Supplement 1: Variables included in predictive models

For each category of variable, we list for each model it was included in.

Continuous variables are marked with an asterisk (*). All continuous variables were standardized so that the odds ratio reflects the difference in shifting one standard deviation from the mean on that variable.

Demographic variables (included in all models)

Gender

Age

Religion (coded as Hindu, Christian, Muslim/Other … non-respondents coded to Hindu)

Caste (coded as Scheduled Caste/Scheduled Tribe, Other Backwards Caste, None of the above, or No response)

Literate (Can you read and write? Yes/No [This was asked as a joint question, so we can't separate out read and write])

Years of education* (0 = 0-4 years; 1 = 5-9 years; 2 = 10-12 years; 3 = some diploma/college or completed diploma/college)

Job (Homemaker; Laborer (Laborer/brick layer, Driver); Professional (Teacher, Owns Business, Private Sector Job, Government Job, Professionals (self-emp)); Other (Student, Housemaid, Other); Not working (retired/unemployed/disabled)

Wealth score* (Based on PCA analysis of assets detailed in Supplement 2. Assets include: Own bike, own scooter, own car, own smartphone, own cellular phone only, own landline, house is private flat, house is housing board apt, house is hut, house is standalone house, house is unit in building, house is other)

Sample Original Chennai-representative sample or slum-based booster sample

General Drivers (Included in all models)

Pre-existing illness (Reports currently having: Sugar, BP, heart disease, high cholesterol, arthritis, cancer, pneumonia, malaria, typhoid, malnutrition, HIV/AIDS, TB, other)

Relative with pre-existing illness (Reports parents or grandparents having or having had any of the illnesses above)

Value good health*

- "I do whatever I can to stay healthy"
- "I get as much information as possible on my medications"
- "Good health is essential for a good life"

(1-5 points for each statement based on answer of Never/Not Often/Half the time/Most of the time/All of the time)

Smoker (Y/N – yes if reports smoking 1 or more days in the last week)

Problem drinker (Y/N – yes if reports drinking 1 or more days in the last week AND answers yes to either of the following questions: "Do you feel like drinking when you are tired or stressed?" or "Is it hard to stop drinking once you start?")

Exercise (Y/N – yes if reports exercising 1 or more days in the last week)

Get enough sleep (Y/N – yes if reports sleeping more than 6 hours/night)

Worry about healthcare cost* (1-5 depending on endorsement of statement "How often do you worry about expenses due to health problems?")

Worry about transportation cost* (1-5 depending on endorsement of statement "How often do you worry about expenses due to transportation?")

General financial stress* (1-5 depending on average endorsement of statements "How often do you worry about expenses due to costs of food and household?", "How often do you worry about expenses due to costs of rent and utilities?", "How often do you worry about expenses due to general expenditures?")

Tendency to delay gratification* (-1, 0, or 1, based on response to the following question: Now, I am going to describe two different treatments.

> [show each card and read option]
>
> Treatment X: This treatment will keep you in excellent health for 1 year, and you will feel better right away.
>
> Treatment Y: This treatment will keep you in excellent health for 3 years, but you have to wait 6 months before you feel better.
>
> If you have to choose today, which treatment will you choose?
>
> +1 for Treatment Y, -1 for Treatment X, 0 for no preference)

Level of perceived family/community care when sick* (1 point each for having a spouse who cares for you when you're sick, a parent who cares for you when you're sick, a child who cares for you when you're sick, another family member who cares for you when you're sick, and a non-family member who cares from you when you're sick)

Family push to seek who care* (The total number of people who will urge you to go see a doctor or take medication when you feel sick (from C04) divided by the total number of people who live with you (from C01) plus 5 (for the 5 categories of people who don't live with you who could potentially urge you to seek care asked about in C04))

Facility preference (government, private, no difference)

Facility preference mismatch w/community – Yes if:

- you think public is better (B01) and those who urge you to seek care when you're sick urge you to go somewhere other than a "government health facility" or "nowhere in particular – they let me choose for myself" (C05), or

- you think private is better (B01) and those who urge you to seek care when you're sick urge you to go somewhere other than "private doctors" or "nowhere in particular – they let me choose for myself" (C05)

Otherwise, No.

Perceived number of people who help with obligations*– Based on: "When you have an urgent need, how many people can help you with work or family obligations?". Re-coded to groups for 0, 1, 2, 3-4, or 5+

Get health info from doctor -- Y/N for getting health info from doctor

Get health info from community*– 1 point each for getting info from family, friends, community elders

Get health info from govt– 1 point for getting health info from the government

Get health info from media*– 1 point each for getting info from television, newspapers/magazines, billboards, radio, internet

Aware of TB– Y/N for having heard of TB

Other disease knowledge* – 1 point each for having heard of the following diseases: malnutrition, HIV/AIDS, pneumonia, malaria, typhoid

TB-specific Drivers (Included in TB knowledge and beliefs model

Believe TB is curable– Y/N for knowing TB is curable

Know free TB treatment exists– Y/N for knowing treatment of TB is available for free in Chennai

Believe nearby facility tests for TB – Y/N for reporting there is a facility nearby that tests for TB

Perceived % in community who will get TB*– Log of answer to "out of 1,00 people in your area, how many will get TB?"

Believe more likely to get TB than community– "Compared to everyone who lives in this area, are you more likely to get TB, less likely to get TB, or about the same?"

Perceived TB stigma*-- 1 point for each group of people endorsed for "If a person has TB, would they keep it a secret from their…" (options are family who live with them, family who live elsewhere, friends, neighbours, co-workers)

Symptom-specific Drivers (included in symptom characteristics model)

Symptom days* – log of # of days since symptom onset for longest-reported symptom (of cough, fever or blood in sputum)

Symptom recurrence score* -- based on the number of times in the past 12 months you've had these symptoms, not counting this time. 0 = 0 times; 1 = 1 time; 2 = 2 times; 3 = 3 times; 4 = 4-5 times; 5 = 6-10 times; 6 = 11+ times

Symptom Count* – 1 point for each symptom reported of: cough, fever, weight loss, blood in sputum

*Continuous variable. All continuous variables were standardized so that the odds ratio reflects the difference in shifting one standard deviation from the mean on that variable.

Supplement 2: Wealth score methodology and variables

The wealth index is a composite measure of a household's cumulative living standard. The wealth index is calculated using data on a household's ownership of selected assets. We followed the methodology outlined by the Demographic and Health Surveys Program.[1]

We calculated our wealth index based on the following conditions: ownership of phone and type of phone; ownership of bike; ownership of scooter; ownership of car; and living conditions (private flat, unit in housing board apartment, hut, standalone house, unit in building, or other). We discretized variables and then used a Principal Components Analysis to calculate the magnitude of the loadings of the first PCA component for each variable. We estimated the contribution of each variable to the overall wealth index using the following formula: Wealth Index value = ((1 – mean value)/std.dev)*component 1 loading score. Values were summed over all wealth index variables to obtain a wealth index score for each variable.

[1]https://www.dhsprogram.com/topics/wealth-index/Wealth-Index-Construction.cfm

https://dhsprogram.com/programming/wealth%20index/Steps_to_constructing_the_new_DHS_Wealth_Index.pdf

Supplement 3: Diagram of recruitment process in Chennai-representative sample and slum-based sample

# Households listed	Households screened	# Individuals screened	# Individuals eligible based on symptom screener	# Individuals enrolled (with completed survey)	# enrolled TB presumptives	# enrolled TB pre-presumptives
Chennai-representative sample	49401	49401	705	533	383	150
Slum-based sample	35224	85600	1217	1134	597	537
Overall	84625	135001	1922	1667	980	687

Supplement 4: Results tables for logistic regression models

We performed a series of multivariate logistic regressions were performed to identify characteristics that were associated with a higher likelihood of having sought care from a government or private doctor at the time of the initial interview. Results from each regression are reported below. We report the odds ratio, upper and lower bounds of the 95% confidence interval for the odds ratio, and p-value. Full descriptions of variables can be found in Supplement 1.

a. Model 1: Predictors of care seeking from the "general" logistic regression which included data from all surveyed participants (n = 1667).

Variable	Odds Ratio	Lower CI	Upper CI	p-value	Category
Gender: male	1.72	1.22	2.43	**0.002**	Demographics
Job: retired/unemployed/disabled vs. homemaker	1.29	0.87	1.92	0.202	Demographics
Age*	1.20	1.04	1.38	**0.012**	Demographics
Caste: SC/ST vs. refused	1.17	0.68	1.99	0.572	Demographics
Caste: none of the above vs. refused	1.15	0.66	2.01	0.618	Demographics
Religion: Hindu vs. Christian	1.15	0.83	1.59	0.387	Demographics
Literate	1.14	0.87	1.51	0.343	Demographics
Religion: Muslim/other vs. Christian	1.10	0.68	1.77	0.699	Demographics
Wealth score*	0.99	0.88	1.10	0.831	Demographics
Years of education*	0.99	0.86	1.14	0.851	Demographics
Caste: OBC vs. refused	0.95	0.55	1.63	0.858	Demographics
Job: other vs. homemaker	0.74	0.51	1.09	0.130	Demographics
Sample: Original Chennai-representative	0.74	0.58	0.93	**0.010**	Demographics
Job: laborer vs. homemaker	0.74	0.53	1.03	0.072	Demographics
Job: professional vs. homemaker	0.58	0.39	0.87	**0.008**	Demographics
Get health info from doctor	1.43	1.02	2.01	**0.038**	General drivers
Pre-existing illness	1.28	1.02	1.61	**0.035**	General drivers
Value good health*	1.19	1.07	1.33	**0.001**	General drivers
Facility preference: private vs. government	1.18	0.92	1.50	0.197	General drivers
Worry about transportation cost*	1.17	1.02	1.34	**0.021**	General drivers
Get health info from govt	1.16	0.91	1.49	0.233	General drivers
Other disease knowledge*	1.12	0.98	1.28	0.111	General drivers
Relative with pre-existing illness	1.08	0.85	1.38	0.510	General drivers
Level of perceived family/community care when sick*	1.07	0.95	1.22	0.282	General drivers
Worry about healthcare cost*	1.07	0.94	1.21	0.313	General drivers
Perceived number of people who help with obligations*	1.03	0.92	1.16	0.574	General drivers
General financial stress*	1.01	0.88	1.15	0.917	General drivers
Tendency to delay gratification*	0.95	0.85	1.06	0.341	General drivers

Family push to seek care*	0.94	0.84	1.06	0.343	General drivers
Get health info from community*	0.94	0.83	1.07	0.351	General drivers
Get health info from media*	0.89	0.78	1.01	0.062	General drivers
Get enough sleep	0.78	0.63	0.98	**0.029**	General drivers
Exercise	0.69	0.52	0.92	**0.012**	General drivers
Facility preference: no difference vs. government	0.66	0.48	0.90	**0.010**	General drivers
Smoker	0.60	0.41	0.87	**0.007**	General drivers
Facility preference mismatch w/ community	0.54	0.40	0.73	**0.000**	General drivers
Problem drinker	0.53	0.35	0.80	**0.002**	General drivers
Aware of TB	0.98	0.68	1.41	0.919	TB-specific drivers

*Continuous variables are scaled so that the odds ratio reflects the impact of one standard deviation increase in the variable.

b. **Model 2:** Impact of TB belief and information predictors on odds of seeking formal care from a doctor for symptoms associated with tuberculosis. Model includes data from all surveyed participants with TB-indicative symptoms who had heard of TB (n = 1427).

Variable	Odds Ratio	Lower CI	Upper CI	p-value	Category
Religion: Hindu vs. Christian	1.35	0.96	1.89	0.084	Demographics
Religion: Muslim/other vs. Christian	1.24	0.76	2.03	0.396	Demographics
Age*	1.19	1.04	1.36	**0.013**	Demographics
Caste: SC/ST vs. refused	1.15	0.64	2.06	0.637	Demographics
Job: retired/unemployed/disabled vs. homemaker	1.12	0.75	1.71	0.582	Demographics
Literate	1.09	0.81	1.46	0.582	Demographics
Gender: male	1.07	0.79	1.46	0.646	Demographics
Caste: none of the above vs. refused	1.02	0.56	1.87	0.938	Demographics
Wealth score*	1.01	0.90	1.13	0.866	Demographics
Years of education*	1.00	0.87	1.15	0.981	Demographics
Caste: OBC vs. refused	0.97	0.54	1.73	0.914	Demographics
Sample: Original Chennai-representative	0.68	0.54	0.87	**0.002**	Demographics
Job: other vs. homemaker	0.62	0.42	0.92	**0.016**	Demographics
Job: laborer vs. homemaker	0.60	0.43	0.85	**0.004**	Demographics
Job: professional vs. homemaker	0.52	0.35	0.78	**0.002**	Demographics
Believe nearby facility tests for TB	1.17	0.90	1.51	0.241	TB-specific drivers
Believe more likely to get TB than community	1.13	1.01	1.26	**0.039**	TB-specific drivers
Know free TB treatment exists	1.06	0.81	1.39	0.670	TB-specific drivers
Perceived TB stigma	0.94	0.84	1.05	0.249	TB-specific drivers
Believe TB is curable	0.92	0.72	1.17	0.503	TB-specific drivers
Perceived % in community who will get TB*	0.84	0.75	0.94	**0.003**	TB-specific drivers

*Continuous variables are scaled so that the odds ratio reflects the impact of one standard deviation increase in the variable.

c. **Model 3:** Impact of symptom characteristics predictors on odds of seeking formal care from a doctor for symptoms associated with tuberculosis. In order to prevent confounding due to the different symptom enrollment criteria used in the two samples, this regression included data only from respondents with TB-indicative symptoms in the slum-based sample (n = 1134).

Variable	Odds Ratio	Lower CI	Upper CI	p-value	Category
Caste: SC/ST vs. refused	1.76	0.70	4.31	0.219	Demographics
Religion: Muslim/other vs. Christian	1.42	0.79	2.55	0.238	Demographics
Caste: OBC vs. refused	1.36	0.54	3.35	0.502	Demographics
Caste: none of the above vs. refused	1.32	0.52	3.31	0.555	Demographics
Literate	1.30	0.93	1.82	0.121	Demographics
Age*	1.22	1.04	1.43	**0.013**	Demographics
Years of education*	1.14	0.97	1.35	0.110	Demographics
Religion: Hindu vs. Christian	1.04	0.71	1.52	0.851	Demographics
Job: retired/unemployed/disabled vs. homemaker	0.96	0.61	1.53	0.876	Demographics
Wealth score*	0.94	0.83	1.07	0.358	Demographics
Gender: male	0.81	0.57	1.15	0.246	Demographics
Job: other vs. homemaker	0.69	0.43	1.11	0.122	Demographics
Job: laborer vs. homemaker	0.61	0.41	0.90	**0.012**	Demographics
Job: professional vs. homemaker	0.60	0.37	0.97	**0.035**	Demographics
Symptom days*	1.59	1.37	1.85	**0.000**	Symptom-specific drivers
Symptom count*	1.28	1.12	1.48	**0.000**	Symptom-specific drivers
Symptom recurrence score*	0.87	0.76	0.99	**0.029**	Symptom-specific drivers

*Continuous variables are scaled so that the odds ratio reflects the impact of one standard deviation increase in the variable.