

Gender Disparity in Risk Factors of COVID-19 Mortality Rate

Uma Ranjan (✉ usranjan@gmail.com)

Indian Institute of Science <https://orcid.org/0000-0001-6258-4513>

Anush Kini

PES University

Harish P B

PES University

Monica Anand

M S Ramaiah Institute of Technology

Brief Communication

Keywords: COVID-19, Gender Disparity, Mortality Rates, Risk Factors

Posted Date: May 3rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-101850/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Gender Disparity in Risk Factors of COVID-19 Mortality Rates

Abstract

Incidence and mortality rates due to COVID-19 have varied widely in different parts of the world and placed a huge strain on hospital resources. Understanding the underlying reasons behind such variation is crucial to developing population-specific or even individual-specific management strategies. This paper presents a comprehensive analysis of incidence and mortality rates from data collected over a cumulative period of approximately 6.5 months from February to August 2020 across 411 districts of India, totalling over 2 million individuals. We identify the health factors which have both positive as well as negative correlates with high mortality rates, using data obtained from district-wise aggregated COVID-19 incidence and mortality rates and health data obtained from National Family Health Survey (NFHS).

To obtain robust indicators, we apply both machine learning techniques as well as classical statistical methods and show that the same factors are identified by both methods. We also identify positive and negative correlates at multiple population scales by dividing the cohort into sub-cohorts formed from two Indian states which were further segregated by gender.

We show that there is a disparity of risk factors among males and females. While obesity is the highest risk factor for men, anaemia is the highest risk factor for women.

Hence, to better manage the health of a specific group of people, it is important to consider gender-wise heterogeneity in health risk factors which could contribute to differing vulnerabilities.

Keywords: COVID-19, Gender Disparity, Mortality Rates, Risk Factors.

1 Introduction

Coronavirus disease 2019 (COVID-19) has been a subject of intense study since its outbreak in December 2019. Incidence and mortality rates of the disease have varied widely, both across regions and time-periods. Understanding the underlying risks, especially of mortality due to the disease could help in developing better management strategies at population and individual levels. The rapid spread of the disease has put enormous stress on the healthcare system, including hospitals and health workers and has precluded the

25 gathering of detailed statistics of affected individuals. This is seen in the pattern of data collection in countries
26 like India where details such as longitudinal follow-up of affected individuals and extensive contact tracing
27 were carried out from March to May by most of the states [1]. However, as the number of cases began to
28 spike, this was stopped. The advisory guidelines released by the Indian Council of Medical Research (ICMR)
29 on 23rd of June, 2020 [2] touch on the need to increase the outreach of testing by deploying rapid antigen
30 Point-of-Care (PoC) tests in all containment zones and hospitals while maintaining the previous methods of
31 testing. This indicates a shift of focus from contact tracing towards mass testing coupled with follow-ups on
32 high-risk contact cases.

33 Population risk factors for COVID-19 have so far been studied from small cohorts or case studies un-
34 der hospitalization conditions or isolation centres where parameters can be collected and monitored in a
35 controlled environment [3, 4, 5, 6].

36 Due to India's large and diverse population, it has been the centre for several prediction and containment
37 strategies related to COVID-19. As of October 1, India has a cumulative total of 6,388,010 positive cases
38 and a death toll of 99,720. India has been one of the few countries which had enforced early lockdown and
39 large scale restrictions. Given the geographic spread and the number of affected individuals, this is one of
40 the largest exercises in the world, impacting the most number of people. Hence, it is important to understand
41 the risk factors underlying mortality from COVID-19, especially in vulnerable populations, which can help
42 plan effective distribution of healthcare resources.

43 Prior studies on risk factors associated with COVID-19 have considered the impact on mortality rates of
44 a single risk factor such as obesity [7, 8, 9, 10] or nutrition [11, 12]. In a first gender-wise study, Nakeshbandi
45 et al. [13] showed with data collected from hospitalized patients over a month, that obesity is associated with
46 higher mortality in men but not in women.

47 In India, maternal and child health is an area of concern. The government has focused efforts on mon-
48 itoring and providing interventions for women in the age group of 15-49 and children. Gender-wise data of
49 both men and women in this age group has been collected periodically by the National Family Health Survey
50 (NFHS) [14] on several health parameters. Health data aggregated over districts are published under the
51 stewardship of the Ministry of Health and Family Welfare (MoHFW), coordinated by the International Institute
52 for Population Sciences (IIPS), Mumbai, and implemented by a group of survey organizations and Population
53 Research Centres, following a rigorous selection procedure.

54 In this observational study, the impact of the health parameters as obtained from the NFHS data on the
55 mortality rate of COVID-19 in the reproductive age group of 15-49 at different population levels is studied.
56 The analysis of positive and negative correlates is performed on data spanning 411 districts across 23 states
57 and 3 union territories. We also repeat the analysis on a sub-cohort of two Indian states to analyze possible
58 heterogeneity of risk factors in different states. We perform aggregated as well as gender-wise analysis
59 on these two states and show that there are disparities in the positive and negative correlates of COVID-19
60 among men and women. We also discuss how these compare with gender-related risk and protective factors
61 in other diseases.

62 To our knowledge, this is the first study that considers both positive and negative correlates of mortality
63 rates, which could correspond to the risk and protective factors. This is also the first study that considers a
64 population at multiple levels of aggregation. Moreover, this is also the first time such an analysis has been
65 carried out on a large aggregate, totalling about 2,350,000 individuals.

66 **2 Methods**

67 The data processing pipeline for this study consists of the collection, cleaning, and analysis. Data extraction
68 from multiple sources constituted a significant portion of the effort. The Scikit-learn [15] and SciPy[16]
69 module in Python was used for the analysis.

70 **2.1 Data Sources**

71 Data related to incidence and mortality of COVID-19 was collected from multiple sources since no single
72 published source of such data is available. The data collected from the period of January 30 to August 18
73 include those put out by the government agencies, crowd-sourced data and daily media bulletins. Gender-
74 wise data related to COVID-19 incidence and mortality is not published by the Ministry of Health and Family
75 Welfare, Government of India. However, some state governments issue official daily bulletins through formal
76 releases to the media. Data from these media bulletins were extracted for the period of April 15 to August
77 18, 2020 for the state of Karnataka and for the period of May 1 to August 18, 2020 for the state of Tamil
78 Nadu. A sample data of such a bulletin can be found in [17, 18]. The sources from which data for this study
79 was sourced include:

- 80 • The Open Government Data Platform India website [19] contains data officially published by the Gov-
81 ernment of India. Serial follow up of people who tested positive for COVID-19 was done and details
82 of age, sex and status (hospitalized, discharged, tested negative) was captured for a few selected
83 cities[20, 21, 22, 23, 24, 25]. This site was last updated on July 30, 2020 and contains details of a total
84 of 16,981 cases and 683 deaths.

- 85 • COVID-19 India [26] is a crowd-sourced initiative to extract and update data related to COVID-19 cases
86 and deaths across India. The data collected here contains district-wise numbers of confirmed cases,
87 active cases, recovered cases and deceased cases for 800 districts across India starting from 30th of
88 January, 2020. We have used the data available until the 18th of August, 2020.

- 89 • Media bulletins issued by the states of Karnataka and Tamil Nadu : Details of incidence in the state
90 of Karnataka were curated by Siva Athreya et al. [27]. This data contains details of 71,068 positive
91 cases extracted from the Karnataka government media bulletins. Additionally, district-wise numbers
92 for 3763 deaths in Karnataka were collected by manually going through the official media bulletins.
93 Similarly, the details of 20,731 positive cases and 5161 deaths were extracted for the state of Tamil
94 Nadu. Of these, the data related to the age group 15-49 was extracted and resulted in 47,213 cases,
95 756 deaths from Karnataka and 14,174 cases, 677 deaths from Tamil Nadu. Gender information for
96 positive cases was available until 31st of May, 2020 for Tamil Nadu media bulletins and 20th of July,
97 2020 for Karnataka media bulletins. Thereafter, only total infected cases related to COVID-19 cases
98 were published, and gender-wise segregation was not available.

- 99 • National Family Health Survey Data : The National Family Health Survey (NFHS) is a large-scale,
100 multi-round survey conducted in a representative sample of households throughout India. In 2014-
101 2015, India implemented the fourth National Family Health Survey (NFHS-4). NFHS-4 had a sample
102 size of approximately 568,200 households and a total sample of 625,014 women and 93,065 men
103 eligible for the interview. The NFHS [14, 28] data provides updates and evidence of trends in key
104 population, health and nutrition indicators. The data spans across 29 states and 6 union territories, a
105 total of 640 districts across India. For each of these districts, 93 key indicators are recorded ranging
106 from various issues during pregnancy, nutrition, population, literacy and more.

107 **2.2 Data Curation**

108 Details of total confirmed cases (which included the number of active, recovered, and deceased cases)
109 and the number of deaths were available for 800 districts. Of these, 535 districts overlapped with the 640
110 districts for which NFHS data was available. We considered the subset of these districts which had at least
111 5 deaths, resulting in a total of 411 districts, spanning across 23 states and 3 union territories of India on
112 which final analysis was done. These 411 districts represent a geographical extent of more than 3 million
113 square kilometres and a cumulative total of 2,331,363 cases and 46,239 deaths.

114 In the sub-cohort of the states of Karnataka and Tamil Nadu, gender-wise numbers of COVID-19 positive
115 cases was available only until 20th of July and 31st of May, 2020 respectively. Thereafter, only total positive
116 cases related to COVID-19 cases were published, and gender-wise segregation was not available. We
117 compute the gender-wise fraction in the age group 15-49 from this initial data, estimate the gender-wise
118 numbers for each district for later periods using this fraction and the total number of positive cases published
119 in the bulletins as of August 18, 2020.

120 Among the 93 key indicators included in the NFHS data, we selected the ones corresponding to adult
121 health indicators and further limited it to factors discussed in case reports and medical opinions in literature.
122 We also took into account chronic conditions that are unlikely to have changed in the period between the
123 collection of the health data and the COVID-19 pandemic. The broad categories of factors considered were
124 those related to low BMI, obesity, anaemia, blood pressure and diabetes. The set of health factors chosen
125 for analysis are enumerated in Table 1.

126 The health factors extracted from the NFHS data are published gender-wise. To obtain the gender-
127 desegregated population value, the weighted average of each factor was computed according to the sex ratio
128 of each district. Through this process, aggregated health factors corresponding to the gender-desegregated
129 population of each district was obtained.

130 **2.3 Analysis**

131 A Lasso regression[29] of the health factors on mortality rates calculated from the COVID-19 India data
132 was conducted on districts that reported at least 5 deaths. The health data was standardized before the
133 regression. For each Lasso test, the λ (regularisation parameter) with the best R^2 value was taken through

134 a search of the results from the Lasso_path function. Residual plots corresponding to this value of λ were
135 inspected visually to ensure there was no bias.

136 Independently, factors which differed significantly between the districts with high mortality rates and those
137 with low mortality rates were identified via the Mann-Whitney U test with a significance level of 0.05 corrected
138 by the Bonferroni criterion for each health factor. The districts were classified into two categories of low and
139 high mortality, depending on whether they fell below or above the second quartile in mortality rates. The
140 effect size was also calculated for all the factors between the two sub-groups using Cohen's d effect size and
141 interpreted according to the thresholds defined in [30] i.e. $|d| \leq 0.2$ is 'negligible' effect size, $0.2 < |d| \leq 0.5$ is
142 'small', $0.5 < |d| \leq 0.8$ is 'medium' and otherwise 'large'.

143 The factors obtained from the Mann-Whitney U test and the Lasso test were compared and common
144 factors identified as the risk and protective factors of the population. The same procedure was also followed
145 in the sub-cohort of the two states of Karnataka and Tamil Nadu.

146 Further, each state was analyzed independently and positive and negative correlative factors were iden-
147 tified. For the gender-wise correlates using the Mann-Whitney U test, the data of each gender across the
148 two states were combined to yield significant numbers.

149 **3 Results**

150 A preliminary two-tailed test performed on the Open Government Data, results of which presented in Table
151 2, shows that there is a significant difference between male and female mortality rates in some regions while
152 other regions do not show a significant difference. Further, though female mortality rates are higher than
153 male mortality rates in many regions, it is not possible to infer that this is uniformly true for all regions.

154 **3.1 Aggregated population across the country**

155 The distribution of the aggregated health factors in the 411 districts is presented in Figure 1. District-wise
156 counts of 2,331,363 COVID-19 cases and 46,239 deaths are used to draw correlations between the health
157 factors and mortality from these districts.

158 The Lasso plot on this data (Figure 2) indicates that the most positive correlates are overweight or obesity
159 and BMI below normal. Anaemia, high blood sugar level and very high blood pressure show a negative

160 correlation with mortality, with anaemia having the highest negative correlation at higher values of λ .

161 The results of the Mann-Whitney U test on the same data between the two categories of districts of
162 high and low values of mortality rates are shown in Table 3. Anaemia and obesity emerged as statistically
163 significant with small effect sizes.

164 By considering the factors that were identified by both the tests, we conclude that obesity (DM=-3.2,
165 95%CI (-4.8,-1.7), P<0.0001, ES=0.2026) is positively correlated and anaemia (DM=4.00, 95%CI (2.3,5.6),
166 P<0.0001, ES=0.2306) is negatively correlated with high mortality on the cohort of nation-wide population.

167 **3.2 Sub-cohort of two Indian states**

168 A sub-cohort of 29 districts, consisting of 16 districts from the state of Karnataka and 13 districts from the
169 state of Tamil Nadu was analysed. Using available case data from the bulletins for the age group of 15-49,
170 we estimated the number of cases for 240,912 cases from Karnataka and 298,046 cases from Tamil Nadu.
171 Only those districts that had reported at least 5 deaths for each gender in the age group 15-49 from the
172 death data was considered for this analysis. This resulted in a total of 667 deaths from Karnataka and 556
173 deaths from Tamil Nadu. The distribution for the aggregated health factors in these districts can be seen in
174 Figure 3.

175 From the Lasso plot in Figure 4, BMI below normal, high blood sugar level, very high blood pressure are
176 the positive correlates while obesity and anaemia are the negative correlates.

177 The results of the Mann-Whitney U test on the same data between the two categories of districts of high
178 and low values of mortality rates are shown in Table 4. Obesity, BMI below normal and anaemia were found
179 to be statistically significant with medium effect sizes.

180 By considering the factors that were identified by both the tests, it is seen that *BMI below normal*
181 (DM=-6.4, 95%CI (-9.9,-2.4), P=0.013, ES=0.5836) is positively correlated while *Obesity* (DM=7.4, 95%CI
182 (2.7,11.6), P=0.013, ES=0.5836) and *Anaemia* (DM=5.4, 95%CI (1.8,8.8), P=0.02, ES=0.5570) are nega-
183 tively correlated with high mortality in the sub-cohort.

184 **3.3 Gender-wise analysis of the sub-cohort**

185 A gender-segregated analysis was performed on the districts of Karnataka and Tamil Nadu to determine
186 the association between gender-wise mortality and health factors from these two states. The gender-wise

187 distribution of each district was estimated for 240,912 cases from Karnataka and 298,046 cases from Tamil
188 Nadu using the initial case data. From the data on deaths in Karnataka and Tamil Nadu, we considered
189 only districts with at least 5 deaths separately for each gender. Karnataka had 224 female deaths across 16
190 districts and 496 male deaths across 23 districts. Tamil Nadu data had 186 female deaths across 13 districts
191 and 434 male deaths across 21 districts.

192 The distribution of the gender-wise health factors across districts of Karnataka and Tamil Nadu is pre-
193 sented in Figure 5. The results of the gender-segregated Lasso tests on Karnataka and Tamil Nadu can be
194 seen in Figure 6 and Figure 7 respectively. In the case of males, overweight or obesity is positively correlated
195 and anaemia is negatively correlated with the high mortality rate in both the states. In the case of females,
196 anaemia and high blood sugar level show a positive correlation while obesity shows a strong negative corre-
197 lation, and very high blood pressure a weak negative correlation. In order to ascertain that the correlations
198 were not due to multi-collinearity effects, the slope from the linear fit of scatter plots was examined for each
199 factor separately.

200 The Mann-Whitney U test was performed by combining the data of both the states to improve the sample
201 size. 44 districts for male and 29 districts for female were considered for analysis based on the condition that
202 the number of deaths for the gender under consideration was at least 5. From the results presented in Table
203 5 for females and Table 6 for males, none of the factors showed statistical significance. However, the factors
204 of BMI below normal, obesity and anaemia for women and anaemia for men had non-negligible effect sizes.

205 By considering the factors that were identified by both the tests, it is seen that *Anaemia* (DM=4.3, 95%CI
206 (0.3,8.5), P=0.2275, ES=0.3891) is positively correlated and *Obesity* (DM=6.05, 95%CI (0.2,11.5), P=0.163,
207 ES=0.4156) is negatively correlated with high mortality in females. Conversely, *Anaemia* (DM=2.4, 95%CI
208 (-1.0,5.6), P=0.6495, ES=0.2327) is negatively correlated with high mortality in males.

209 **3.4 Summary of Results**

210 The factors identified by each test and the common factors at each level of population aggregation is shown
211 in Figure 8.

212 **4 Conclusion**

213 In this paper, we have reported the first large scale analysis of health factors correlated positively as well
214 as negatively with mortality rates of COVID-19 in India. The factors affecting mortality rates show a distinct
215 difference among males and females in the age group of 15-49. Among males, obesity has been shown to
216 be a risk factor in both regression as well as statistical analysis. Other studies investigating obesity have
217 also concluded that there is difference in the impact of obesity in males and females on COVID-19 mortality
218 [13]. Pre-menopausal obesity has been shown to be associated with a lower risk of breast cancer [31, 32],
219 thus suggesting that fat distribution could play a significant role in health conditions [33] and pre-menopausal
220 obesity could serve as a protective factor.

221 We conclude through this study that anaemia is positively correlated with COVID-19-related mortality
222 in women, but negatively correlated with mortality in men. The difference in the role played by anaemia
223 in women vs. men can be explained by the fact that anaemia, when present in men, is mild or moderate,
224 whereas the prevalence of severe anaemia is higher in women [34]. It is also evident from the gender-
225 wise distribution in Figure 5 that anaemia is twice as prevalent in females as in males. Recent estimates
226 of iron-deficiency anaemia (IDA) show that 52% of women aged 15-49 are anaemic [35]. This difference in
227 prevalence could be significantly higher during menstruation and pregnancies. Severe anaemia has been
228 associated with higher maternal mortality [36]. Severe anaemia has also been associated with higher rates
229 of ICU admission in COVID-19 [37, 38, 39]. This first study of the effect of anaemia in COVID-19 suggests
230 that haemodilution could play a role in COVID-19 mortality.

231 The findings indicate that risk factors for COVID-19 mortality are by themselves heterogeneous, and their
232 effects need to be investigated in conjunction with gender, menopausal status, and severity of the condition
233 to understand them better.

234 **Competing Interests Statement**

235 The authors declare no competing interests.

236 **Financial disclosures**

The authors received no financial support for the research, authorship, and/or publication of this article.

References

- [1] Bhatia R, Abraham P. Lessons learnt during the first 100 days of COVID-19 pandemic in India *Indian Journal of Medical Research*. 2020;151:387.
- [2] Indian Council of Medical Research . Newer Additional Strategies for COVID-19 Testing https://www.icmr.gov.in/pdf/covid/strategy/New_additional_Advisory_23062020_3.pdf 2020.
- [3] NCIRD . Evidence used to update the list of underlying medical conditions that increase a person's risk of severe illness from COVID-19 <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/evidence-table.html> 2020.
- [4] Chang MC, Park YK, Kim BO, Park D. Risk factors for disease progression in COVID-19 patients *BMC Infectious Diseases*. 2020;20:1–6.
- [5] Lee S, Kim T, Lee E, et al. Clinical Course and Molecular Viral Shedding Among Asymptomatic and Symptomatic Patients With SARS-CoV-2 Infection in a Community Treatment Center in the Republic of Korea *JAMA Internal Medicine*. 2020.
- [6] Liang W, Liang H, Ou L, et al. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19 *JAMA Internal Medicine*. 2020;180:1081-1089.
- [7] Stefan N, Birkenfeld AL, Schulze MB, Ludwig DS. Obesity and impaired metabolic health in patients with COVID-19 *Nature Reviews Endocrinology*. 2020:1–2.
- [8] Curtin KM, Pawloski LR, Mitchell P, Dunbar J. COVID-19 and Morbid Obesity: Associations and Consequences for Policy and Practice *World Medical & Health Policy*. 2020.
- [9] Simonnet A, Chetboun M, Poissy J, et al. High Prevalence of Obesity in Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) Requiring Invasive Mechanical Ventilation *Obesity*. 2020;28:1195-1199.

- [10] Chaudhry R, Dranitsaris G, Mubashir T, Bartoszko J, Riazi S. A country level analysis measuring the impact of government actions, country preparedness and socioeconomic factors on COVID-19 mortality and related health outcomes *EClinicalMedicine*. 2020;25:100464.
- [11] Zhao X, Li Y, Ge Y, et al. Evaluation of Nutrition Risk and Its Association With Mortality Risk in Severely and Critically Ill COVID-19 Patients *Journal of Parenteral and Enteral Nutrition*. 2020.
- [12] Hiedra R, Lo KB, Mohammad E, et al. The Use of IV vitamin C for patients with COVID-19: a single center observational study *Expert Review of Anti-Infective Therapy*. 2020.
- [13] Nakeshbandi M, Maini R, Daniel P, et al. The impact of obesity on COVID-19 complications: a retrospective cohort study *International Journal of Obesity*. 2020;44:1832–1837.
- [14] The Ministry of Health and Family Welfare, India . The National Family Health Survey (NFHS) <http://rchiips.org/nfhs/> 2015-2016.
- [15] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python *Journal of Machine Learning Research*. 2011;12:2825–2830.
- [16] Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python *Nature Methods*. 2020;17:261–272.
- [17] Health And Family Welfare Department, Government of Karnataka . Health Department Bulletin https://covid19.karnataka.gov.in/govt_bulletin/en 2020.
- [18] Health & Family Welfare Department Government of Tamil Nadu . Daily Bulletin <https://stopcorona.tn.gov.in/daily-bulletin/> 2020.
- [19] Government of India . Open Government Data Platform India <https://data.gov.in/> 2020.
- [20] Government of India . COVID-19 Details: : Surat <https://smartcities.data.gov.in/catalog/covid-19-details-surat> 2020.
- [21] Government of India . COVID-19 Thane <https://smartcities.data.gov.in/catalog/covid-19-thane> 2020.

- [22] Government of India . Covid_19 : Jabalpur <https://smartcities.data.gov.in/catalog/covid19-jabalpur> 2020.
- [23] Government of India . COVID Rajkot <https://smartcities.data.gov.in/catalog/covid-rajkot-0> 2020.
- [24] Government of India . Covid 19 Indore <https://smartcities.data.gov.in/catalog/covid-19-indore> 2020.
- [25] Government of India . COVID-19 Varanasi <https://smartcities.data.gov.in/catalog/covid-19-varanasi> 2020.
- [26] COVID19-India Ops . crowd-sourced data <https://www.covid19india.org/> 2020.
- [27] Athreya S, Gadhiwala N, Mishra A. COVID-19 India-Timeline an understanding across States and Union Territories <https://www.isibang.ac.in/~athreya/incovid19/data.html> 2020.
- [28] Das D. NFHS-4: Opening, Structuring and Visualizing India's latest and largest National Health Dataset <http://debarghyadas.com/writes/nfhs-4/> 2017.
- [29] Tibshirani R. Regression Shrinkage and Selection Via the Lasso *Journal of the Royal Statistical Society: Series B (Methodological)*. 1996;58:267-288.
- [30] Cohen J. Statistical power analysis for the behavioral sciences, 2nd edn 1988.
- [31] Mohanty SS, Mohanty PK. Obesity as potential breast cancer risk factor for postmenopausal women *Genes & Diseases*. 2019.
- [32] The Premenopausal Breast Cancer Collaborative Group . Association of Body Mass Index and Age With Subsequent Breast Cancer Risk in Premenopausal Women *JAMA Oncology*. 2018;4:e181771-e181771.
- [33] Park JK, Lim YH, Kim KS, et al. Changes in body fat distribution through menopause increase blood pressure independently of total body fat in middle-aged women: the Korean National Health and Nutrition Examination Survey 2007–2010 *Hypertension Research*. 2013;36:444–449.

- [34] Didzun O, De Neve JW, Awasthi A, et al. Anaemia among men in India: a nationally representative cross-sectional study *The Lancet Global Health*. 2019;7:e1685–e1694.
- [35] Rammohan A, Awofeso N, Robitaille MC. Addressing female iron-deficiency anaemia in india: is vegetarianism the major obstacle? *ISRN Public Health*. 2012;2012.
- [36] Daru J, Zamora J, Fernández-Félix BM, et al. Risk of maternal mortality in women with severe anaemia during pregnancy and post partum: a multilevel analysis *The Lancet Global Health*. 2018;6:e548–e554.
- [37] Mendy A, Apewokin S, Wells AA, Morrow AL. Factors Associated with Hospitalization and Disease Severity in a Racially and Ethnically Diverse Population of COVID-19 Patients *medRxiv*. 2020.
- [38] Taneri Petek Eylül, Gomez-Ochoa Sergio Alejandro, Llanaj Erand, et al. Anemia and iron metabolism in COVID-19: A systematic review and meta-analysis *European Journal of Epidemiology*. 2020;35:763–773.
- [39] Tao Z, Xu J, Chen W, et al. Anaemia is associated with severe illness in COVID-19: a retrospective cohort study *Journal of Medical Virology*. 2020.

Table 1. Selected health factors from NFHS data.

<i>Factor (%)</i>	<i>Characteristic</i>
Men whose Body Mass Index (BMI) is below normal	BMI < 18.5 kg/m ²
Women whose Body Mass Index (BMI) is below normal	BMI < 18.5 kg/m ²
Men who are overweight or obese	BMI ≥ 25.0 kg/m ²
Women who are overweight or obese	BMI ≥ 25.0 kg/m ²
Men who are anaemic	<13.0 g/dl
Women who are anaemic	<12.0 g/dl(Non-pregnant) and <11.0 g/dl(Pregnant)
Men who have a high blood sugar level	>140 mg/dl
Women who have a high blood sugar level	>140 mg/dl
Men who have very high blood pressure	Systolic ≥180 mm of Hg and/or Diastolic ≥110 mm of Hg
Women who have very high blood pressure	Systolic ≥ 180 mm of Hg and/or Diastolic ≥ 110 mm of Hg

Table 2. Results of the hypothesis tests on gender-wise mortality rates across different regions in India.

<i>Region</i>	<i>Difference in mortality rates (%)</i>	<i>95 % CI</i>	<i>P Value</i>
Surat	0.8613	(-0.0837,1.8063)	0.362
Thane	-0.8337	(-1.2795, -0.3880)	0.0614
Jabalpur	3.6054	(1.5874,5.6234)	0.074
Rajkot	2.5332	(0.6437,4.4227)	0.18
Indore	-1.6341	(-2.1142,-1.1540)	< 0.001
Varanasi	1.0449	(-0.2144,2.3041)	0.4067
Karnataka	-0.3217	(-0.3737,-0.2698)	< 0.001
Tamil Nadu	-0.5387	(-0.5849,-0.4924)	< 0.001

Significant regions are marked in bold.

Table 3. Results of Mann-Whitney U test on aggregated health parameters across India.

<i>Aggregated Factor (%)</i>	<i>Difference in Median</i>	<i>95 % CI</i>	<i>P Value</i>	<i>Effect size</i>
Body Mass Index (BMI) below normal	1.9	(0.4 , 3.4)	0.0675	0.1221
Overweight or obese	-3.2	(-4.8 , -1.7)	<0.0001	0.2026
Anaemia	4.0	(2.3 , 5.6)	<0.0001	0.2306
High blood sugar level	0.2	(-0.2 , 0.7)	1.0	0.0472
Very high blood pressure	-0.03	(-0.1 , 0.05)	1.0	0.0299

Significant regions are marked in bold.

Table 4. Results of Mann-Whitney U test on aggregated health parameters across Karnataka and Tamil Nadu.

<i>Aggregated health Factor (%)</i>	<i>Difference in median</i>	<i>95% CI</i>	<i>P Value</i>	<i>Effect size</i>
Body Mass Index (BMI) below normal	-6.4	(-9.9 , -2.4)	0.013	0.5836
Overweight or obese	7.4	(2.7 , 11.6)	0.013	0.5836
Anaemia	5.4	(1.8 , 8.8)	0.02	0.5570
High blood sugar level	0.6	(-0.6 , 1.8)	1.0	0.1415
Very high blood pressure	0.08	(-0.4 , 0.5)	1.0	0.0530

Significant regions are marked in bold.

Table 5. Results of Mann-Whitney U test on female health parameters across Karnataka and Tamil Nadu.

<i>Female health Factor (%)</i>	<i>Difference in median</i>	<i>95% CI</i>	<i>P Value</i>	<i>Effect size</i>
Body Mass Index (BMI) below normal	-4.1	(-8.1 , 0.2)	0.2975	0.3670
Overweight or obese	6.05	(0.2 , 11.5)	0.163	0.4156
Anaemia	4.3	(0.3 , 8.5)	0.2275	0.3891
High blood sugar level	0.45	(-0.8 , 2.0)	1.0	0.1503
Very high blood pressure	0.0	(-0.4 , 0.3)	1.0	0.0354

Table 6. Results of Mann-Whitney U test on male health parameters across Karnataka and Tamil Nadu.

<i>Male health Factor (%)</i>	<i>Difference in median</i>	<i>95% CI</i>	<i>P Value</i>	<i>Effect size</i>
Body Mass Index (BMI) below normal	-1.45	(-5.5 , 2.7)	1.0	0.1271
Overweight or obese	1.0	(-3.4 , 4.3)	1.0	0.0913
Anaemia	2.4	(-1.0 , 5.6)	0.6495	0.2327
High blood sugar level	-0.35	(-1.9 , 1.4)	1.0	0.0501
Very high blood pressure	0.2	(-0.4 , 0.8)	1.0	0.1020

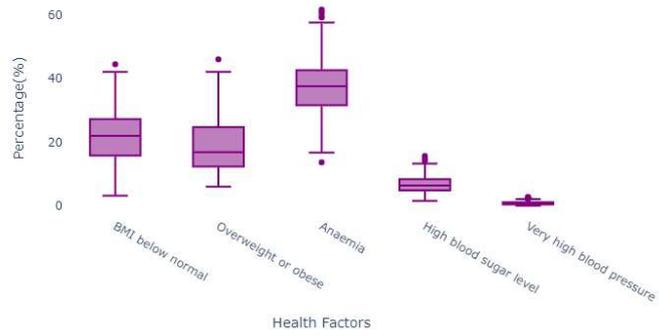


Figure 1. Distribution of aggregated health factors on pan India data.

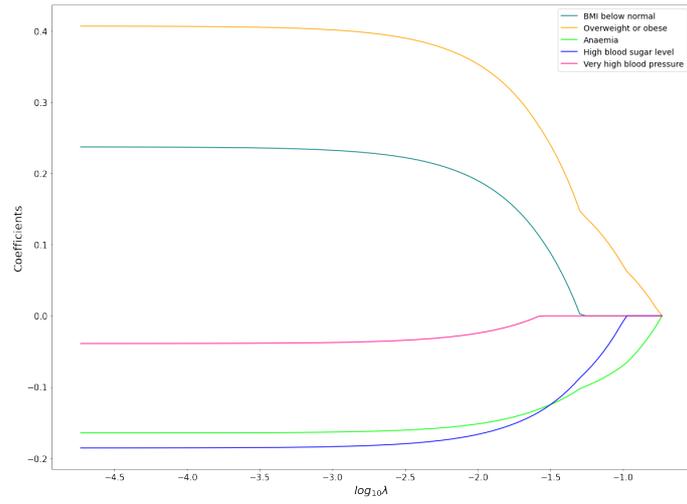


Figure 2. Lasso regression on pan India data. $R^2 = 0.1035$; Teal - BMI below normal, Orange - Overweight or obese, Green - Anaemia, Blue - High blood sugar level, Pink - Very high blood pressure.

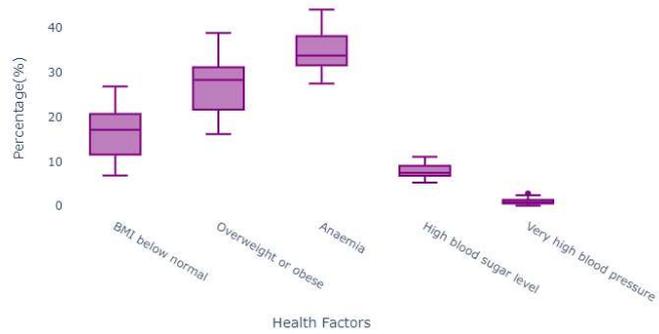


Figure 3. Distribution of aggregated health factors on sub-cohorts of Karnataka and Tamil Nadu.

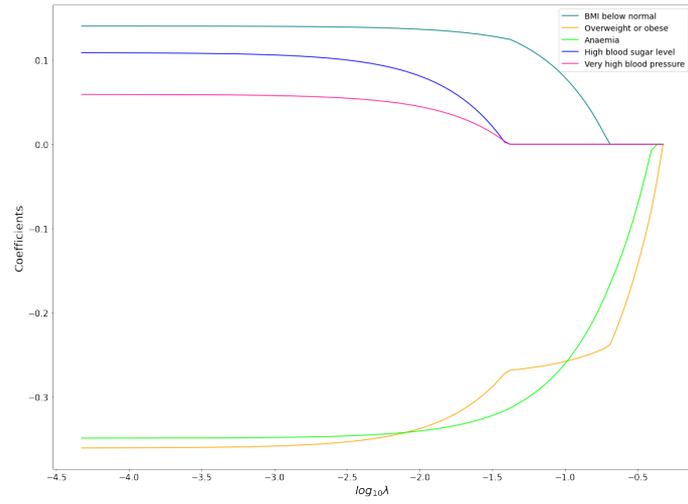


Figure 4. Lasso regression on aggregated sub-cohorts of Karnataka and Tamil Nadu. $R^2 = 0.3509$; *Teal - BMI below normal, Orange - Overweight or obese, Green - Anaemia, Blue - High blood sugar level, Pink - Very high blood pressure.*

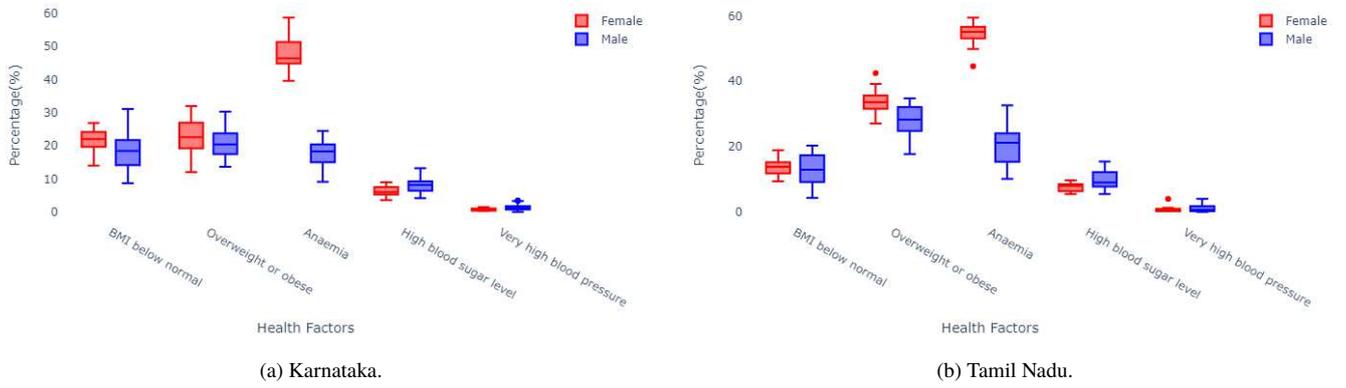
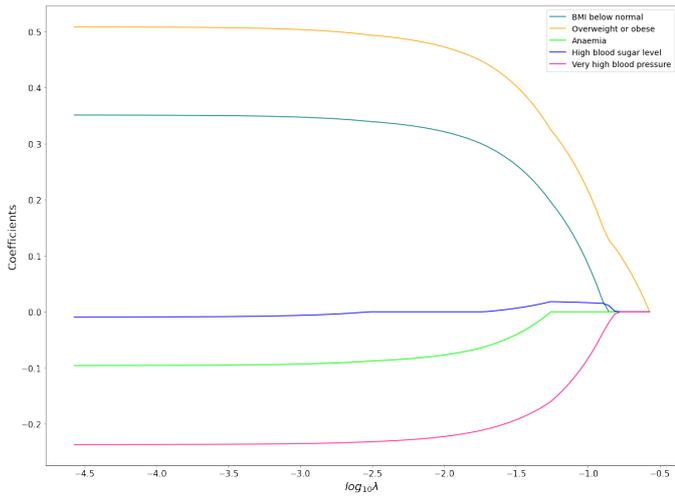
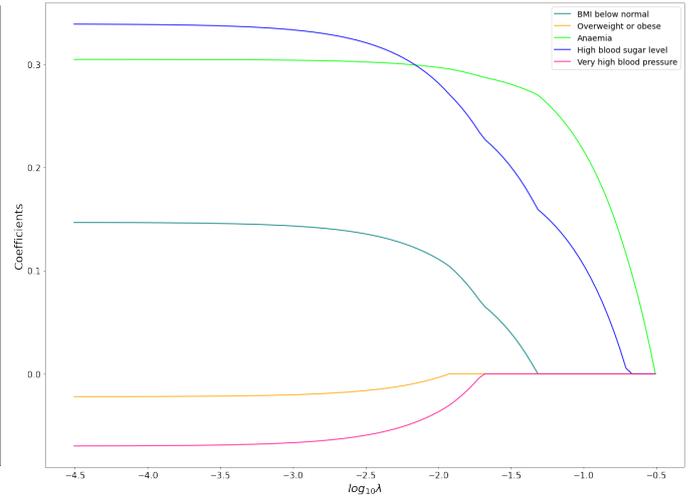


Figure 5. Distribution of gender-wise health factors on sub-cohorts of Karnataka and Tamil Nadu. *Red - Female; Blue - Male;*

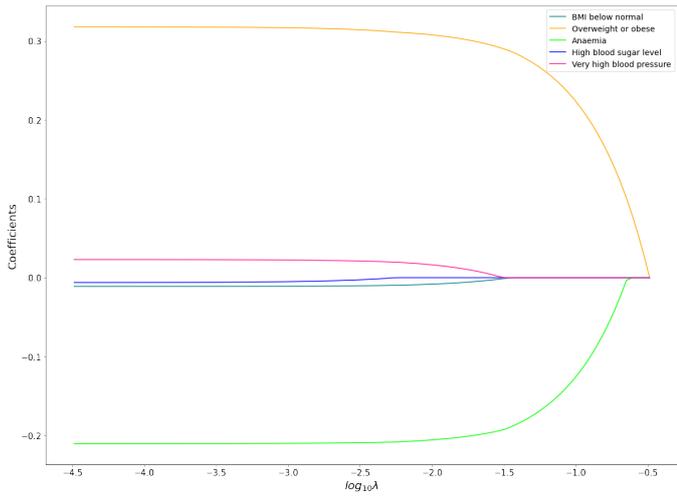


(a) sub cohort of males.

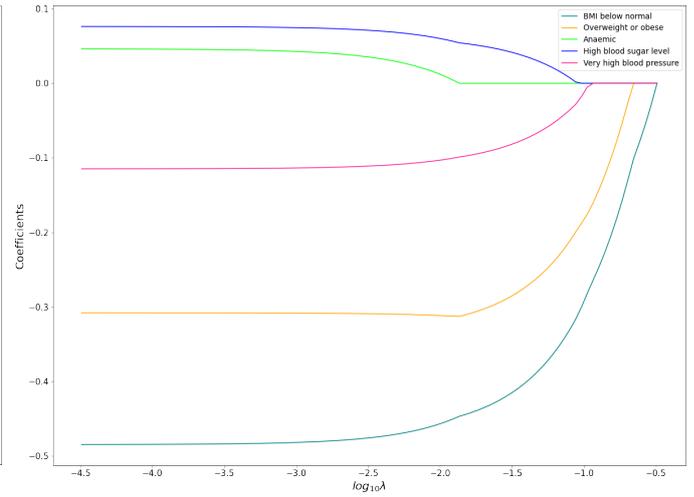


(b) sub cohort of females.

Figure 6. Lasso regression on gender-wise sub-cohorts of Karnataka. (a) $R^2 = 0.1917$ (b) $R^2 = 0.1529$; Teal - BMI below normal, Orange - Overweight or obese, Green - Anaemia, Blue - High blood sugar level, Pink - Very high blood pressure.



(a) sub cohort of males.



(b) sub cohort of females.

Figure 7. Lasso regression on gender-wise sub-cohorts of Tamil Nadu. (a) $R^2 = 0.1567$ (b) $R^2 = 0.2270$; Teal - BMI below normal, Orange - Overweight or obese, Green - Anaemia, Blue - High blood sugar level, Pink - Very high blood pressure.

	Positive factors from lasso tests	Negative factors from lasso tests	Selected factors from Mann-Whitney U test
All India	Obesity BMI below normal	Anaemia Blood sugar level high Blood pressure very high	Anaemia Obesity
Common correlates	Common positive correlates		Common negative correlates
	Obesity		Anaemia

	Positive factors from lasso tests	Negative factors from lasso tests	Selected factors from Mann-Whitney U test
Sub cohort of 2 states	BMI below normal Blood sugar level high Blood pressure very high	Obesity Anaemia	Obesity BMI below normal Anaemia
Common correlates	Common positive correlates		Common negative correlates
	BMI below normal		Obesity, Anaemia

Male

Female

	Positive factors from lasso tests	Negative factors from lasso tests	Selected factors from Mann-Whitney U test
Karnataka	Obesity BMI below normal	Blood pressure very high Anaemia	Anaemia
Tamil Nadu	Obesity Blood pressure very high	Anaemia	
Common correlates	Common positive correlates		Common negative correlates
	-		Anaemia

	Positive factors from lasso tests	Negative factors from lasso tests	Selected factors from Mann-Whitney U test
Karnataka	Anaemia Blood sugar level high BMI below normal	Blood pressure very high Obesity	Obesity Anaemia BMI below normal
Tamil Nadu	Blood sugar level high Anaemia	BMI below normal Obesity Blood pressure very high	
Common correlates	Common positive correlates		Common negative correlates
	Anaemia		Obesity

Figure 8. Summary of influencing factors across regions and gender.

Figures

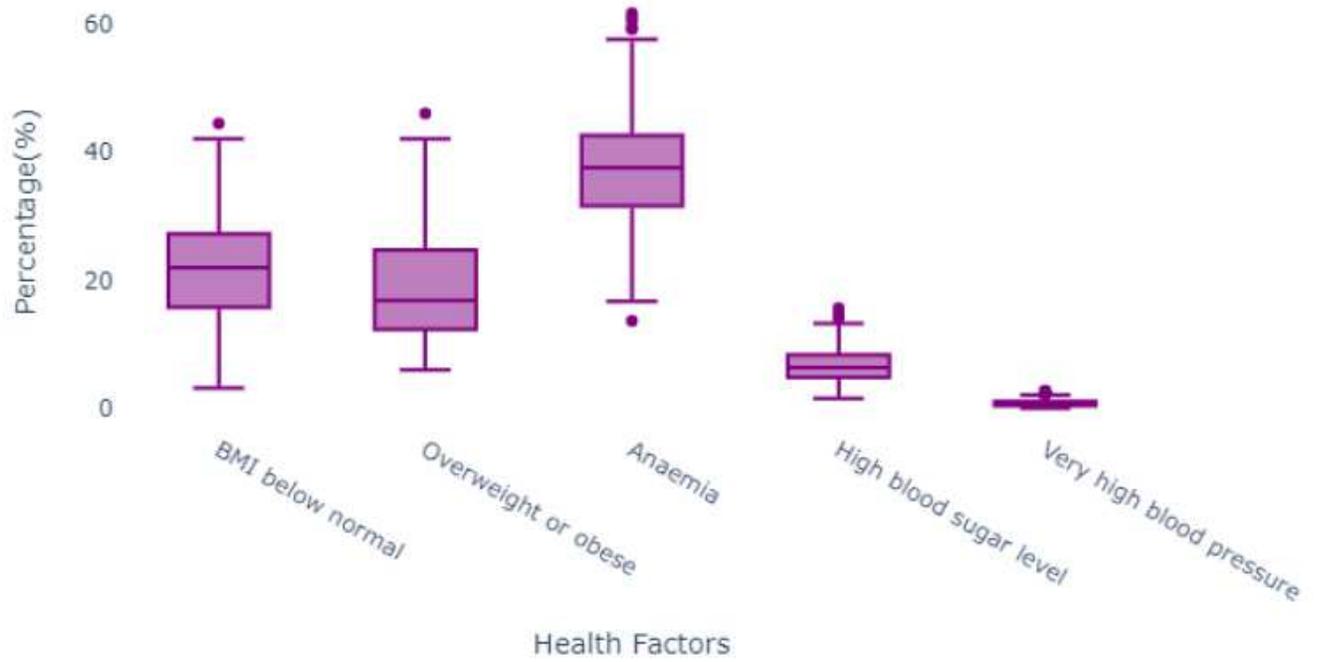


Figure 1

Distribution of aggregated health factors on pan India data.

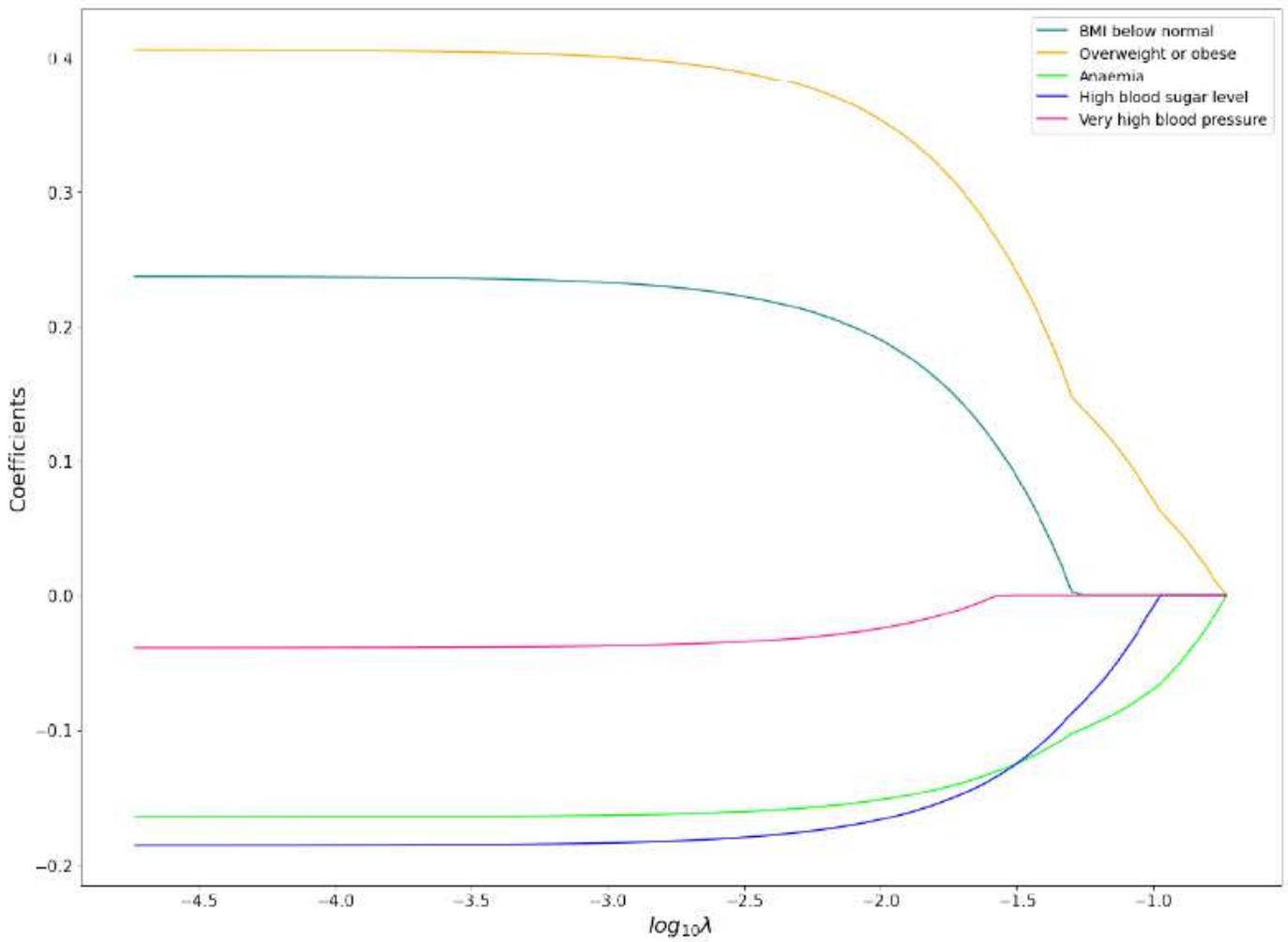


Figure 2

Lasso regression on pan India data. $R^2 = 0.1035$; Teal - BMI below normal, Orange - Overweight or obese, Green - Anaemia, Blue - High blood sugar level, Pink - Very high blood pressure.

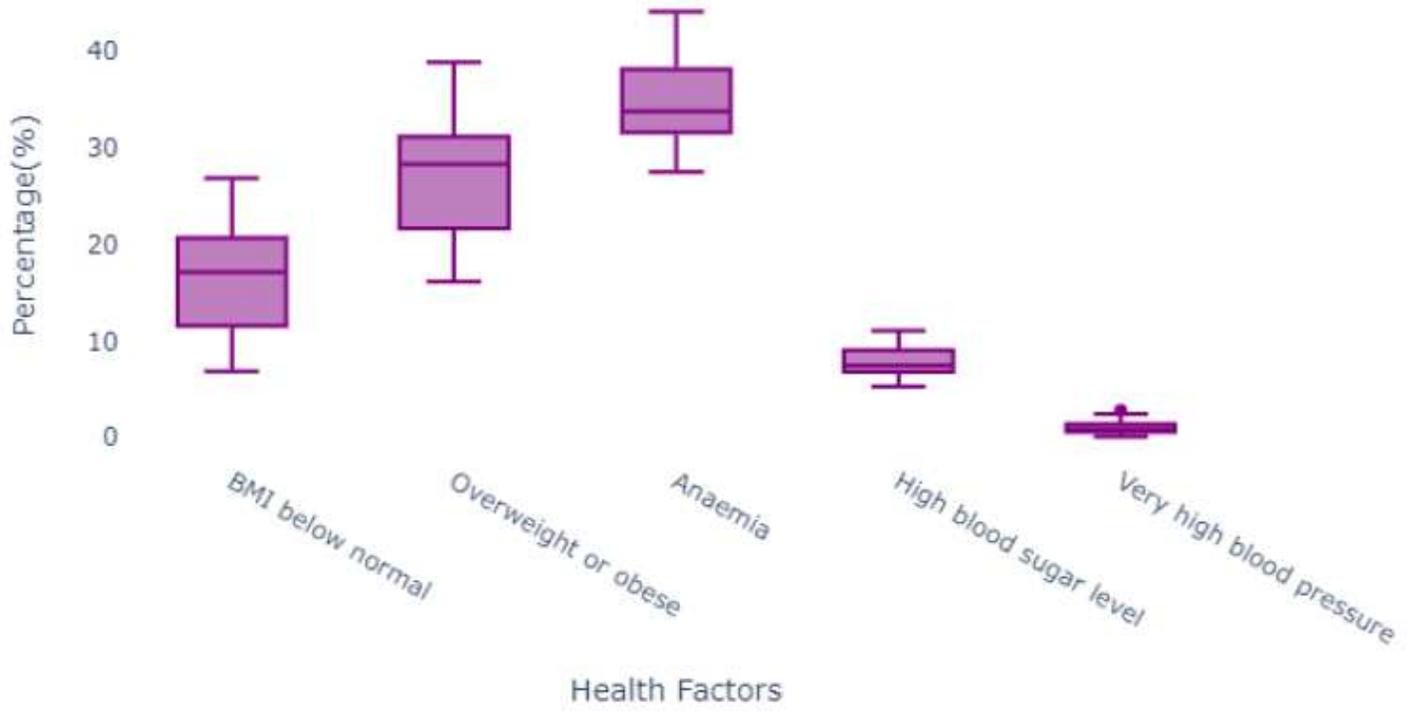


Figure 3

Distribution of aggregated health factors on sub-cohorts of Karnataka and Tamil Nadu.

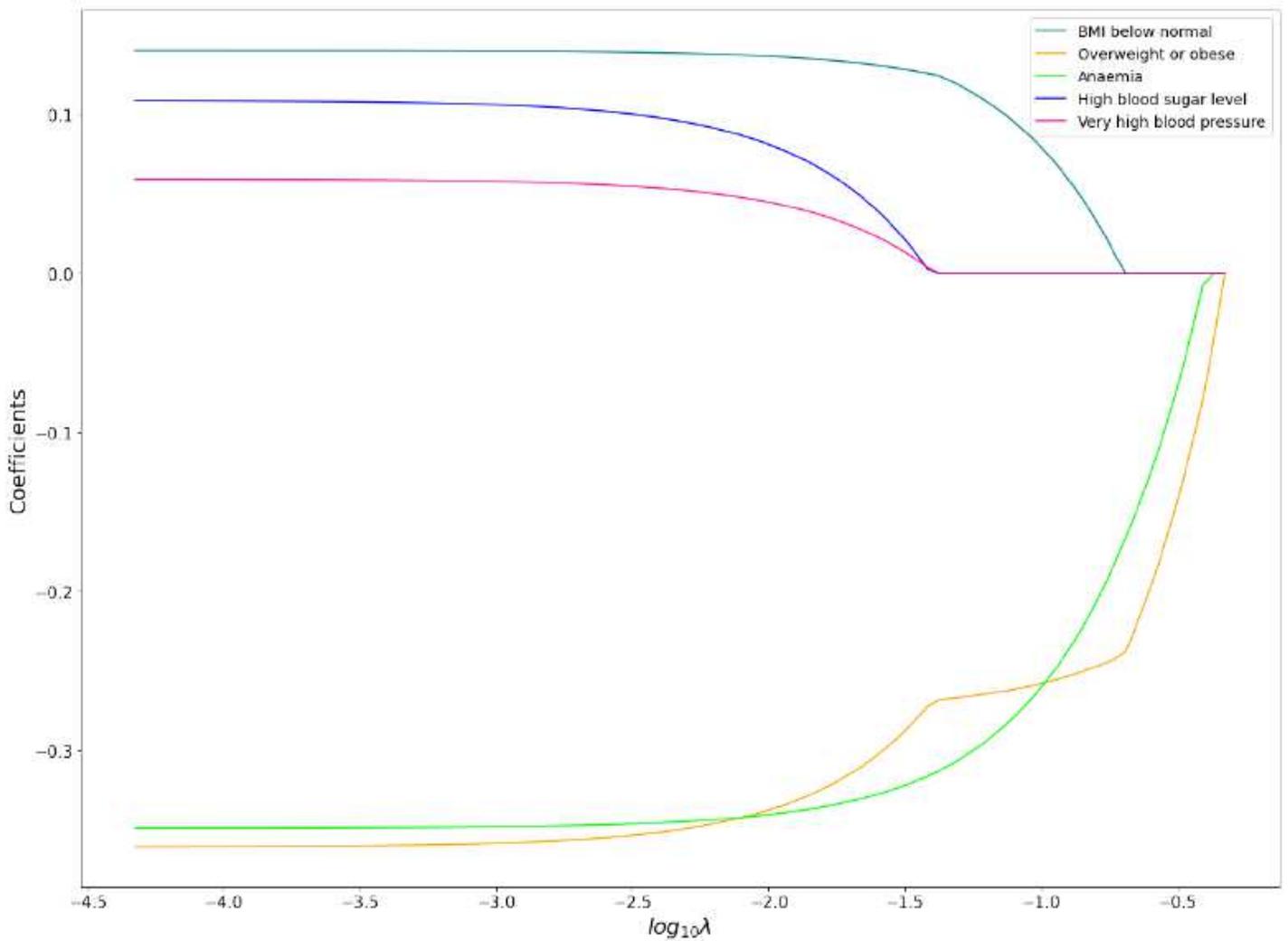


Figure 4

Lasso regression on aggregated sub-cohorts of Karnataka and Tamil Nadu. $R^2 = 0.3509$; Teal - BMI below normal, Orange - Overweight or obese, Green - Anaemia, Blue - High blood sugar level, Pink - Very high blood pressure.

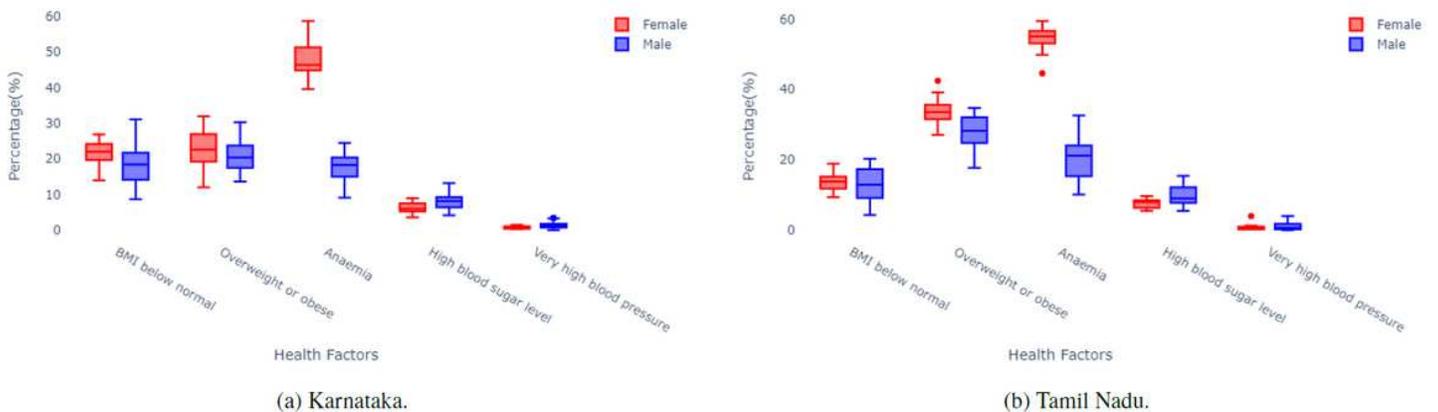


Figure 5

Distribution of gender-wise health factors on sub-cohorts of Karnataka and Tamil Nadu. Red - Female; Blue - Male;

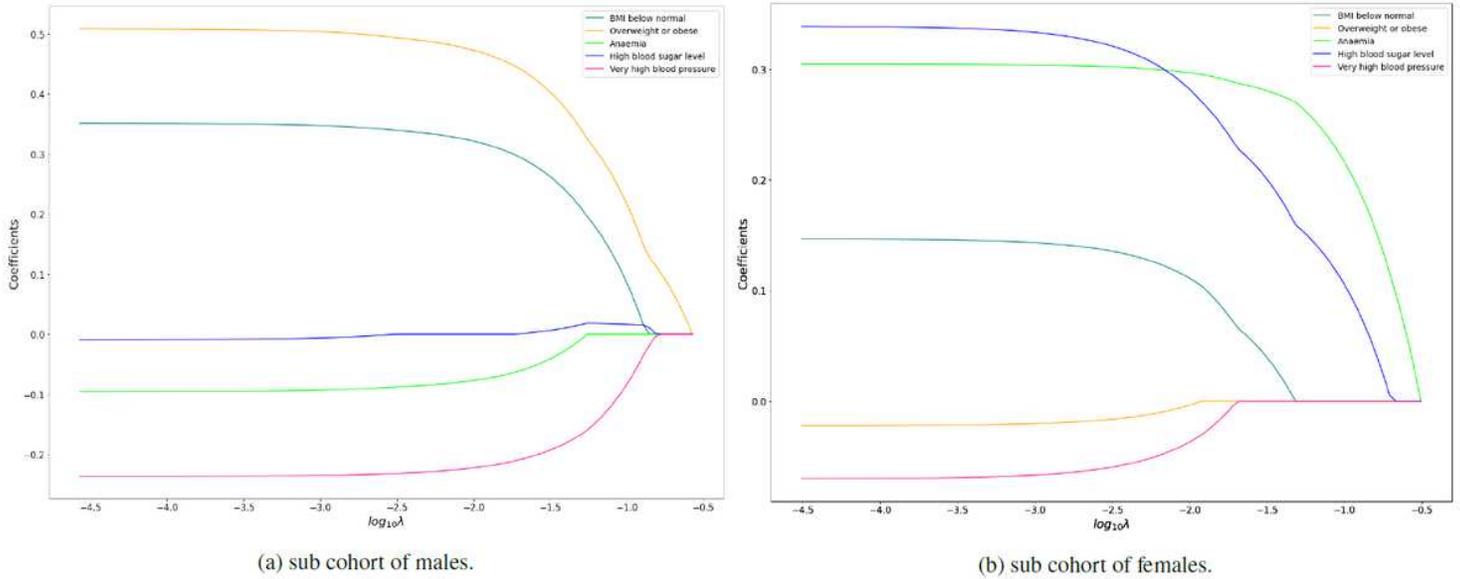


Figure 6

Lasso regression on gender-wise sub-cohorts of Karnataka. (a) $R^2 = 0.1917$ (b) $R^2 = 0.1529$; Teal - BMI below normal, Orange - Overweight or obese, Green - Anaemia, Blue - High blood sugar level, Pink - Very high blood pressure.

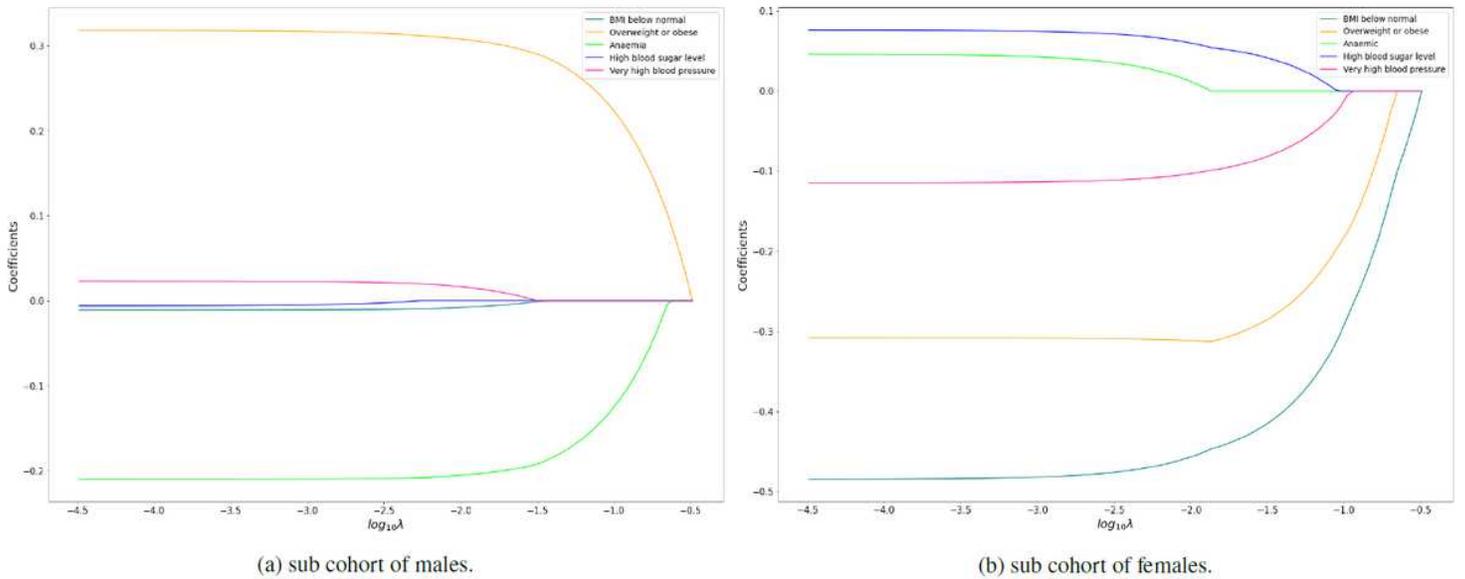


Figure 7

Lasso regression on gender-wise sub-cohorts of Tamil Nadu. (a) $R^2 = 0.1567$ (b) $R^2 = 0.2270$; Teal - BMI below normal, Orange - Overweight or obese, Green - Anaemia, Blue - High blood sugar level, Pink - Very

high blood pressure.

	Positive factors from lasso tests	Negative factors from lasso tests	Selected factors from Mann-Whitney U test
All India	Obesity BMI below normal	Anaemia Blood sugar level high Blood pressure very high	Anaemia Obesity
Common correlates	Common positive correlates		Common negative correlates
	Obesity		Anaemia

	Positive factors from lasso tests	Negative factors from lasso tests	Selected factors from Mann-Whitney U test
Sub cohort of 2 states	BMI below normal Blood sugar level high Blood pressure very high	Obesity Anaemia	Obesity BMI below normal Anaemia
Common correlates	Common positive correlates		Common negative correlates
	BMI below normal		Obesity, Anaemia

Male

Female

	Positive factors from lasso tests	Negative factors from lasso tests	Selected factors from Mann-Whitney U test
Karnataka	Obesity BMI below normal	Blood pressure very high Anaemia	Anaemia
Tamil Nadu	Obesity Blood pressure very high	Anaemia	
Common correlates	Common positive correlates		Common negative correlates
	-		Anaemia

	Positive factors from lasso tests	Negative factors from lasso tests	Selected factors from Mann-Whitney U test
Karnataka	Anaemia Blood sugar level high BMI below normal	Blood pressure very high Obesity	Obesity Anaemia BMI below normal
Tamil Nadu	Blood sugar level high Anaemia	BMI below normal Obesity Blood pressure very high	
Common correlates	Common positive correlates		Common negative correlates
	Anaemia		Obesity

Figure 8

Summary of influencing factors across regions and gender.