported the statistical analysis on SMR performed by OF. OF, MP, CM, SR, FL, and SC contributed equally in conceptualization, literature search, and writing, and SC and OF took care of the figures. All authors contributed equally in study design, data interpretation, and revision of the manuscript.

- in-sardegna-le-mappe-della-situazione-all-8-aprile-a-sassari-piu-di-500-contagi-1.38695370 (accessed April 09, 2020).
17. ISTAT. *Impatto dell'epidemia covid-19 sulla mortalità totale della popolazione residente primo trimestre 2020* (2020). Available online at: https://www.istat.it/it/files//2020/05/Rapporto_Istat_ISS.pdf (accessed April 22, 2020).
 18. WHO. *Pandemic Influenza Risk Management Guidance* (2017). Available online at: [https://apps.who.int/iris/bitstream/handle/10665/259893/WHO-WHE-IHM-GIP-2017.1-eng.pdf;jsessionid=\\$979C2680AD203C5E52B84C6A50E5BA57?sequence=\\$1](https://apps.who.int/iris/bitstream/handle/10665/259893/WHO-WHE-IHM-GIP-2017.1-eng.pdf;jsessionid=$979C2680AD203C5E52B84C6A50E5BA57?sequence=$1) (accessed April 22, 2020).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Ferraro, Puci, Montomoli, Rolesu, Cappai and Loi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.