


# Understanding why at-risk population segments do not seek care for tuberculosis: a precision public health approach in South India

Sarah Helfinstein,<sup>1</sup> Elisabeth Engl,<sup>1</sup> Beena E Thomas,<sup>2</sup> Gayathri Natarajan,<sup>3</sup> Prithivi Prakash,<sup>3</sup> Mokshada Jain,<sup>1</sup> Jayabal Lavanya,<sup>4</sup> Murugesan Jagadeesan,<sup>5</sup> Rebekah Chang,<sup>3</sup> Tichakunda Mangono,<sup>1</sup> Hannah Kemp,<sup>1</sup> Shamim Mannan,<sup>3</sup> Harkesh Dabas,<sup>3</sup> Grace K Charles,<sup>1</sup> Sema K Sgaier <sup>1,6,7</sup>

**To cite:** Helfinstein S, Engl E, Thomas BE, *et al.* Understanding why at-risk population segments do not seek care for tuberculosis: a precision public health approach in South India. *BMJ Global Health* 2020;**5**:e002555. doi:10.1136/bmjgh-2020-002555

**Handling editor** Stephanie M Topp

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjgh-2020-002555>).

SH and EE contributed equally.

Received 31 March 2020  
Revised 15 July 2020  
Accepted 16 July 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr Sema K Sgaier;  
semasgaier@surgofoundation.org

## ABSTRACT

**Introduction** Delaying care-seeking for tuberculosis (TB) symptoms is a major contributor to mortality, leading to worse outcomes and spread. To reduce delays, it is essential to identify barriers to care-seeking and target populations most at risk of delaying. Previous work identifies barriers only in people within the health system, often long after initial care-seeking.

**Methods** We conducted a community-based survey of 84 625 households in Chennai, India, to identify 1667 people with TB-indicative symptoms in 2018–2019. Cases were followed prospectively to observe care-seeking behaviour. We used a comprehensive survey to identify care-seeking drivers, then performed multivariate analyses to identify care-seeking predictors. To identify profiles of individuals most at risk to delay care-seeking, we segmented the sample using unsupervised clustering. We then estimated the per cent of the TB-diagnosed population in Chennai in each segment.

**Results** Delayed care-seeking characteristics include smoking, drinking, being employed, preferring different facilities than the community, believing to be at lower risk of TB and believing TB is common. Respondents who reported fever or unintended weight loss were more likely to seek care. Clustering analysis revealed seven population segments differing in care-seeking, from a retired/unemployed/disabled cluster, where 70% promptly sought care, to a cluster of employed men who problem-drink and smoke, where only 42% did so. Modelling showed 54% of TB-diagnosed people who delay care-seeking might belong to the latter segment, which is most likely to acquire TB and least likely to promptly seek care.

**Conclusion** Interventions to increase care-seeking should move from building general awareness to addressing treatment barriers such as lack of time and low-risk perception. Care-seeking interventions should address specific beliefs through a mix of educational, risk perception-targeting and social norms-based campaigns. Employed men who problem-drink and smoke are a prime target for interventions. Reducing delays in this group could dramatically reduce TB spread.

## Key questions

### What is already known?

- Tuberculosis (TB) is a leading cause of death worldwide, with India contributing the largest number of patients with TB.
- Delays in promptly seeking care for symptoms suggestive of TB are common in India and are a major contributor to ongoing TB mortality; retrospective studies on patients suggest they are influenced by demographic as well as access-related and awareness-related factors.

### What are the new findings?

- In the South Indian city of Chennai, we identified a broad spectrum of barriers to prompt care-seeking, including demographic factors such as occupation, beliefs around perceived risk (higher care-seeking), normalisation of TB (lower care-seeking) and having different facility preferences than one's community (lower care-seeking), as well as symptom-related factors such as experiencing sustained unintentional weight loss or having a fever for at least 7 days (higher care-seeking).
- TB awareness is high and (contrary to previous research) not associated with care-seeking delays.
- Segmentation revealed that a high-impact subpopulation of employed men who problem-drink and smoke may comprise over half of Chennai's TB-diagnosed individuals who do not promptly seek care, and that this group is both particularly unlikely to seek care and particularly likely to have TB.

## INTRODUCTION

Tuberculosis (TB) is the 10th leading cause of death in the world and, with 1.5 million deaths annually, the infectious disease responsible for the most deaths.<sup>1</sup> India is the largest contributor to the global TB burden, accounting for 2.7 million of the 10 million new TB cases in 2018.<sup>1</sup> With the establishment of the Sustainable

## Key questions

### What do the new findings imply?

- ▶ TB awareness campaigns alone may not drive care-seeking, and normalisation of TB may be counterproductive; instead, making service delivery more convenient could be an immediate priority.
- ▶ Addressing high-impact segments such as employed men who problem-drink and smoke with targeted campaigns around their risk perception and other key barriers, rather than employing population-level, one-size-fits-all interventions, would allow programmes to focus efforts on the subset of the population that is most at risk of acquiring TB and least likely to seek care.

Development Goals, the global TB community has set ambitious targets to combat the global TB epidemic, aiming to reduce TB incidence by 90% and TB deaths by 95% between 2015 and 2035.<sup>2</sup> Achieving these targets in the face of limited resources will require national and local programmes to take a ‘precision public health approach’. We define precision public health as efficiently delivering the right intervention, to the right person at the right place and time to achieve the desired impact using the best available data and technologies.<sup>3</sup> In this context, interventions should therefore focus on the people most at risk of contracting and spreading the disease, and target the key underlying barriers to care-seeking and treatment.<sup>4</sup>

Recent research suggests that the delay between the onset of symptoms and first contacting a provider to seek care is one of the greatest contributors to ongoing TB mortality and incidence.<sup>5 6</sup> Reducing the length of time between symptom onset and care-seeking can be particularly impactful in reducing TB incidence because it shortens the window during which people with symptoms are contagious and can spread the disease to others. Indeed, modelling indicates that, in India, if the average care-seeking delay for those infected with TB were reduced by 25%, TB mortality would be reduced by roughly 6%, and incidence of new cases would similarly decline.<sup>5</sup> Indian symptomatics experience significant delays in care-seeking, with people exhibiting symptoms for a median of 18.4 days without seeking care, and a subsegment of TB symptomatics delaying care-seeking for up to several months.<sup>7</sup>

Understanding what drives those with TB symptoms to delay care is essential to designing effective interventions. With limited resources, this requires a precision public health approach: programmes need to know whether certain segments of the population are most at risk for not seeking care promptly and understand their potential barriers to care-seeking. Then, efforts can be targeted to the right people with the right interventions based on what segment is most prevalent, most at risk, most easily accessed and most easily ‘converted’ to the right behaviour.<sup>8</sup> Many studies have been conducted both globally<sup>9–14</sup> and in the Indian context<sup>15–20</sup> to better understand TB patient care-seeking pathways. These studies find a number of factors to be commonly

associated with delayed care-seeking, including lack of knowledge about TB,<sup>10–13</sup> being a woman,<sup>11 14 20</sup> lack of education,<sup>11</sup> living far from health facilities,<sup>11 12 15 20</sup> visiting informal care providers,<sup>11 13</sup> low income in Africa and Asia,<sup>14 20</sup> high income in South India,<sup>16</sup> age,<sup>9 10 13 19</sup> and tobacco and alcohol use.<sup>16 21–25</sup> Although existing research provides some insight into what drives people to seek care for their TB symptoms, these studies often only focus on a limited set of potential drivers, many of which are demographic variables. These variables alone are difficult to address without knowing associated differences in knowledge, attitudes, resources, and health or information-seeking behaviours. Additionally, the majority of studies survey patients in hospitals or other health facilities after they have already chosen to seek care.<sup>10–17 21 26</sup> Thus, importantly these studies miss people who have failed to seek care at all, and are subject to participant recall bias since they retrospectively survey patients about care-seeking decisions that often happened months in the past.

Our study fills existing knowledge gaps in several ways. To our knowledge, this study is the first that uses a prospective design to capture early care-seeking behaviour in TB. This allowed us to systematically identify people with TB symptoms who may never seek care, or seek care only informally, to develop a more representative sample of TB symptomatics. The longitudinal, prospective design allowed us to observe care-seeking in real time to reduce recall biases. Our community-based survey is representative of Chennai, the sixth largest metropolis in India, which accounts for a quarter of the world’s TB cases.<sup>27</sup> Finally, we used a comprehensive survey built on the CUBES framework of behaviour<sup>28</sup> that allowed us to systematically assess a wide range of possible drivers of and barriers to care-seeking, including beliefs, knowledge, social support, other health behaviour patterns and demographic characteristics. After identifying key predictors of TB care-seeking, we used unsupervised clustering to create distinct and actionable segments of people that differed in their care-seeking behaviour and its drivers.

The goal of segmentation was to identify subgroups of the population that may have a particularly low propensity to seek care. We show that such a segment, while small relative to the general population, may comprise a large section of Chennai’s TB-diagnosed population (ie, appears to have an enhanced risk for becoming diagnosed with TB). Therefore, targeting it could lead to outsize improvements in care-seeking. Segmentation has been effective for identifying and targeting individuals for health behaviour change in other contexts,<sup>29 30</sup> and can be used to target those most in need using interventions that address their specific barriers to seek care. The insights from this study point to the need for segment-targeted programmatic interventions to drive timely care-seeking, so that limited resources can be used efficiently to reduce TB incidence and mortality.

## METHODS

Written informed consent was obtained from all study participants. Patients or the public were not involved in the design or reporting of this research. At the end of the initial survey, all participants who met the definition of 'TB presumptive' based on their symptoms (see Community-representative sampling) were given a card with the address of the nearest public health clinic, urged to see a doctor for their current illness and provided with directions to the facility if they did not know where it was located.

This community-based study was conducted in the city of Chennai, India. Two samples of households were collected: a community-representative sample and a slum-based sample. Each sample had different sampling and selection criteria. The community-representative sample was collected as part of an initiative by the city of Chennai to obtain a more up-to-date estimate of TB prevalence across the city. The slum-based sample was then sampled specifically from slum areas to capture a larger proportion of TB symptomatics relative to the main sample and to achieve an adequate sample size for the proposed analyses examining predictors of care-seeking.

### Patient and public involvement

This study did not involve patients in the design, recruitment or implementation phase. However, cognitive pretests of survey questions were conducted in the community from which participants were selected. TB patient and healthcare provider reports of delays in the care-seeking pathway, revealed in previous work, informed the development of the research question. Due to the large-scale, community-based recruitment, results will not be disseminated to each participant. Instead, study results are being implemented in Chennai-based programmes to optimise care-seeking outreach.

### Community-representative sampling

Data from the community-representative sample were collected between August and December 2018. For this sample, 100 out of 200 wards from the city of Chennai were randomly selected based on probability proportional to size. Within each selected ward, a random sample of four streets were selected as starting points, and the right-hand rule was used to list 200 houses per starting point. In each household, one adult (age 18+) was randomly selected to be screened for symptoms and three recontact attempts were made if the person was not present. In line with medical guidance in Chennai for symptoms requiring referral to get tested for TB, individuals who reported blood in the sputum in the last month, and/or a cough for 14 or more days, and/or fever for 14 or more days were classed and enrolled as 'TB presumptive' symptomatics. This category also included individuals who reported unintended substantial weight loss in the last 3 months (from 6 October onwards only, due to a change in question wording). Individuals with milder symptoms, reporting a cough or fever for between 7 and 14 days,

were enrolled as 'TB pre-presumptives'. To reduce the potential for TB-related social desirability bias, participants were also asked about a set of unrelated symptoms, which were not used as screening criteria.

### Slum-based sampling

Data from the slum-based sample were collected between March and July 2019. Due to the nature of slum layouts, the main sampling method was adjusted. Slums from a list by the Tamil Nadu Slum Rehabilitation Board were sorted by the number of households in descending order, and starting from the most populous 2000 households or less were listed in a given slum to avoid over-representation. Slums were mapped and checked against a set of qualitative criteria, such as the presence of open sewers and of shacks with metal or asbestos roofs. A total of 39 slums were sampled. To further increase identification of TB presumptives and pre-presumptives, initial screening by proxy (other household members) was employed, with all adults in the household being first screened directly or by proxy for TB symptoms. If multiple TB presumptives or pre-presumptives were identified in the household, one was randomly selected to be enrolled in the study.

### Data collection

#### Initial interview

To maximise the chance of enrolling a representative sample, data collection took place on weekdays and weekends, in daytime and evening hours. After symptom screening, all TB presumptives and pre-presumptives were invited to enrol in the study. If they agreed to participate, they were immediately given ~45 min interview assessing both their care-seeking to date for their symptoms, as well as a wide variety of possible drivers of care-seeking behaviours. Questionnaire items were developed based on the CUBES framework<sup>28</sup> to systematically assess possible perceptual and contextual drivers, as well as household, community and media influencers. In addition to the items assessing care-seeking, survey items addressed general health facilities perceptions; social support and influencers; general health status, beliefs and biases; information channels; disease knowledge, beliefs and stigma; and a sociodemographic profile.

'Not seeking care promptly' was defined as people who had not yet sought care by the time they were initially surveyed and is used as the outcome variable for the subsequent analyses. 'Formal' care was defined as seeking care from a doctor at either a government or private facility.

#### Follow-up care-seeking assessments

While not the focus of the current study, participants were then followed up to determine what care-seeking practices they engaged in, if any, for 6 weeks after the initial contact.

### Data analysis

Data were analysed using R. Unless otherwise indicated below, analyses incorporated data from both samples



and all enrolled participants (TB presumptives and pre-presumptives) for statistical power.

### Predictors of care-seeking

First, 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. Predictor variable selection was based around the CUBES framework.<sup>28</sup> Drivers considered included demographic variables (eg, gender, occupation, wealth score), symptom characteristics (eg, type and quantity of symptoms), perceptions of health facilities (eg, preference for government or private facilities), social support and influencers (eg, number of people who can help with obligations in a time of need), health practices and beliefs (eg, smoking, drinking), information channels (eg, getting health information from the government), and disease knowledge and beliefs (eg, knowing that treatment for TB is available for free in Chennai). Details of all variables can be found in online supplementary 1. Details of the calculation of the wealth score are detailed in online supplementary 2. All linear variables were standardised so that the OR reflects the difference in shifting 1 SD from the mean on that variable. Three logistic regressions were performed on slightly different samples: (1) a 'general' logistic regression, which included data from all surveyed participants; (2) a 'TB knowledge and beliefs' logistic regression, which included data only from the 85% of participants who reported having heard of TB; and (3) a 'symptom characteristics' logistic regression, which included data only from the 68% of the sample that were collected as part of the slum-based sample, in order to prevent confounding due to the different symptom enrolment criteria used in the two samples. All three regressions included demographic variables as controls. Variables were tested for multicollinearity and no variables used in any regression were found to have a variance inflation factor greater than 2.5.

### Care-seeking segmentation

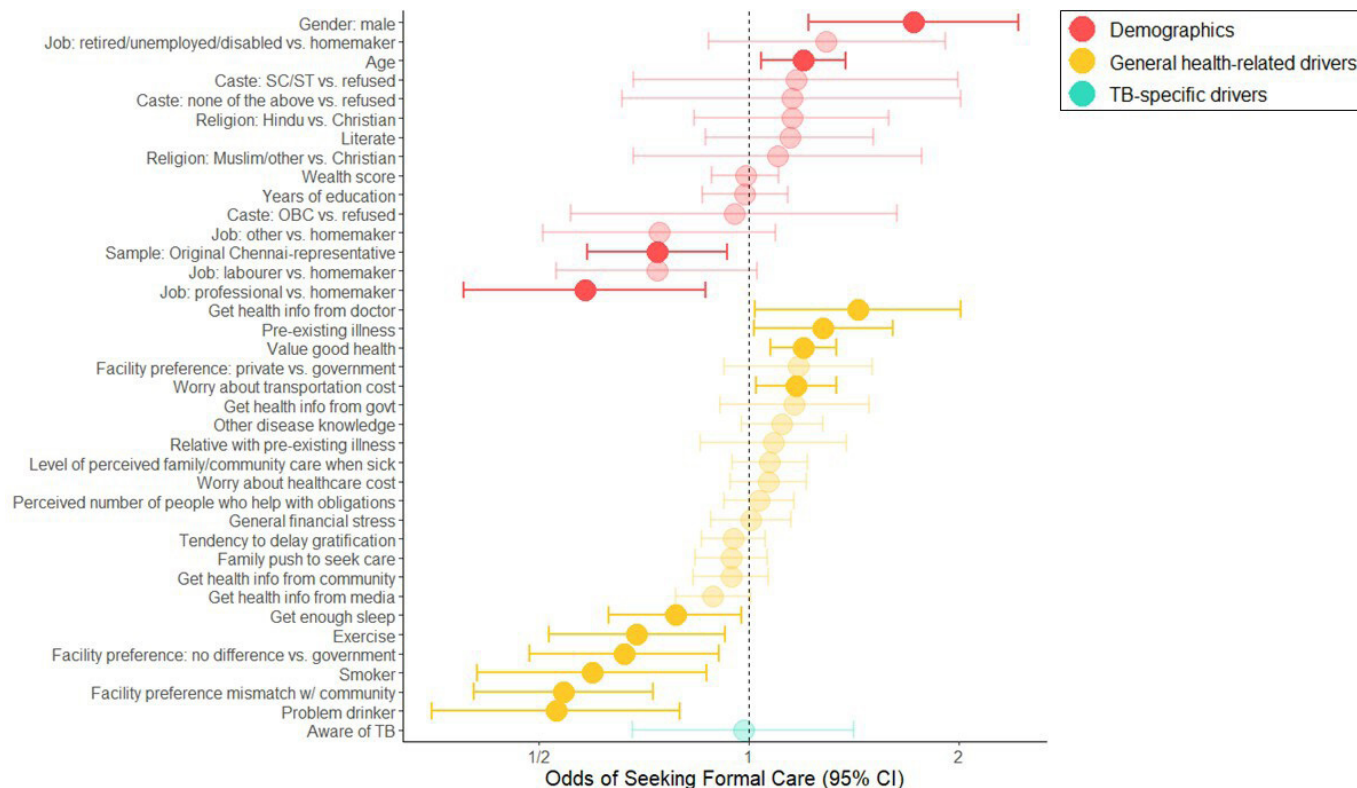
Second, a k-medoid partitioning around medoids (PAM) clustering algorithm (with a Gower distance metric) was used to identify clusters of TB presumptives and pre-presumptives that differed on the following seven variables: gender, smoking status (having smoked in the past week), problem-drinker status (having had a drink in the past week and answering yes to either 'Do you feel like drinking when tired or stressed?' or 'Is it hard to stop drinking once you stop?'), awareness of a nearby facility that tests for TB and two variables related to employment status (paid job vs no paid employment, and home maker vs retired/unemployed/disabled). These variables were selected for segmentation based on their relations to care-seeking behaviour observed in the predictive models as well as possible actionability. Segments were profiled on a variety of demographic and other characteristics.

### Care-seeking segment prevalence

Finally, to determine the likely prevalence of these segments in the TB-diagnosed population (ie, to be able to reprioritise segments based on TB infection risk<sup>8</sup>), we used data from a recent TB prevalence survey conducted in the city of Chennai to determine age, gender, smoking and alcohol use status prevalence among individuals who were culture-positive for TB.<sup>21</sup> Because information on awareness and employment of Chennai patients with TB was not available in the prevalence study, we could not use this information when estimating the distribution of the population into their risk groups. We then 'raked' the data, iteratively reweighting the sample to make it representative of the TB-diagnosed population in Chennai on the same four variables, using the *anesrake* package in R. This generated a weighted sample matching that of the TB-diagnosed population in Chennai which could be used to estimate the per cent of the TB-diagnosed

**Table 1** Participant profile

	n (%)
<b>Age (years)</b>	
18–34	371 (22)
35–54	668 (40)
55+	628 (38)
<b>Gender</b>	
Male	558 (33)
Female	1109 (67)
<b>Occupation</b>	
Home maker	756 (45)
Professional	182 (11)
Labourer	351 (21)
Retired/unemployed/disabled	221 (13)
Other	157 (9)
<b>Education</b>	
Primary (fourth standard or less)	582 (35)
Secondary (ninth standard or less)	595 (36)
Senior secondary	305 (18)
Postgraduate	185 (11)
<b>Literate</b>	
Yes	1164 (70)
No	503 (30)
<b>Classification of participants</b>	
Pre-presumptive	687 (41)
Presumptive	980 (59)
Smoked in the past week	239 (14)
Had alcohol in the past week	322 (19)
<b>Care-seeking by initial interview</b>	
Sought care from doctor by initial interview	942 (57)
Sought informal care by initial interview	771 (46)
Sought any care by initial interview	1379 (83)



**Figure 1** Impact of predictors on odds of seeking formal care from a doctor for symptoms associated with TB. Variables are coloured by category, and those in bold are significant at a cut-off of  $p < 0.05$ , uncorrected for multiple comparisons. Error bars represent 95% CI. A 'vs' on the y axis denotes a comparison relative to the right argument. For example, 'job: professional vs homemaker' indicates the changed odds of someone with a job seeking care relative to home makers. The model includes data from all surveyed participants with TB-indicative symptoms ( $n=1667$ ). OBC, other backward class; SC/ST, scheduled caste/scheduled tribe; TB, tuberculosis.

population that would fall into each segment. Based on the care-seeking rates observed for each segment in our sample, we then calculated the per cent of TB-diagnosed non-care-seekers that can likely be found within each segment.

## RESULTS

### Participant demographics and health behaviour characteristics

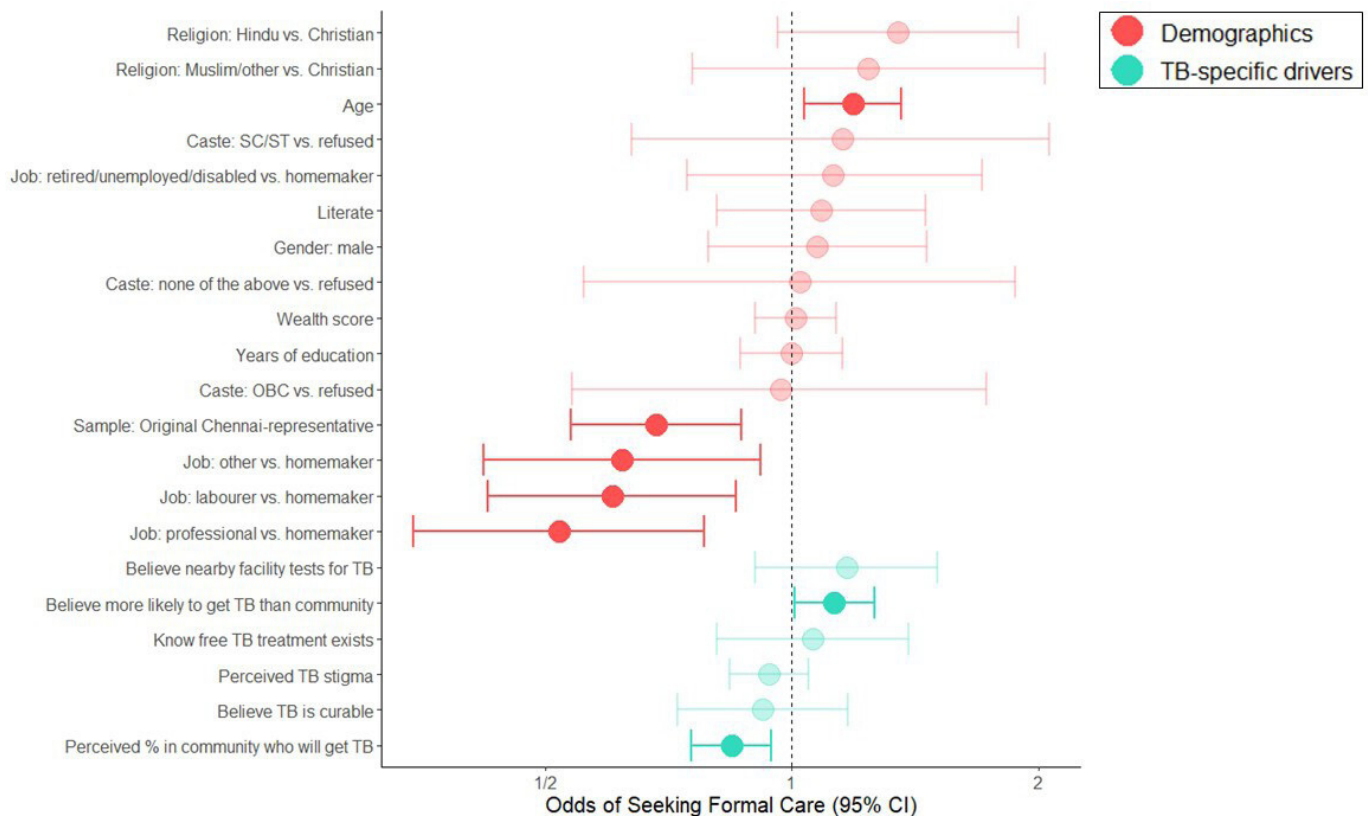
Altogether, 84 625 households were screened and 1667 TB presumptives or pre-presumptives were enrolled in our study. **Table 1** reports the demographic characteristics of all enrolled participants. Online supplementary 3 details the recruitment process for the Chennai-representative sample and the slum-based sample. Two-thirds of the enrollees ( $n=1109$ ) were female, and nearly half (45%,  $n=756$ ) were home makers (who could also, very rarely, be male). The majority were in the middle age (40%,  $n=668$ ) or older age (38%,  $n=628$ ) group. Most enrollees (70%,  $n=1164$ ) were able to read and write, but relatively few had more than secondary level education (senior secondary: 18%,  $n=305$ ; postgraduate: 11%,  $n=185$ ). Of the enrollees, 19% had consumed alcohol in the last week ( $n=322$ ) and 14% had smoked ( $n=239$ ). Altogether, 57% of the enrollees ( $n=942$ ) had sought care from a doctor for their symptoms at the time of the initial interview, and

83% ( $n=1379$ ) had sought either care from a doctor or some sort of informal care, such as treatment at a pharmacy or home remedies.

### Predictors of prompt care-seeking

#### Demographics

Across all three models, we found that both age and job type were significant predictors of care-seeking behaviour. Specifically, a 1 SD increase in age was associated with a 1.19–1.22 increase in odds of seeking care by the time of the survey across all three logistic regressions ( $p < 0.01$  for all; see online supplementary 4 for regression tables). In addition, professionals were less likely to seek care than home makers across all three regressions (adjusted OR (aOR): 0.0.52–0.60;  $p < 0.05$  for all; see online supplementary 4 for regression tables) and labourers were less likely to seek care than home makers in our models of care-seeking among those who had heard of TB (aOR=0.60,  $p=0.004$ ) and the slum-based 'symptom characteristics' regression (aOR=0.61,  $p=0.01$ ). In our general model, which included samples from the Chennai-representative sample and the slum-based sample, respondents from the Chennai-representative sample were less likely to seek care (aOR=0.74 and 0.68, respectively;  $p < 0.01$  for both). We believe this is due to the inclusion of survey respondents that only listed weight loss as a symptom in



**Figure 2** Impact of TB belief and information predictors on odds of seeking formal care from a doctor for symptoms associated with TB. Variables are coloured by category, and those in bold are significant at a cut-off of  $p < 0.05$ , uncorrected for multiple comparisons. Error bars represent 95% CI. The model includes data from all surveyed participants with TB-indicative symptoms who had heard of TB ( $n=1427$ ). OBC, other backward class; SC/ST, scheduled caste/scheduled tribe; TB, tuberculosis.

the Chennai-representative sample, but not the slum-based sample. Respondents who only listed weight loss as a symptom sought care at a rate of 34% compared with respondents with other symptoms, who sought care at a rate of 59%. Care-seeking rates for those with other symptoms were comparable between the Chennai-representative sample (59%) and the slum-based sample (58%). No other demographic variables were found to be significant predictors of care-seeking.

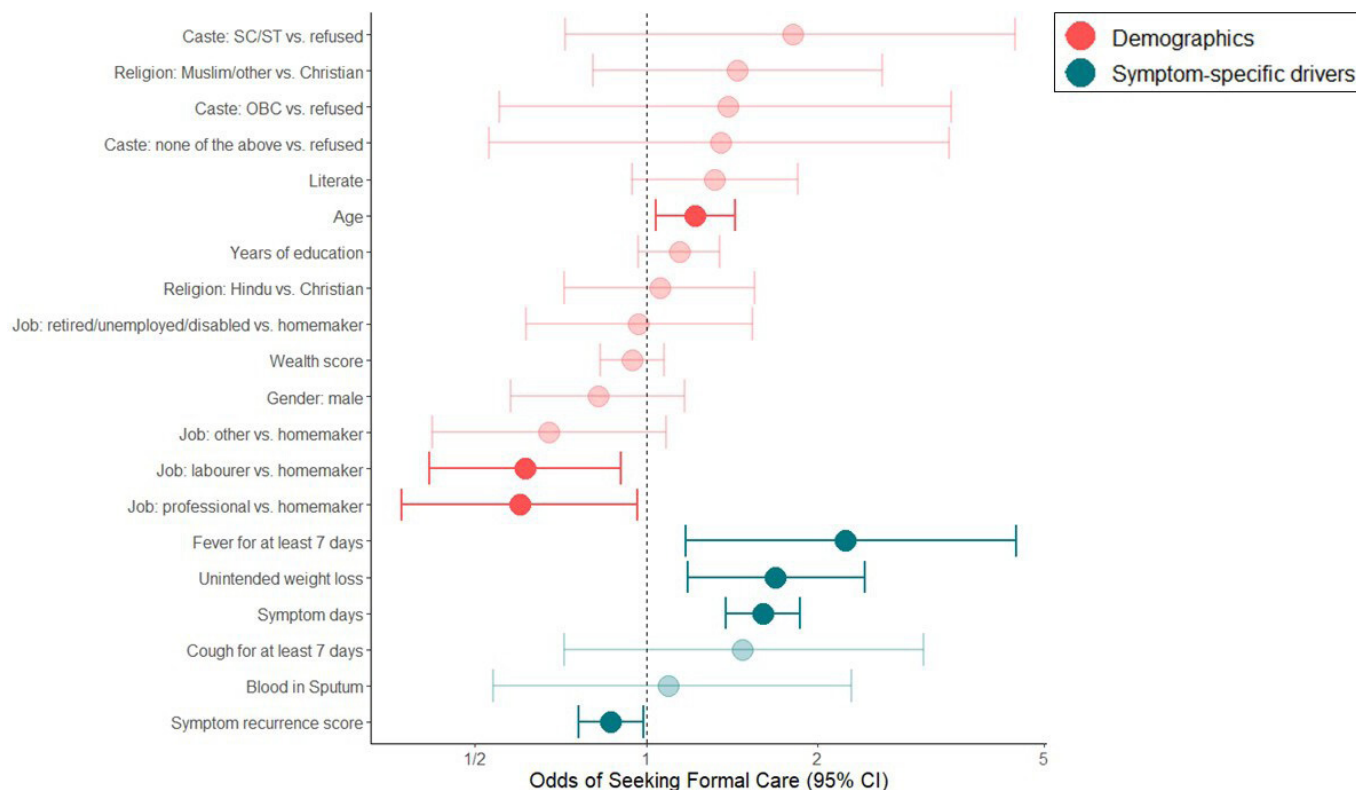
**General**

**Figure 1** displays all drivers considered as predictors of care-seeking in the ‘general’ regression and their impact on the odds of seeking medical care from a doctor at the time of the survey. In addition to demographic variables, the regression reveals that some of the strongest negative predictors of care-seeking are related to symptom type, addictive behaviours and facility type preferences. In addition, those who smoke (aOR: 0.60, 95% CI 0.41 to 0.87) and drink (aOR: 0.53, 95% CI 0.35 to 0.80) are significantly less likely to seek care than those who do not. Two factors linked to facility preferences are related to care-seeking. First, those who have no preference between government and private facilities seek care less than those who prefer government facilities (aOR: 0.66, 95% CI 0.48 to 0.90). Second, those who have a preference mismatch

with their community (ie, those who themselves prefer government facilities but believe their community wants them to seek care at a private facility, or vice versa) are substantially less likely to seek care than those who do not (aOR: 0.54, 95% CI 0.40 to 0.73). Finally, those who report that they exercise at least once per week (aOR: 0.69, 95% CI 0.52 to 0.92) and those who report that they get 7+ hours of sleep per week (aOR: 0.78, 95% CI 0.63 to 0.98) are slightly less likely to seek care than those who say they do not.

Individuals who report getting health information from a doctor (aOR: 1.45, 95% CI 1.03 to 2.04), who report greater stress related to transportation costs (aOR: 1.17, 95% CI 1.02 to 1.34), who report valuing good health more highly (aOR: 1.19, 95% CI 1.07 to 1.33) or who report having a pre-existing illness (aOR: 1.28, 95% CI 1.02 to 1.61) all are more likely to have sought care from a doctor for their symptoms. Being male is a positive predictor of care-seeking (aOR: 1.72, 95% CI 1.22 to 2.43,  $p=0.003$ ). This is likely because this analysis disaggregates gender from smoking and drinking behaviour, which were reported almost exclusively by men and which are negative predictors of care-seeking. In other words, while men as a whole are no more likely to seek care than women, a non-smoking, non-drinking man of a





**Figure 3** Impact of symptom characteristics predictors on odds of seeking formal care from a doctor for symptoms associated with TB. Variables are coloured by category, and those in bold are significant at a cut-off of  $p < 0.05$ , uncorrected for multiple comparisons. Error bars represent 95% CI. In order to prevent confounding due to the different symptom enrolment criteria used in the two samples, this regression included data only from respondents with TB-indicative symptoms in the slum-based sample ( $n=1134$ ). OBC, other backward class; SC/ST, scheduled caste/scheduled tribe; TB, tuberculosis.

given occupation is more likely to seek care than a non-smoking, non-drinking woman of the same occupation.

#### TB knowledge and beliefs

**Figure 2** outlines the effect size of predictors of care-seeking related to TB knowledge and beliefs. When considering this set of variables, two additional significant predictors of care-seeking emerge: believing that you are more likely than others in your community to get TB, which increases care-seeking behaviour (aOR: 1.13, 95% CI 1.01 to 1.26), and believing that a large number of people within your community get TB, which decreases care-seeking behaviour (aOR: 0.84, 95% CI 0.75 to 0.94).

#### Symptom characteristics

**Figure 3** outlines the effect size of predictors related to symptom type and duration on care-seeking. The duration of symptoms (aOR: 1.60 for 1 SD increase in symptom duration; 95% CI 1.38 to 1.86) increased the likelihood of care-seeking. By contrast, having had more recurrences of the symptoms over the course of the past year slightly decreased the likelihood of care-seeking (aOR: 0.87, 95% CI 0.76 to 0.99), possibly due to the normalisation of the symptoms for the individual. Two symptoms, having a fever for at least 7 days (aOR: 2.23, 95% CI 1.17 to 4.48) and experiencing unintended weight loss (aOR: 1.69, 95% CI 1.18 to 2.42), increased the odds

of care-seeking. Surprisingly, two of the most classic TB symptoms, having a cough for 7 days and having blood in sputum, did not significantly increase the odds of care-seeking ( $p=0.29$  and  $p=0.81$ , respectively).

Tables that show detailed results for all models are provided in online supplementary 4.

#### Population segments for care-seeking

Overall, we found a wide variety of care-seeking predictors. Next, we wanted to gain insight into the potential heterogeneity within the population to understand differences in barriers to care-seeking that could then be addressed with targeted interventions. We used an unsupervised clustering method (PAM) to segment participants into seven groups based on their gender, occupation, problem drinking/smoking and knowledge of nearby facilities where free TB testing was available. Alternative cluster solutions from three to ten groups were also explored, and seven groups were found to yield the most actionable segmentation. We felt the seven-cluster solution offered segments that had good variability between the groups in terms of their likelihood to seek care. In addition, through discussions between analysts and programme leaders, we determined that groups in the seven-cluster solution differed from one another based on actionable characteristics and would be easy to identify in terms of

**Table 2** Characteristics of the seven population segments identified

Segment	Retired/unemployed/ disabled	Home makers with nearby facility	Home makers without nearby facility	Employed women with nearby facility	Employed women without nearby facility	Employed men who do not smoke or drink	Employed men who problem-drink and smoke
Segment size (% of sample)	13	14	34	5	11	13	10
% male	83	0	3	0	0	100	100
% retired/unemployed/disabled	100	0	10	0	1	1	4
% in paid employment	0	0	0	100	100	91	91
% who smoke	27	0	0	0	1	15	88
% who problem-drink	11	0	0	0	1	15	63
% aware of nearby facility that provides tuberculosis testing	61	100	0	100	0	63	24
% who seek formal care	70	61	58	55	47	59	42

typing people for designing future interventions. We also found that the seven-cluster solution had the highest average silhouette width of all segmentation solutions we tried, an indicator of good clustering. Table 2 shows the characteristics of each segment on each of the variables used in the clustering, the percentage of the sample that belongs to each segment and the care-seeking rates for each segment.

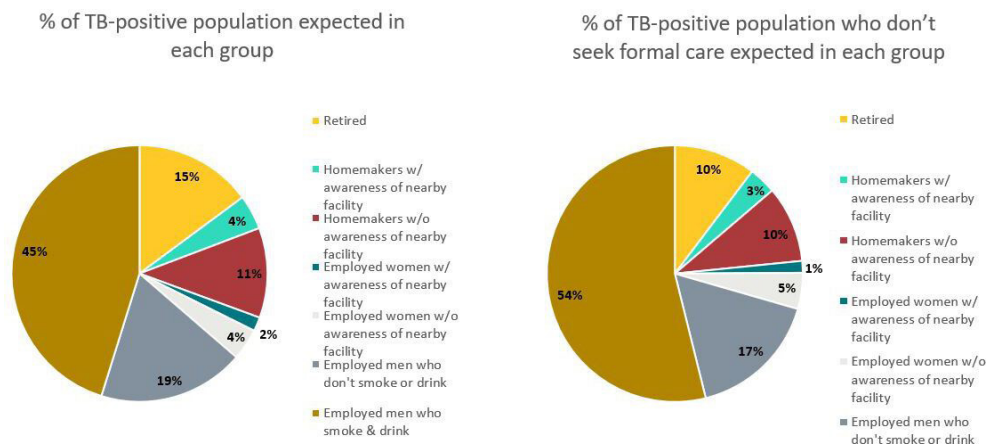
Segments were found to differ significantly in their care-seeking rates, in a pattern consistent with the relation between the segmenting variables and care-seeking seen in the predictive model. For example, the segment of individuals who were retired, unemployed or disabled sought care at the highest rates, followed by segments with home makers, followed by segments with individuals who had paid employment, consistent with the notion that those who have more flexible schedules are most likely to seek care. In addition, knowledge of a free TB facility nearby seems to be an important differentiator for care-seeking among employed individuals, but not home makers, further suggesting that time flexibility may be an important barrier to care-seeking; among those with less flexibility, knowing they can quickly and easily seek care affects care-seeking, but less so for those who have more time flexibility. Finally, consistent with the findings of the predictive analysis, the group most vulnerable for low care-seeking is men who both smoke and problem-drink. This effect can be seen clearly when we reweighted our sample population based on TB prevalence rates in Chennai by age, gender, smoking and drinking status according to a recent community-based prevalence study in the city of Chennai (see figure 4, right panel). Among those in the city who have TB, it is likely that 45% of them belong to the smoking and drinking segment, and 54% of those who have TB but do not promptly seek care for their symptoms fall into this segment.

## DISCUSSION

In this study, we used a novel large-scale, prospective, community-based survey to identify people with TB-indicative symptoms in South India and reveal key barriers to delayed care-seeking. We then identified a subpopulation of employed men who problem-drink and smoke, who are particularly at risk of TB and unlikely to seek care. While they comprise only 10% of our sample, we model that they may account for 54% of all TB-diagnosed individuals in Chennai who do not promptly seek care, making them a prime target for targeted interventions to increase care-seeking behaviour.

Characteristics around occupation and health behaviours emerged as strong overall predictors of delayed care-seeking. In particular, problem drinking and smoking were both strongly negatively associated with seeking care for TB symptoms. We also found large differences in care-seeking behaviour by occupation, with individuals who are not in the labour force (home makers and the retired/unemployed/disabled) being





**Figure 4** Expected prevalence of segments among the TB-diagnosed and the TB-diagnosed non-care-seeking population in Chennai, after ‘raking’ the sample (see the Methods section). TB, tuberculosis.

more likely to seek care than those who are (professionals and labourers), a pattern that was not yet seen in other studies. This finding suggests that one major barrier to seeking care may be a lack of time and convenience, rather than a lack of money, with those with greater time flexibility showing a greater propensity to seek care. Broadly, these findings tie in with other studies in South India, where problem drinking, smoking and higher incomes, but not disease knowledge, were associated with care-seeking delays.<sup>15 16</sup> That employment could play a role in care-seeking has also been indicated by a previous study in South India, where ‘pressure of work’ was associated with delays<sup>19</sup>; another study in Chennai had found that 31% of employed patients with TB, versus just 24% of non-employed patients, had taken action more than 30 days after symptom onset.<sup>15</sup> Our results may not all generalise to other geographies; for example, studies in several African countries<sup>10 11 13</sup> and China<sup>12</sup> did find that TB disease knowledge predicted care-seeking delays.

Surprisingly, two of the most classic TB symptoms, having a cough for 7 days and having blood in sputum, did not significantly increase the odds of care-seeking. Two other symptoms, having a fever for 7 days or experiencing unintended weight loss, did significantly increase the odds of care-seeking. Having symptoms for longer prompted individuals to seek care, but having multiple recurrences of symptoms decreased the likelihood of care-seeking.

We also uncovered a set of beliefs related to care-seeking that had not previously been examined. In particular, when people believed their community had different attitudes about where it would be best to seek care than they did, it suppressed care-seeking, with individuals being less likely to go to either their own preferred location or the location preferred by their community. In addition, beliefs about TB risk and prevalence were also linked to care-seeking behaviour, with individuals who believed that they were more at risk for TB being more likely to seek care and those who believed that TB was more common within their community being less likely to seek care, perhaps due to a sense of normalisation of the illness.

At present, many campaigns are informational, but in our study mere awareness of TB does not predict care-seeking. These findings therefore offer the opportunity of novel types of interventions addressing specific beliefs through a mix of educational, risk perception-targeting and social norms-based campaigns.

Our study also has limitations. For example, the two-sample populations are likely to be rather different on many dimensions, but needed to be merged for analytic power even with the large number of households screened. Further research should also be powered enough to stratify the sample for subgroup analyses. For example, care-seeking drivers could be evaluated relative to where people eventually seek care, as more than half of Indian patients do so in the formal and informal private sector.<sup>26</sup> Despite our efforts to knock on doors on weekends and in evenings, our sample likely skews women. Because TB is more prevalent in men, and in many geographies men are less likely to seek care and/or be diagnosed, obtaining insight into the ‘missing men’ would be important in future studies.<sup>31</sup> In our study, we also find that the segment at highest risk for not seeking prompt care are employed men who problem-drink and smoke, which are male-dominant behaviours in South India.<sup>32</sup> Targeted recruiting strategies, for example at workplaces, might facilitate enrolling men in future research. Finally, a community-based study enrolling only a small subset of the screened population can be expensive. More generally, initial care-seeking is just one part of the care pathway. Standardised patient studies in China, India, Kenya and South Africa show that people presenting with the classic TB symptoms receive a wide range of often inadequate care, including inappropriate diagnostic tests or unnecessary medication. To achieve swift TB detection and treatment, addressing provider behaviour must therefore be part of the research and intervention package.<sup>33</sup>

One major objective of this study was to identify segments at particularly high risk for delayed care-seeking, as well as their underlying drivers, so that these populations could potentially be targeted for

case-finding, messaging and other interventions designed to increase care-seeking behaviour. For this clustering, we focused mainly on characteristics related to occupation and smoking and drinking status: these were both highly related to care-seeking behaviour and are characteristics that are relatively easy to identify within the population and thus target on.

This approach was successful in identifying groups that differed substantially in their risky low care-seeking behaviour, including populations that were not initially obvious from the predictive analysis. For example, the segmentation revealed that individuals who were retired, unemployed or disabled were significantly more likely to seek care than nearly all other population segments. Segmentation also more clearly disaggregated the data, showing the large difference in care-seeking between employed men who do and do not problem-drink and smoke, and the also sizeable difference in care-seeking between home makers and employed women. Employed women who knew of a nearby facility that tested TB were more likely to seek care than employed women who did not. This pattern was not present in home makers (where those without that knowledge were similarly likely to seek care), further suggesting that having the time to visit a facility may be a major barrier to care-seeking.

The segmentation analysis also revealed that the population segment least likely to seek prompt care for their symptoms was employed men who problem-drink and smoke. This was a particularly worrisome segment to engage in low care-seeking behaviour, as being male and tobacco and alcohol use are all risk factors for TB diagnosis.<sup>21–25</sup> In other words, among all our TB presumptives, this group both was the most likely to actually have TB and the least likely to seek prompt care for it. Indeed, by reweighting our sample to match the characteristics of the TB-diagnosed population in Chennai, we found that 45% of the TB-positive population and 54% of the TB-diagnosed and non-care-seeking population might fall into this segment, even though they only comprised 10% of our TB-presumptive population. Thus, targeting this relatively small portion of the population could have a disproportionate impact on driving care-seeking among those who likely most need it. With limited resources, rather than trying to deploy interventions to the whole population, targeted efforts should be directed to this small but high-risk group. These efforts could greatly increase the prompt identification and treatment of TB cases.

Segmentation has been used successfully in other contexts to identify and target high-risk populations,<sup>29</sup> and we believe it can be effective in the TB context as well. Here, we identified a priority group that is not reached well by existing efforts. This segmentation requires knowledge about only a handful of easy-to-measure items, and clear guidance on finding the target population can be given. Potential hotspots for identification could entail targeting workplaces, screening in evenings and weekends (in data collection, men were less

likely to be home during the day), and targeting places frequented by alcohol and tobacco users. Targeted interventions to address care-seeking in this priority group will require creative solutions. Key barriers around time and convenience of care-seeking should be addressed, and interventions might be merged with existing efforts to address general low health-seeking behaviours. Further qualitative research with individuals in this population would generate a deeper understanding of the barriers this group sees to seeking care, providing information to fine-tune future interventions.

## CONCLUSION

In this study, we conducted a novel large-scale, prospective, community-based survey in people with TB-indicative symptoms. We surveyed participants about their care-seeking behaviour, as well as a comprehensive set of potential drivers of care-seeking behaviour, including demographics, knowledge, beliefs, health behaviours and social support. Through a combination of multivariate analysis and clustering, we identified characteristics associated with a failure to promptly seek care and segments of the population at particularly high or low risk for seeking care. We found that smoking and alcohol use were both negatively associated with care-seeking, as was employment status. Moreover, we found that a segment of employed men who problem-drink and smoke were both less likely to seek care and more likely to have TB than others. This group may comprise more than half of the individuals with TB in Chennai who do not promptly seek care. Interventions for this priority group should target barriers around time and convenience, rather than awareness or cost. The study's findings suggest that finding ways to effectively target priority populations and shift their care-seeking behaviour could have a disproportionate impact on the spread of TB throughout the city of Chennai, and that the methods introduced here can be applied to identify promising interventions in other contexts.

## Author affiliations

<sup>1</sup>Surgo Foundation, Washington, DC, USA

<sup>2</sup>Department of Social and Behavioral Research, National Institute for Research in Tuberculosis, Chennai, Tamil Nadu, India

<sup>3</sup>Clinton Health Access Initiative, Chennai/Delhi/NYC, India/USA

<sup>4</sup>District Tuberculosis Centre, Greater Chennai Corporation, Chennai, India

<sup>5</sup>Public Health Department, Greater Chennai Corporation, Chennai, India

<sup>6</sup>Harvard University T H Chan School of Public Health, Boston, Massachusetts, USA

<sup>7</sup>Department of Global Health, University of Washington, Seattle, WA, United States

**Contributors** SKS, EE, RC, BET, MJ, JL, GN, PP and HK conceptualised and designed the study. GN, PP, RC, MJ and EE were involved in the acquisition of data. SH, PP, GN and GKC analysed the data. SKS, SH, EE, RC, BET, MJ, JL, GN, PP, HK, TM, SM, MJ, HD and GKC contributed to the interpretation of the data. SH and EE drafted the manuscript. All authors provided critical comments to the manuscript. All authors provided final approval for the version to be published. SKS acts as guarantor of this paper.

**Funding** This study was funded by Surgo Foundation.

**Competing interests** None declared.

**Patient and public involvement** Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

**Patient consent for publication** Not required.

**Ethics approval** Ethics approval was obtained from the Centre for Media Studies Institutional Review Board, New Delhi, India (IRB number: 00006230) and from the Advarra Institutional Review Board, Columbia, Maryland (IRB number: 00023824).

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** De-identified participant data are available upon request from SKS (semasgaier@surgofoundation.org, ORCID: <https://orcid.org/0000-0002-8311-2686>). Reuse is permitted on a case-by-case basis for academic purposes.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iD

Sema K Sgaier <http://orcid.org/0000-0002-8311-2686>

## REFERENCES

- World Health Organization. *Global tuberculosis report 2019*. Geneva, 2019.
- World Health Organization. *The end TB strategy*. Geneva, 2020: 105.
- Horton R. Offline: in defence of precision public health. *Lancet* 2018;392:1504.
- Temesgen Z, Cirillo DM, Raviglione MC. Precision medicine and public health interventions: tuberculosis as a model? *Lancet Public Health* 2019;4:e374.
- Vesga JF, Hallett TB, Reid MJA, *et al*. Assessing tuberculosis control priorities in high-burden settings: a modelling approach. *Lancet Glob Health* 2019;7:e585–95.
- Madebo T, Lindtjorn B. Delay in treatment of pulmonary tuberculosis: an analysis of symptom duration among Ethiopian patients. *MedGenMed* 1999;18:E6.
- Sreeramareddy CT, Qin ZZ, Satyanarayana S, *et al*. Delays in diagnosis and treatment of pulmonary tuberculosis in India: a systematic review. *Int J Tuberc Lung Dis* 2014;18:255–66.
- Engl E, Smittenaar P, Sgaier SK. Identifying population segments for effective intervention design and targeting using unsupervised machine learning: an end-to-end guide. *Gates Open Res* 2019;3:1503.
- Yirgu R, Lemessa F, Hirpa S, *et al*. Determinants of delayed care seeking for TB suggestive symptoms in Seru district, Oromiya region, Ethiopia: a community based unmatched case-control study. *BMC Infect Dis* 2017;17:292.
- Biya O, Gidado S, Abraham A, *et al*. Knowledge, care-seeking behavior, and factors associated with patient delay among newly-diagnosed pulmonary tuberculosis patients, federal Capital Territory, Nigeria, 2010. *Pan Afr Med J* 2014;18 Suppl 1:6.
- Awoke N, Dulo B, Wudneh F. Total delay in treatment of tuberculosis and associated factors among new pulmonary TB patients in selected health facilities of Gedeo zone, southern Ethiopia, 2017/18. *Interdiscip Perspect Infect Dis* 2019;2019:2154240
- Tong Y, Guan X, Hou S, *et al*. Determinants of health Care-Seeking delay among tuberculosis patients in rural area of central China. *Int J Environ Res Public Health* 2018;15. doi:10.3390/ijerph15091998. [Epub ahead of print: 13 Sep 2018].
- Asres M, Gedefaw M, Kahsay A, *et al*. Patients' delay in seeking health care for tuberculosis diagnosis in East Gojjam zone, Northwest Ethiopia. *Am J Trop Med Hyg* 2017;96:1071–5.
- Chanda-Kapata P, Kapata N, Masiye F, *et al*. Health seeking behaviour among individuals with presumptive tuberculosis in Zambia. *PLoS One* 2016;11:e0163975.
- Rajeswari R, Chandrasekaran V, Suhadev M, *et al*. Factors associated with patient and health system delays in the diagnosis of tuberculosis in South India. *Int J Tuberc Lung Dis* 2002;6:789–95.
- Van Ness SE, Chandra A, Sarkar S, *et al*. Predictors of delayed care seeking for tuberculosis in southern India: an observational study. *BMC Infect Dis* 2017;17:567.
- Bronner Murrison L, Ananthakrishnan R, Swaminathan A, *et al*. How do patients access the private sector in Chennai, India? an evaluation of delays in tuberculosis diagnosis. *Int J Tuberc Lung Dis* 2016;20:544–51.
- Samal J. Health seeking behaviour among tuberculosis patients in India: a systematic review. *J Clin Diagn Res* 2016;10:Le01–6.
- Charles N, Thomas B, Watson B, *et al*. Care seeking behavior of chest symptomatics: a community based study done in South India after the implementation of the RNTCP. *PLoS One* 2010;5. doi:10.1371/journal.pone.0012379. [Epub ahead of print: 20 Sep 2010].
- Cai J, Wang X, Ma A, *et al*. Factors associated with patient and provider delays for tuberculosis diagnosis and treatment in Asia: a systematic review and meta-analysis. *PLoS One* 2015;10:e0120088.
- Dhanaraj B, Papanna MK, Adinarayanan S, *et al*. Prevalence and risk factors for adult pulmonary tuberculosis in a metropolitan city of South India. *PLoS One* 2015;10:e0124260.
- Kolappan C, Gopi PG. Tobacco smoking and pulmonary tuberculosis. *Thorax* 2002;57:964–6.
- Kolappan C, Gopi PG, Subramani R, *et al*. Selected biological and behavioural risk factors associated with pulmonary tuberculosis. *Int J Tuberc Lung Dis* 2007;11:999–1003.
- Rehm J, Samokhvalov AV, Neuman MG, *et al*. The association between alcohol use, alcohol use disorders and tuberculosis (TB). A systematic review. *BMC Public Health* 2009;9:450.
- Lönnroth K, Williams BG, Stadlin S, *et al*. Alcohol use as a risk factor for tuberculosis - a systematic review. *BMC Public Health* 2008;8:289.
- Mistry N, Rangan S, Dholakia Y, *et al*. Durations and delays in care seeking, diagnosis and treatment initiation in uncomplicated pulmonary tuberculosis patients in Mumbai, India. *PLoS One* 2016;11:e0152287.
- India TB Report. *New Delhi: central TB division, Directorate General of health services, Ministry of health and family welfare*, 2018.
- Engl E, Sgaier SK. Cubes: a practical toolkit to measure enablers and barriers to behavior for effective intervention design. *Gates Open Res* 2019;3:886.
- Sgaier SK, Eletskeya M, Engl E, *et al*. A case study for a psychographic-behavioral segmentation approach for targeted demand generation in voluntary medical male circumcision. *Elife* 2017;6:e25923.
- Sgaier SK, Engl E, Kretschmer S. *Time to scale psycho-behavioral segmentation in global development*. Stanford University, and Stanford center on philanthropy and civil society: Stanford social innovation review, 2018.
- Horton KC, MacPherson P, Houben RMGJ, *et al*. Sex differences in tuberculosis burden and notifications in low- and middle-income countries: a systematic review and meta-analysis. *PLoS Med* 2016;13:e1002119.
- Gajalakshmi V, Peto R, Smoking PR. Smoking, drinking and incident tuberculosis in rural India: population-based case-control study. *Int J Epidemiol* 2009;38:1018–25.
- Daniels B, Kwan A, Pai M, *et al*. Lessons on the quality of tuberculosis diagnosis from standardized patients in China, India, Kenya, and South Africa. *J Clin Tuberc Other Mycobact Dis* 2019;16:100109.