Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Biological attributes of age and gender variations in Indian COVID-19 cases: A retrospective data analysis

Savitesh Kushwaha, Poonam Khanna, Vineeth Rajagopal, Tanvi Kiran *

Department of Community Medicine and School of Public Health, Post Graduate Institute of Medical Education and Research, Chandigarh, 160012, India

ARTICLE INFO

Keywords:
- Age
- COVID-19
- Gender
- Mortality
- Recovery

ABSTRACT

Background: The associated risk factors, co-morbid conditions and biological differences varying with gender and age might be the cause of higher COVID-19 infection and deaths among males and older persons. The objective of this study was to predict and specify the biological attributes of variation in age and gender-based on COVID-19 status (deceased/recovered).

Methods: In this retrospective study, the data was extracted from a recognised web-based portal. A total of 112,860 patients' record was filtered out and an additional 9131 records were separately analysed to examine age and gender relationship with patient’s COVID-19 status (recovered/deceased). Chi-square, t-test, binary logistic regression, and longitudinal regression analysis were conducted.

Results: The male COVID-19 cases (65.39%) were more than females (34.61%) and mean age of infected and recovered patients was 39.47 ± 17.59 years and 36.85 ± 18.51 years respectively. The odds for infection was significantly higher among females for lower age categories, which declines with age. The age-adjusted odds for recovery were significantly higher among females (O.R. = 1.779) and odds for recovery was highest in 5–17 years age category (O.R. = 88.286) independent of gender.

Conclusion: The chances of being COVID-19 infected was higher for females of lower age categories (<35 years) which decreases with age. The odds for recovery among females was significantly higher than males. The chances of recovery declines with increasing age and the variation could be attributed to the biological differences between age categories and gender.

1. Introduction

The COVID-19 pandemic has challenged the global health systems and medical sciences. The SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus-2) virus has infected about 10% of the global population and since the end of May 2020, India is facing an upsurge of COVID-19 cases and has been ranked as 2nd most affected country in the world. Several assumptions were made for the differential impact of COVID-19 on age and gender. The environmental and socio-economic factors were also utilised for depicting the vulnerability to COVID-19. The sociological factors like jobs, daily activities, economy, literacy etc. put individual differentially towards diseases. The susceptibility to external pathogen also differs due to biological differences at various age and between the genders (male and female).

The immune system plays a crucial role in the prevention from various microorganisms, including viruses. The normal human immune system adapts during the fetal to infant stage, matures during adolescent to the adult stage with variability during pregnancy and decreases as the senescence approaches. These fluctuations in the immune system throughout life poses a higher risk for complications in infants, pregnant women, and elderly. There are several components involved in differentiating of immune system based on gender and age. The variations in the level/count of immunoglobins, CD4 and CD8 cells, B-cells, T-cells among males and females might be causing the variation in COVID-19 cases and deaths.

The publication of studies stating the binding of SARS-CoV-2 with Angiotensin-converting enzyme 2 (ACE2) and some studies claiming a strong binding affinity to human ACE2, evoked a debate on continuation of ACEIs and ARBs (Angiotensin II receptor blockers) drugs among Cardiovascular Diseases patients. However, there is a lack of evidence to prove a significant increase in ACE2 levels, subsequent ACEIs and ARB’s dosages. Due to COVID-19 the association of ACE2 with gender and...
age also came in light after some studies showed differences in COVID-19 infection and death rates based on age and gender.\textsuperscript{10}

The COVID-19 positive cases are continuously increasing in India, and little evidence is available highlighting the age and gender perspectives of this disease. Therefore, the present study was an attempt to study the association between age gender and patients’ status (recovered/deceased) among positive COVID-19 cases and also discussed the possible biological reasons for the variations among different age categories and gender. The present analysis is expected to provide evidence for framing age and gender-specific public health policies and treatment of COVID-19 infections.

2. Methods

The present study is based on secondary data which was extracted from a recognised web-based portal (https://api.covid19india.org/documentation/csv/), which is the open-source portal/page of COVID-19 tracking website (https://www.covid19india.org/). The portal keeps a real-time track record of COVID-19 cases in India and allows the viewer to access the raw (MS Excel) form of data under various categories like time-series cases, state-wise, district-wise etc. The data available on the portal was in accordance with the official estimates and was sourced mainly from various government websites, government Twitter accounts updates and newspapers. The time-series cases data available in MS Excel Sheets from ‘raw_data1’ to ‘raw_data16’ were downloaded. The cases reported between 01 and 03-2020 to 30-09-2020 (7 months) were filtered in a separate file. The raw and coded data was uploaded to Mendeley data.\textsuperscript{17} A total of 112,860 records were separated based on nationality, date of COVID-19 positive, age, gender (male/female), and patient’s COVID-19 status (recovered/hospitalised/deceased). From these extracted records, 9131 records were separated based on the status (recovered/deceased) of COVID-19 patients.

3. Data analysis

A retrospective secondary data analysis was performed for COVID-19 cases in India. Descriptive statistics were used for reporting the mean, standard deviation, percentage and count under various categories and sub-categories. After cleaning and coding of data, appropriate statistical analysis was performed for group comparisons and estimating relationships. The age and gender records for N = 112,860 and age, gender, and patient’s status records for n = 9131 were analysed separately (Fig. 1). The age was divided into five categories (0–4 years, 5–17 years, 18–35 years, 36–55 years and 56 years & above). The recovery and deceased estimates were not in accordance with national estimates therefore, longitudinal regression modelling was conducted for n = 9131 records. The chi-square test was used to estimate the differences between categorical variables (age categories, gender) and independent t-test was performed to estimate the differences in mean values of the continuous variable (age) across gender. The binary logistic regression analysis was also used to examine the relationship between the dependent and independent variables. The Nagelkerke R\textsuperscript{2} (pseudo R\textsuperscript{2}) and Hosmer Lemeshow tests were further used for the estimation of model performance and goodness of fit respectively. The Odds ratio (ORs) along with 95% confidence interval were used to assess the risk among gender and age.

For all these analyses the females and males were coded as ‘1’ and ‘2’ respectively. The deceased was coded as ‘1’ and recovered as ‘2’. Among age categories, the lowest age category i.e., 0–4 year were coded as ‘1’ then, 5–17 years as ‘2’, 18–35 years as ‘3’, 36–55 years as ‘4’ and 56 years and above as ‘5’. The statistical significance was considered at p-value of <0.05 and all the statistical tests were performed using IBM SPSS Version 26.0.

4. Results

The extracted data includes the details of 73,797 (65.39%) males and 39,063 (34.61%) females COVID-19 patients. Most COVID-19 cases (37.48%) were found in the 18–35 years age category (Table 1). The Highest distribution of infected male (37.21%) and female (37.99%) patients was within 18–35 years age category and the lowest distribution of male (1.28%) and female (2.19%) patients were found among 0–4 years age category (Fig. 2). Moreover, the difference between gender among age categories was found to be statistically significant (p < 0.001) (Table 1). The mean age of all COVID-19 patients was 39.47 ± 17.59 years and the difference between the mean age of males (39.98 ± 17.19 years) and females (38.50 ± 18.29 years) was also statistically significant (p < 0.001) (Table 1), with a mean age of COVID-19 infected males being significantly higher than that of the female counterparts.

![Fig. 1. STROBE flowchart of study methodology.](image-url)
The logistic regression analysis between age categories and gender shows significantly higher odds for infection among females in all age categories except for 36–55 years age category after considering 56 years & above age categories as constant. The odds of being COVID-19 infected females was high in 0–4 years (O.R. = 1.723, 95% C.I. = 1.564–1.898), 5–17 years (O.R. = 1.439, 95% C.I. = 1.365–1.517) and 18–35 years (O.R. = 1.030, 95% C.I. = 0.996–1.066) age category and it decreases among 36–55 years (O.R. = 0.894, 95% C.I. = 0.864–0.926) age categories (Table 2). Therefore, the odds of COVID-19 infection are higher among females in comparison to males for lower age categories and it decreases with the increasing age. The Nagelkerke R² shows that the model is explaining 0.6% of the results and Hosmer Lemeshow tests show that the model is a good fit (p = 1.000) (Table 2).

Another set of longitudinal analysis was carried out for n = 9131 records which include the patient’s COVID-19 status (recovered/deceased) details along with gender and age. Among the total number of deceased and recovered patients, the proportion of deceased patients increases with age, whereas the proportion of recovered patients increases until 18–35 years (Fig. 3). The significant mean age difference (p < 0.001) was observed between recovered (36.85 ± 18.51) and deceased (59.99 ± 14.36) categories. Overall patient deceased (94.50%) was greater than the patient recovered (5.50%) in currently assessed records (Supplementary File).

The binary logistic regression model predicted significant odds of recovery among 0–4 years (O.R. = 67.604, 95% C.I. = 34.872–131.059), 5–17 years (O.R. = 88.286, 95% C.I. = 54.423–143.217), 18–35 years (O.R. = 25.121, 95% C.I. = 19.202–32.865) and 36–55 years (O.R. = 5.50%, 95% C.I. = 1.765, 95% C.I. = 1.442–2.159) and increases for males (O.R. = 0.562, 95% C.I. = 0.460–0.687).

5. Discussion

The SARS-CoV-2 virus being a novel virus can infect human race irrespective of age categories and gender. However, there exist individual variations in physiological functions, immune responses and risk factors across gender and age. Therefore, the chances of getting infected might vary among gender and different age categories. The present study attempted to explore the category wise (age and gender) chances of getting infected and predicted chances of getting recovered or deaths with increasing age in the Indian population. In this study, 112,860 COVID-19 patients records were extracted and analysed from a web-based portal to determine the role of age and gender determining COVID-19 status among Indian population. Various studies across the globe show that the older males were more susceptible (>50%) in getting infected by SARS-CoV-2.4,11,12,20 60.3% of all cases were found to be males in a study among 5700 hospitalised COVID-19 patients in the U.

Table 1

| Age categories/ Gender | Male Mean ± S.D./ N (%) | Female Mean ± S.D./ N (%) | Total Mean ± S.D./ N (%) | p-value |
|------------------------|-------------------------|---------------------------|-------------------------|---------|
| Age                    |                         |                           |                         |         |
| 0–4 years              | 39.98 ± 17.19           | 38.50 ± 18.29             | 39.47 ± 17.59           | <0.001  |
| 5–17 years             | 946 (52.53)             | 855 (47.47)               | 1801 (1.60)             | <0.001  |
| 18–35 years            | 4457 (56.98)            | 3365 (43.02)              | 7822 (1.60)             |         |
| 36–55 years            | 27459 (64.91)           | 14841 (35.09)             | 42300 (37.48)           |         |
| 56 years & above       | 26545 (68.07)           | 12454 (31.93)             | 38999 (34.56)           |         |
| Total                  | 73797 (65.39)           | 39063 (34.61)             | 112860 (100.00)         |         |

The COVID-19 status among Indian population. Various studies across the globe show that the older males were more susceptible (>50%) in getting infected by SARS-CoV-2.4,11,12,20 60.3% of all cases were found to be males in a study among 5700 hospitalised COVID-19 patients in the U.
The Nagelkerke R² shows that the model is explaining 26.4% of the results and Hosmer-Lemeshow tests show that the model is a good fit (p = 0.895). Reference Category = “males” and “56 years & above” Predicted probabilities for “recovered”.

The results can be substantiated and explained based on the evidence that; the male adaptive immune system weakens with increasing age as compared to females.

To extend the current study with additional findings, 9131 records were separated from previous extracted records based on the patient’s status (recovered/deceased) and analysed. The results of the present study have also significantly predicted the chances of being recovered in different gender groups and age categories. The females and 5–17 years age category have the highest chances of being recovered followed by 0–4 years age category. The prognosis of COVID-19 can be greatly increased by the genetic factors (presence of an extra X chromosome). The regression analysis in the present study depicted age-category wise variation in COVID-19 cases. The younger females and older males were at higher risk for getting infected and overall chances of recovery decreases as age increases. This may be because the adaptive immunity plays a significant role in response to viral infection and it declines after a certain age which makes us vulnerable to infections. Among males and females, adaptive immunity also varies differentially with age. During the childhood stage or infant stage, the males possess a greater level of IgA, IgM, and TRECs cell count and an equal number of CD4/CD8 ratio, CD8+ T cell and B cells. But as the age increases (after puberty/adulthood) the CD4/CD8 ratio, B cells, immunoglobins and T cell proliferation/activation enhances more in females. The results can be substantiated and explained based on the evidence that; the male adaptive immune system weakens with increasing age as compared to females.

To extend the current study with additional findings, 9131 records were separated from previous extracted records based on the patient’s status (recovered/deceased) and analysed. The results of the present study have also significantly predicted the chances of being recovered in different gender groups and age categories. The females and 5–17 years age category have the highest chances of being recovered followed by 0–4 years age category. The prognosis of COVID-19 can be greatly affected by comorbidities and greater risk of developing critical and mortal conditions among male, elderly (> 65 years) and smoking patients. A study among 1590 COVID-19 patients showed a hazard ratio of 1.79 and 2.59 among patients reporting one and two co-morbid conditions (hypertension and diabetes), respectively. NK (natural killer) cells increase greatly among females with age as compared to males and they experience a decline in lymphocytes elements during/- after menopause, but the elderly males experience a greater rapid decline in B cell count, CD4+ T cells and T cell proliferation than females. In recent studies, males were reported to have increased level of plasma ACE2 concentration and ACE2 is the receptor required for cellular entry of SARS CoV-2. The plasma ACE2 level was found to be higher among females with increased ACE2 expression observed among women with hypertension and diabetes.

### Table 3

| Categories | Coef. | p-value | O.R. | 95% C.I. for O.R. Lower | Upper |
|------------|-------|---------|------|------------------------|-------|
| Gender     | 0.576 | <0.001  | 1.779| 1.456                  | 2.175 |
| 0-4 years  | 4.214 | <0.001  | 67.694| 34.872                 | 131.059|
| 5-17 years | 4.481 | <0.001  | 88.286| 54.423                 | 143.217|
| 18-35 years| 3.224 | <0.001  | 25.121| 19.202                 | 32.865|
| 36-55 years| 1.230 | <0.001  | 3.421 | 2.631                  | 4.450 |
| Constant   | -4.297| <0.001  | 0.014| –                      | –      |

The results can be substantiated and explained based on the evidence that; the male adaptive immune system weakens with increasing age as compared to females.

### Table 4

| Age (in years) | Adjusted O.R. (95% C.I.) | Unadjusted O.R. (95% C.I.) |
|---------------|--------------------------|----------------------------|
| Male          | 0.920 (0.914-0.925)       | 0.919 (0.914-0.925)       |
| Female        | 1.765 (1.442-2.159)       | 1.877 (1.565-2.250)       |

Adjusted for age and gender.

S. In the present study, the male (65.39%) COVID-19 patients were higher than females (34.61%), and this trend remains consistent among all age categories. The odds ratio shows that the chances for getting recovered among females were high as compared to males independent of age. The researchers consider higher rates of smoking, lower hand washing rates, prior respiratory conditions, biological difference between sexes as a driving force for higher infection and mortality among males. The current study suggests higher chances of infection among females with lower age and males with higher age.

Hormonal response elements like putative androgen response elements (AREs) and oestrogen response elements (ORE) produces several innate immunity responses through a genetic mechanism which results in dimorphic innate immunity. Several studies emphasized the higher susceptibility of males to viral infection and produce lower antibodies than women. The higher level of TRL7 (Toll-like receptor 7 - protein sensor of RNA viruses) among women produces high interferon-α which provides higher innate immunity to women. In other diseases like cancer and HIV, the women also show greater innate immunity and greater response to vaccines. The child and adult females show high CD4+ T cells and CD8+ T cells and, higher CD4/CD8 ratio and proliferating T-cells compared to males. Regardless of age, the females tend to show higher antibody response, immunoglobin levels and B cells which are further enhanced by the genetic factors (presence of an extra X chromosome). The regression analysis in the present study depicted age-category wise variation in COVID-19 cases. The younger females and older males were at higher risk for getting infected and overall chances of recovery decreases as age increases. This may be because the adaptive immunity plays a significant role in response to viral infection and it declines after a certain age which makes us vulnerable to infections. Among males and females, adaptive immunity also varies differentially with age. During the childhood stage or infant stage, the males possess a greater level of IgA, IgM, and TRECs cell count and an equal number of CD4/CD8 ratio, CD8+ T cell and B cells. But as the age increases (after puberty/adulthood) the CD4/CD8 ratio, B cells, immunoglobins and T cell proliferation/activation enhances more in females. The results can be substantiated and explained based on the evidence that; the male adaptive immune system weakens with increasing age as compared to females.

To extend the current study with additional findings, 9131 records were separated from previous extracted records based on the patient’s status (recovered/deceased) and analysed. The results of the present study have also significantly predicted the chances of being recovered in different gender groups and age categories. The females and 5–17 years age category have the highest chances of being recovered followed by 0–4 years age category. The prognosis of COVID-19 can be greatly affected by comorbidities and greater risk of developing critical and mortal conditions among male, elderly (> 65 years) and smoking patients. A study among 1590 COVID-19 patients showed a hazard ratio of 1.79 and 2.59 among patients reporting one and two co-morbid conditions (hypertension and diabetes), respectively. NK (natural killer) cells increase greatly among females with age as compared to males and they experience a decline in lymphocytes elements during/- after menopause, but the elderly males experience a greater rapid decline in B cell count, CD4+ T cells and T cell proliferation than females. In recent studies, males were reported to have increased level of plasma ACE2 concentration and ACE2 is the receptor required for cellular entry of SARS CoV-2. The plasma ACE2 level was found to be higher among females with increased ACE2 expression observed among women with hypertension and diabetes.
highly correlated with immune signatures in lungs of males and older person and less correlated among females and younger person. So, the chances of infection and death due to COVID-19 was higher among elderly males. This also has been corroborated by the results of the present study, where the chances of being COVID-19 infected were observed to be higher for males with increasing age and chances of recovery increases with decrease in age and for females. After adjustment for age, the chances of recovery from COVID-19 among females slightly decreases, whereas it increases for males. Another perspective for higher infection among males of higher ages was X and Y chromosomes-based variations. Both these chromosomes harbours gene which involved in secreting the immune response elements but in females mosaic form of X chromosomes results into heterogenic ACE2 allele. Thus, efficient form of ACE2 receptor is present only in half of all cells which limit the infection/attachment of SARS-CoV-2 virus and provides relatively greater protection to females.

6. Conclusion

The chances of getting infected with SARS-CoV-2 varies with age-categories and gender. The females of lower age categories (<35 years) have higher chances of getting infected than male counterparts and as age increases (>35 years) the infection rate for male increases. The chances of getting recovered declines with increasing age and it is lower among males as compared to females. The variation in infection rates and chances of recovery and deaths across age and gender might be explained on the basis of biological variations like immunological and genetic differences, which however needs further verifications, considering both social and biological factors separately. The public health policies and new treatment strategies can be developed considering the present findings.

6.1. Strength and limitations

The present study provides preliminary estimates of the association of age and gender with COVID-19 cases in India which were rarely available. Evidence to correlate the role of the various immunological and genetic basis of the observed variations is offered by the present study for pursuing future research. The data used for the current study was from a private source, nevertheless, these data were also used by a government institute for the development of COVID-19 prediction portal (http://covid-tracker.imiv.ac.in:3939/covid/) therefore, the validity of using the aforementioned data persists adequately.

Ethical approval

The authors have used secondary data source for developing the present article therefore, the ethical approval was not required.

Funding source

None.

Acknowledgements

We would like to acknowledge all the associated persons involved in updating https://api.covid19india.org/documentation/csv/web portal.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cegh.2021.100788.
27 Hirokawa K, Utsuyama M, Hayashi Y, Kitagawa M, Makinodan T, Fulop T. Slower immune system aging in women versus men in the Japanese population. *Immun Ageing*. 2013;10. https://doi.org/10.1186/1742-4933-10-19.

28 Osama T, Pankhania B, Majed A. Protecting older people from COVID-19: should the United Kingdom start at age 60? *J R Soc Med*. 2020. https://doi.org/10.1177/014076820921107.

29 Li MY, Li L, Zhang Y, Wang XS. Expression of the SARS-CoV-2 cell receptor gene ACE2 in a wide variety of human tissues. *Infect. Dis. Poverty*. 2020;9:45. https://doi.org/10.1186/s40249-020-00662-x. PubMed PMID: 32345362.

30 Gemmati D, Bramanti B, Serino ML, Secchiero P, Zauli G, Tisato V. COVID-19 and individual genetic susceptibility/receptivity: role of ACE1/ACE2 genes, immunity, inflammation and coagulation. Might the double x-chromosome in females be protective against SARS-COV-2 compared to the single x-chromosome in males? *Int J Mol Sci*. 2020;21:3474. https://doi.org/10.3390/ijms21103474. PubMed PMID: 32423094.