

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

Cohort profile: The Trivandrum non-alcoholic fatty liver disease (NAFLD) cohort

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2018-027244
Article Type:	Cohort profile
Date Submitted by the Author:	18-Oct-2018
Complete List of Authors:	Chalmers, Jane; NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham; Nottingham Digestive Diseases Centre Ban, Lu; NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham; Nottingham Digestive Diseases Centre Leena, Balakumaran; Population Health and Research Institute; Holistic Health and Research Institute Edwards, Kim; Orthopaedics, Trauma and Sports Medicine, School of Medicine, Faculty of Medicine and Health Sciences, University of Nottingham Grove, Jane; NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham; Nottingham Digestive Diseases Centre Aithal, Guruprasad; NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham; Nottingham Digestive Diseases Centre Shenoy, Thri vikrama; Population Health and Research Institute; Sree Gokulam Medical College and Research Foundation
Keywords:	Hepatology < INTERNAL MEDICINE, EPIDEMIOLOGY, Hepatobiliary disease < GASTROENTEROLOGY

SCHOLARONE™
Manuscripts

COHORT PROFILE: THE TRIVANDRUM NON-ALCOHOLIC FATTY LIVER DISEASE (NAFLD) COHORT

J. Chalmers^{1,2}, L. Ban^{1,2}, K.B. Leena^{3,4}, K.L. Edwards⁵, J.I. Grove^{1,2}, G.P Aithal^{1,2**}, K.T Shenoy^{3,6**}

Corresponding author

Professor Guruprasad Aithal

NIHR Nottingham Biomedical Research Centre, W/E 1418, E Floor, West Block, Queen's Medical Centre Campus, Derby Road, Nottingham, NG7 2UH, UK.

Email: guru.aithal@nottingham.ac.uk

Phone number: (+44) 01158231149

Co-authors

Dr Jane Chalmers

Dr Lu Ban

Dr K Balakumaran Leena

Dr Kimberley Edwards

Dr Jane Grove

Prof K Thirivikrama Shenoy

Author affiliations

¹ NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham, UK

² Nottingham Digestive Diseases Centre, Nottingham, UK

³ Population Health and Research Institute (PHRI), Trivandrum, India

⁴ Holistic Health and Research Institute, Trivandrum, India

⁵ Orthopaedics, Trauma and Sports Medicine, School of Medicine, Faculty of Medicine and Health Sciences, University of Nottingham, UK

⁶ Department of Gastroenterology, Sree Gokulam Medical College and Research Foundation, Trivandrum, India

** Joint senior author

Word count: 2543

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ABSTRACT

Purpose

The Trivandrum NAFLD cohort is a population-based study designed to examine the interaction between genetic and lifestyle factors and their association with increased risk of NAFLD within the Indian population.

Participants

Between 2013 and 2016, 2222 participants were recruited to this cohort through multi-stage cluster sampling across the whole population of Trivandrum – a district within the state of Kerala, South India. Data was collected from all inhabitants of randomly-selected households over the age of 25.

Findings to date

Full baseline clinical and pathological data was collected from 2158 participants. This included detailed demographic profiles, anthropometric measures and lifestyle data (food frequency, physical activity and anxiety and depression questionnaires). Biochemical profile and ultrasound assessment of the liver were performed and whole blood aliquots were collected for DNA analysis.

The NAFLD prevalence within this population was 49.8% which is significantly higher than the global pooled prevalence of 25%. This highlights the importance of robust, prospective studies like this to enable collection of longitudinal data on risk factors, disease progression and to facilitate future interventional studies.

Future plans

The complete analysis of data collected from this cohort will give valuable insights into the interaction of the phenotypic and genotypic profiles that result in such a dramatic increased risk of NAFLD within the Indian population. The cohort will also form the basis of future lifestyle interventional studies, aimed at improving liver and metabolic health.

ARTICLE SUMMARY

Strengths and limitations of this study

- This large population-based cohort is designed to examine the interaction between the genetic and lifestyle factors which contribute to increased NAFLD risk in India.
- Though limited to the district of Trivandrum, population-level sampling to include participants from both urban and rural domiciles make it one of the largest cohort studies of its kind.
- Extensive baseline data includes detailed cohort phenotyping (including food frequency and physical activity data) as well as biochemical profiles and whole blood aliquots for DNA analysis.
- Low rates of migration and robust cohort infrastructure will enable comprehensive longitudinal follow-up and form the basis for future interventional studies.

INTRODUCTION

Rising levels of obesity, type 2 diabetes (T2DM) and the metabolic syndrome (MetS) are a global health concern. The risk of T2DM is increasing worldwide, particularly in Indian populations where the risk is higher in both native, and migrant populations compared to other ethnicities,[1] and at a lower body mass index (BMI).[2]

One of the consequences of this metabolic pandemic, is increasing prevalence of non-alcoholic fatty liver disease (NAFLD) which is thought to be the hepatic manifestation of MetS. NAFLD encompasses a spectrum of disease from simple steatosis, non-alcoholic steatohepatitis, through increasing stages of fibrosis to cirrhosis – which may result in liver decompensation, hepatocellular carcinoma and death. Individuals with NAFLD, particularly those with MetS, have higher overall mortality, cardiovascular mortality and liver-related mortality compared to the general population.[3] NAFLD is particularly prevalent in Western populations,[4] but with the westernisation of Asia-Pacific culture, we are seeing increasing levels of obesity, MetS and NAFLD in this group.[5] The rise in MetS and its complications is variable according to region and degree of urbanisation – being induced by a variety of lifestyle factors such as diet and exercise habits.[6]

A recent meta-analysis showed that the pooled global prevalence of NAFLD is 25%.[7] The issue with this generalisation is that it is mainly based on large studies from the US, which include small numbers of people of Indian ethnicity. The documented prevalence of NAFLD also varies depending on the diagnostic tool used. In low-middle income countries (like India) it is difficult to perform large population-based studies, with those that are published being biased towards urban, tertiary centres where ultrasound and biopsy are more readily available.[8] Evidence has shown that within these urban areas – often of higher socioeconomic class – there are higher rates of NAFLD and as such these studies are likely to give an overestimate of NAFLD prevalence in India.[9]

The creation of the Trivandrum NAFLD cohort was funded and supported by the NIHR Nottingham Digestive Diseases Biomedical Research Unit, Nottingham University Hospitals NHS Trust and the University Of Nottingham, Nottingham UK with data collection and management run through the Population Health and Research Institute (PHRI), Trivandrum, Kerala, India. PHRI have specific expertise in population-based studies, having established and run the Trivandrum Tobacco Study (a study of tobacco habits of 0.4 million men in Trivandrum), which has undergone five yearly follow up since 2002 (data can be found through <https://www.ctsu.ox.ac.uk/research/the-trivandrum-tobacco-study>). The Trivandrum NAFLD cohort was created to examine how the interaction between genetics and lifestyle factors result in increased risk of NAFLD within this population. The cohort will combat the paucity of population-based studies in India and enable accurate estimation of NAFLD prevalence within an Indian population as a whole (including both urban and rural domiciles). It will also allow the development of a nested case-control study to look at the impact of different variables on NAFLD risk. Through large, population-based sampling of a relatively static populace, this cohort will also facilitate further disease-specific studies (including interventional studies), genetic studies and has the capability of generating longitudinal follow-up data.

COHORT DESCRIPTION

Cohort creation

Trivandrum (Thiruvananthapuram) is the southern-most district in the state of Kerala, located on the south-western coast of India (Fig. 1). At the time of the Indian census in 2011, it had a population of 3.3 million divided into urban and rural domiciles (urban 54%, rural 46% - census data captured through

https://www.census2011.co.in/). The Trivandrum NAFLD cohort was created between February 2013 and July 2016, through population-based sampling of all inhabitants over the age of 25 years.

Figure 1.

As described, Trivandrum has both urban and rural districts. The rural districts are divided into 12 blocks, which are themselves formed of 78 panchayats (originally denoting a village council). Panchayats are then subdivided further into wards (originally denoting a local authority area – usually indicating a neighbourhood). The urban districts are divided into wards only.

The enrolment of study participants was done through un-weighted multi-stage cluster sampling. Using random number tables, three out of 12 rural blocks were selected (made up of 22 panchayats), four rural panchayats were selected out of 22 (made up of 71 wards), and finally 40 out of 71 wards were selected. Using the same technique seven out of 81 urban wards were selected (Figure 2.).

Figure 2.

Using the Trivandrum electoral roll from 2011, households within these wards were grouped into clusters of seven. Using random number tables 112 of 2,420 rural clusters and 128 of 2,631 urban clusters were identified for sampling. Each household within the chosen clusters were sampled, and through house-to-house survey by local social workers all eligible inhabitants were invited to enrol. For some households this meant repeated visits to ensure all those who were eligible (anyone over the age of 25 years) were given the opportunity to participate. At the time of recruitment, within the household, baseline demographic, anthropometric and past medical history data were collected.

Originally the aim was to recruit around 2000 cases of NAFLD. Based on an estimated NAFLD prevalence of 31.8% from a previous pilot study,[10] this would mean recruiting a total cohort of 10,000 participants. Unfortunately due to time constraints, cohort recruitment at household-level was ceased at around 2000 participants (Figure 3), which based on estimated prevalence would result in 840 NAFLD cases.

Figure 3.

In order to collect pathological data (including bloods and abdominal ultrasound) study camps were held within the area local to the selected wards and those who had been recruited were invited to attend. Of the original 2222 population sample 2161 participants attended the camps – demonstrating a high attendance rate with only 3% drop-out. Three participants did not have an ultrasound so were excluded leaving the total number of participants as 2158. Analysis of the demographic/anthropometric data of those who did and did not have an ultrasound shows that a higher proportion of men (61% vs 40.6% $p<0.001$) and a higher proportion of people within the rural setting (73.4% vs 46.3% $p<0.001$) did not attend. This is likely due to the difficulty in getting to the rural camps (which covered a larger geographical area than the urban camps), and that men were often limited by their need to work.

Data collection

Clinical data were collected through house-to-house survey and pathological data were collected through the local study camps. This is outlined below:

Clinical data

Potential participants were approached and consented to the cohort through house-to-house survey by social workers within their wards. Initial clinical data were collected at the time of consent within the

participant's home. Baseline demographic data included date of birth, gender, level of education, work status, marital status and religion. Information on past medical history was also documented along with history of tobacco smoking/chewing and alcohol intake. Anthropometric data collected included height (cm), weight (kg), waist circumference (cm) and blood pressure (mmHg). Participants also undertook a food frequency questionnaire (FFQ), a global physical activity questionnaire (GPAQ) and hospital anxiety and depression score (HADS).

Pathological data

Participants who attended the study camps had blood taken for assessment of liver function, platelet count, fasting glucose, insulin and lipids. Additional whole blood aliquots were taken for DNA extraction. Participants then underwent abdominal ultrasonography to assess presence and grade of fatty infiltration (Grade 1: Liver echogenicity is increased. Grade 2: Echogenic liver obscures the echogenic walls of portal vein branches. Grade 3: Echogenic liver obscures the diaphragmatic outline).[11] Those who had evidence of NAFLD subsequently underwent Transient Elastography (Fibroscan®) as a surrogate marker to stage degree of fibrosis (≥ 8 kPa indicating presence of significant fibrosis).

Baseline characteristics

Demographic data from the population sample (n=2222) and NAFLD cohort (n=2158) compared to the census data of Trivandrum in 2011 is outlined in Table 1. The NAFLD cohort has a larger proportion of women to men and a larger proportion of Hindu population than the national average. Importantly, the proportion of rural and urban population within the cohort matches the census, implying adequate population-level sampling.

Demography	Census (n=3,301,427)	Population sample (n=2222)	<i>p value</i>	NAFLD Cohort (n=2158)	<i>p value</i>
Gender					
- Male n (%)	1,581,678 (47.9)	1307 (58.8)	*	876 (40.6)	*
- Female n (%)	1,719,749 (52.1)	915 (41.2)		1282 (59.4)	
Domicile					
- Rural n (%)	1,529,831 (46.3)	1046 (47.1)	<i>p=0.49</i>	999 (46.3)	<i>p=1.04</i>
- Urban n (%)	1,771,596 (53.7)	1176 (52.9)		1159 (53.7)	
Literacy (%)	93.0	98.0	<i>p=0.09</i>	98.0	<i>p=0.09</i>
Religion					
- Hindu n (%)	2,194,057 (66.5)	1821 (81.9)	*	1765 (81.8)	*
- Muslim n (%)	452,915 (13.7)	170 (7.7)	*	167 (7.7)	*
- Christian n (%)	630,573 (19.1)	231 (10.4)	*	226 (10.5)	*
- Other n (%)	23,882 (<1)	0		0	

Table 1. Comparison of the population sample demographics to the Trivandrum population census and comparison of the NAFLD cohort demographics to the population sample. (* $p < 0.001$)

FINDINGS TO DATE

A total of 2158 participants were included in the final cohort. The prevalence of NAFLD within this population-based cohort is 49.8% (n=1075).

Within this cohort, a nested NAFLD case-control study has been done to look at the difference in aetiological factors between those with NAFLD and those without. In order to do this, data from the original NAFLD cohort (2158) was cleaned to remove other causes of non-NAFLD related liver disease such as alcohol intake >21 units per week and alternate causes of an echo-bright liver (such as history of viral hepatitis or other liver disease), leaving a sub-cohort of 2089 (NAFLD 49.8%). Baseline demographic data from this sub-cohort is presented in Table 2.

Demography	NAFLD n = 1040	Control n = 1049	P value
Age mean (\pm SD)	48.34 (10.74)	45.98 (12.31)	<0.001
Gender			
- Male n (%)	473 (45.5)	339 (32.3)	<0.001
- Female n (%)	567 (54.5)	710 (67.7)	
Domicile			
- Rural n (%)	417 (40.1)	543 (51.8)	<0.001
- Urban n (%)	623 (59.9)	506 (48.2)	
Literacy (%)	97.9	97.9	0.978
BMI mean (\pm SD)	26.90 (4.16)	24.30 (4.08)	<0.001
- Underweight n (%) <18.5kg/m ²	4 (0.4)	68 (6.5)	
- Normal weight n (%) 18.5-22.9kg/m ²	165 (15.9)	336 (32)	
- Overweight n (%) 23-24.9kg/m ²	195 (18.8)	222 (21.2)	
- Obese n (%) >25kg/m ²	675 (64.9)	423 (40.3)	
Diabetes (%)	67.6	47.1	<0.001

Table 2. Nested NAFLD case-control demographics *p<0.001

Overall, females outnumbered males and constituted 1277 of the total 2089 participants. 55.2% of the urban population had NAFLD compared to 43.4% of the rural population (an unadjusted OR of 1.60, 1.35-1.91) – a finding which is consistent with the literature on increased NAFLD prevalence in areas of higher socioeconomic class.[12] As expected, mean BMI was higher in the NAFLD group (OR 1.79 if overweight, OR 3.25 if obese) as was rate of diabetes (OR 2.34, 1.95-2.81). This data is presented in Table 3. Under logistic regression, for every unit increase in BMI, the odds of having NAFLD increases by 1.17 (17%) (1.15, 1.20 p<0.001).

Variable	Unadjusted OR (95% CI, p value)	Adjusted OR (95% CI, p value)
Male sex	1.75 (1.46, 2.09 p<0.001)	2.04 (1.67, 2.47 p<0.001)
Urban domicile	1.60 (1.35, 1.91 p<0.001)	1.23 (1.02, 1.49 p=0.03)
BMI		
- Underweight <18.5kg/m ²	0.12 (0.04, 0.34 p<0.001)	0.14 (0.05, 0.39 p<0.001)
- Overweight 23-24.9kg/m ²	1.79 (1.36, 2.34 p<0.001)	1.72 (1.31, 2.28 p<0.001)
- Obese	3.25 (2.58, 4.10 p<0.001)	3.34 (2.63, 4.23 p<0.001)

>25kg/m ²		
DM	2.34 (1.95, 2.81 p<0.001)	2.03 (1.68, 2.45 p<0.001)

Table 3. NAFLD odds ratios adjusting for sex, domicile, BMI category (with normal weight as baseline), DM. Variables were mutually adjusted for each other with baseline categories being those variables presented.

This data confirms commonality of NAFLD risk factors between India and the Western world. Increasing levels of obesity and the dramatic rise in diabetes prevalence in India [1] is reflected in the World Health Organisation (WHO) Global Burden of Disease study, which shows that hypertension and raised fasting blood glucose are the leading cause of non-communicable disease in India, unlike other low-middle income countries where childhood malnutrition and unsafe water predominate.[13] Interestingly, presence of obesity conveys the strongest risk of NAFLD within this group (above diabetes) which highlights the “Asian paradox” – whereby cardiovascular risk increases with BMI in Asians, as in other ethnicities, but at a lower BMI. This data supports the significant impact of BMI on risk of disease in this group and the reclassification of BMI cut-offs that was implemented in 2000.[2]

Recruitment to the Trivandrum NAFLD cohort was completed in July 2016. As yet, it has not undergone a complete follow up. The aim of the project is to collect longitudinal data on risk factors, disease progression and to make an assessment of proportional mortality rate using verbal autopsy – a validated instrument used to ascertain time and cause of death.[14] Further lifestyle data that was collected from the cohort is under review with the aim to look at the impact of diet and physical activity on risk of NAFLD, both in native Indians and on those who now live in the UK.

The cohort also forms the basis of a Global Challenges Research Fund recently awarded to the NIHR Nottingham BRC in collaboration with PHRI to develop lifestyle interventions for improved liver and metabolic health in South India.

STRENGTHS AND LIMITATIONS

The biggest strength of the Trivandrum NAFLD cohort is the fact that it was designed to collect detailed anthropometric, clinical and pathological data at a population-level, representing the full spectrum of demography of a large Indian state. Through population-level sampling we have enabled accurate estimation of NAFLD prevalence, which given its significant departure from the estimated global prevalence has filled an important gap in the knowledge-base. Through the utilisation of local social workers within the community, PHRI have created a cohort with excellent uptake and retention of participants, have methods in place to document rates of migration and accurate mortality data as well as having the infrastructure and man-power in place to facilitate longitudinal and interventional studies.

The main limitation of the study from a clinical perspective is the use of ultrasound as the diagnostic tool for NAFLD. Ultrasound has been shown to have a poor yield in the diagnosis of microvesicular fat within the liver, with an overall sensitivity of 43% and specificity of 73%.[15] This means that our NAFLD prevalence of 49.8% may in fact be an under-estimate. Liver biopsy is generally considered to be the gold standard for NAFLD diagnosis, but its invasive nature combined with its inherent risk of sampling error make it unsuitable for population-based studies. Despite this, the message that NAFLD prevalence is significantly higher in this Indian population compared to the global estimates is an important one. Finally, although representative of a large Indian population, the data collected from this cohort is not necessarily representative of the India as a whole – whose diverse population includes thousands of distinct and unique cultures.

COLLABORATION

Both the NIHR Nottingham Biomedical Research Centre, Gastrointestinal and Liver disorder theme and Population Health and Research Institute (PHRI) welcome interest and collaboration with national and international colleagues. For further information please contact the corresponding author.

PATIENT INVOLVEMENT

Participants of the Trivandrum NFALD cohort were not involved in setting the research question or the outcome measures of this project, but during data collection the patient and public involvement (PPI) team from the NIHR Nottingham Biomedical Research Centre ran a 3 day workshop alongside PHRI in Trivandrum. During this workshop, participants were consulted on their opinions of having lifestyle information collected, and the best way to do this. This qualitative feedback has helped shape the way we approach data collection from a cultural standpoint in this ethnic group.

CONTRIBUTORS

KTS and GA contributed the original concept and design of the study. KTS, KBL and JG were involved in data collection and sample management. JC and LB did the data analysis and interpretation whilst JC, LB and KE drafted the work. All the authors have critically revised this article and approved the final version to be published.

FUNDING

The Nottingham Biomedical Research Centre (digestive diseases theme) supports PHRI and is funded by the National Institute for Health Research (NIHR).

COMPETING INTERESTS

None declared

ETHICAL APPROVAL

Ethical approval for the study was granted by the Sree Gokulam Medical College and Research Foundation, Venjaramoodu, Trivandrum, India ethics committee. Written informed consent was obtained from all participants at the time of recruitment.

DATA SHARING STATEMENT

Researchers may request access to data and biomaterial by contacting the corresponding author

1. Whiting, D.R., et al., *IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030*. Diabetes Res Clin Pract, 2011. **94**(3): p. 311-21.
2. WHO, E.C., *Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies*. Lancet, 2004. **363**(9403): p. 157-63.
3. Ong, J.P., A. Pitts, and Z.M. Younossi, *Increased overall mortality and liver-related mortality in non-alcoholic fatty liver disease*. J Hepatol, 2008. **49**(4): p. 608-12.
4. Farrell, G.C. and C.Z. Larter, *Nonalcoholic fatty liver disease: from steatosis to cirrhosis*. Hepatology, 2006. **43**(2 Suppl 1): p. S99-s112.
5. Mohan, V., et al., *Prevalence of non-alcoholic fatty liver disease in urban south Indians in relation to different grades of glucose intolerance and metabolic syndrome*. Diabetes Res Clin Pract, 2009. **84**(1): p. 84-91.
6. Mohan, V., et al., *Intra-urban differences in the prevalence of the metabolic syndrome in southern India -- the Chennai Urban Population Study (CUPS No. 4)*. Diabet Med, 2001. **18**(4): p. 280-7.
7. Younossi, Z.M., et al., *Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes*. Hepatology, 2016. **64**(1): p. 73-84.
8. Prashanth, M., et al., *Prevalence of nonalcoholic fatty liver disease in patients with type 2 diabetes mellitus*. J Assoc Physicians India, 2009. **57**: p. 205-10.
9. Singh, S., G.N. Kuftinec, and S. Sarkar, *Non-alcoholic Fatty Liver Disease in South Asians: A Review of the Literature*. Journal of Clinical and Translational Hepatology, 2017. **5**(1): p. 76-81.
10. K T Shenoy, e.a. *Prevalence and risk factors for non-alcoholic fatty liver in a transition society with high incidence of diabetes mellitus*. in *The 15th Annual Conference of the INASAL*. 2007. Vellore.
11. Saadeh, S., et al., *The utility of radiological imaging in nonalcoholic fatty liver disease*. Gastroenterology, 2002. **123**(3): p. 745-50.
12. Petersen, K.F., et al., *Increased prevalence of insulin resistance and nonalcoholic fatty liver disease in Asian-Indian men*. Proc Natl Acad Sci U S A, 2006. **103**(48): p. 18273-7.
13. Forouzanfar, M.H., et al., *Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015*. The Lancet, 2016. **388**(10053): p. 1659-1724.
14. Jha, P., et al., *Prospective Study of One Million Deaths in India: Rationale, Design, and Validation Results*. PLOS Medicine, 2005. **3**(2): p. e18.
15. Dasarathy, S., et al., *Validity of real time ultrasound in the diagnosis of hepatic steatosis: a prospective study*. J Hepatol, 2009. **51**(6): p. 1061-7.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 1. Map of Kerala

Figure 2. Trivandrum NAFLD cohort sampling framework

Figure 3. Trivandrum NAFLD cohort recruitment diagram

For peer review only



Figure 1. Map of Kerala

282x463mm (300 x 300 DPI)

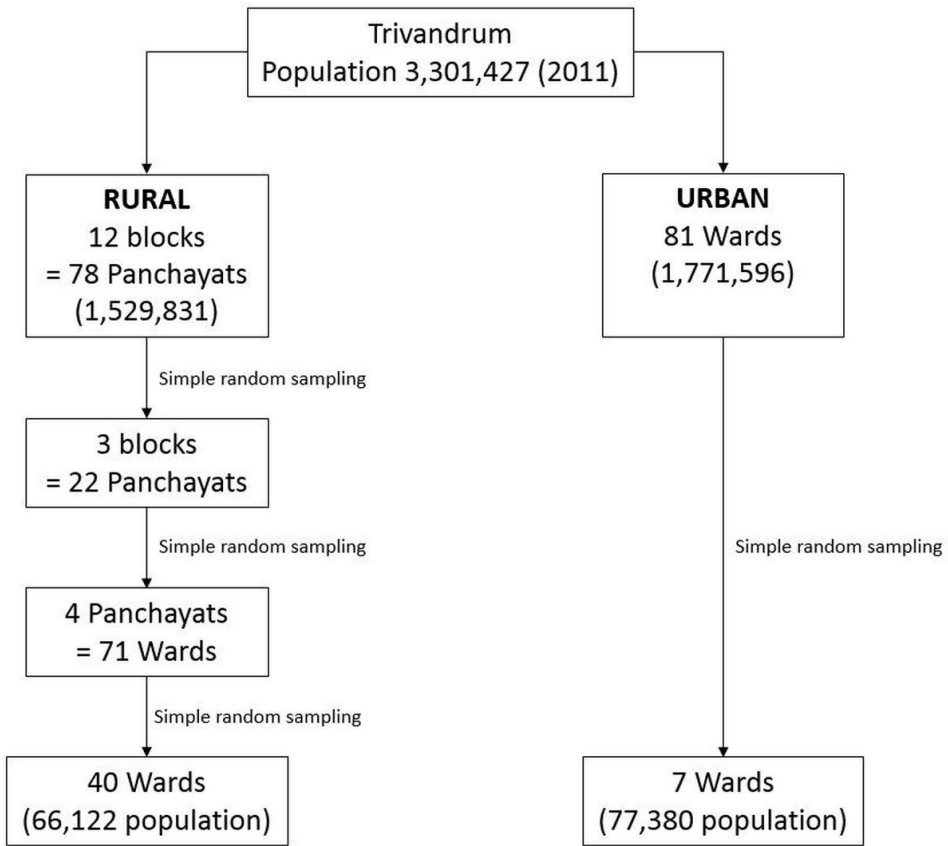


Figure 2. Trivandrum NAFLD cohort sampling framework

106x90mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

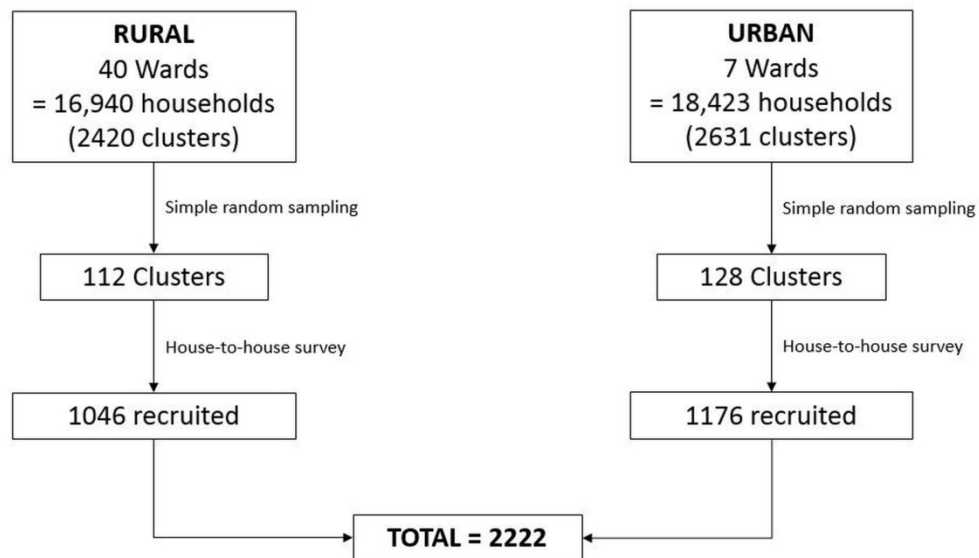


Figure 3. Trivandrum NAFLD cohort recruitment diagram

155x90mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

BMJ Open

Cohort profile: The Trivandrum non-alcoholic fatty liver disease (NAFLD) cohort

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2018-027244.R1
Article Type:	Cohort profile
Date Submitted by the Author:	19-Feb-2019
Complete List of Authors:	Chalmers, Jane; NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham; Nottingham Digestive Diseases Centre Ban, Lu; NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham; Nottingham Digestive Diseases Centre Leena, Balakumaran; Population Health and Research Institute; Holistic Health and Research Institute Edwards, Kim; Orthopaedics, Trauma and Sports Medicine, School of Medicine, Faculty of Medicine and Health Sciences, University of Nottingham Grove, Jane; NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham; Nottingham Digestive Diseases Centre Aithal, Guruprasad; NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham; Nottingham Digestive Diseases Centre Shenoy, Thri vikrama; Population Health and Research Institute; Sree Gokulam Medical College and Research Foundation
Primary Subject Heading:	Gastroenterology and hepatology
Secondary Subject Heading:	Global health
Keywords:	Hepatology < INTERNAL MEDICINE, EPIDEMIOLOGY, Hepatobiliary disease < GASTROENTEROLOGY

SCHOLARONE™
Manuscripts

COHORT PROFILE: THE TRIVANDRUM NON-ALCOHOLIC FATTY LIVER DISEASE (NAFLD) COHORT

J. Chalmers^{1,2}, L. Ban^{1,2}, K.B. Leena^{3,4}, K.L. Edwards⁵, J.I. Grove^{1,2}, G.P. Aithal^{1,2**}, K.T. Shenoy^{3,6**}

Corresponding author

Professor Guruprasad Aithal

NIHR Nottingham Biomedical Research Centre, W/E 1418, E Floor, West Block, Queen's Medical Centre Campus, Derby Road, Nottingham, NG7 2UH, UK.

Email: guru.aithal@nottingham.ac.uk

Phone number: (+44) 01158231149

Co-authors

Dr Jane Chalmers

Dr Lu Ban

Dr K Balakumaran Leena

Dr Kimberley Edwards

Dr Jane Grove

Prof K Thrikrama Shenoy

Author affiliations

¹ NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham, UK

² Nottingham Digestive Diseases Centre, Nottingham, UK

³ Population Health and Research Institute (PHRI), Trivandrum, India

⁴ Holistic Health and Research Institute, Trivandrum, India

⁵ Orthopaedics, Trauma and Sports Medicine, School of Medicine, Faculty of Medicine and Health Sciences, University of Nottingham, UK

⁶ Department of Gastroenterology, Sree Gokulam Medical College and Research Foundation, Trivandrum, India

** Joint senior author

Word count: 2931

ABSTRACT

Purpose

The Trivandrum NAFLD cohort is a population-based study designed to examine the interaction between genetic and lifestyle factors and their association with increased risk of NAFLD within the Indian population.

Participants

Between 2013 and 2016, 2222 participants were recruited to this cohort through multi-stage cluster sampling across the whole population of Trivandrum – a district within the state of Kerala, South India. Data was collected from all inhabitants of randomly-selected households over the age of 25.

Findings to date

Full baseline clinical and pathological data was collected from 2158 participants. This included detailed demographic profiles, anthropometric measures and lifestyle data (food frequency, physical activity and anxiety and depression questionnaires). Biochemical profile and ultrasound assessment of the liver were performed and whole blood aliquots were collected for DNA analysis.

The NAFLD prevalence within this population was 49.8% which is significantly higher than the global pooled prevalence of 25%. This highlights the importance of robust, prospective studies like this to enable collection of longitudinal data on risk factors, disease progression and to facilitate future interventional studies.

Future plans

The complete analysis of data collected from this cohort will give valuable insights into the interaction of the phenotypic and genotypic profiles that result in such a dramatic increased risk of NAFLD within the Indian population. The cohort will also form the basis of future lifestyle interventional studies, aimed at improving liver and metabolic health.

ARTICLE SUMMARY

Strengths and limitations of this study

- This large population-based cohort is designed to examine the interaction between the genetic and lifestyle factors which contribute to increased NAFLD risk in India.
- Though limited to the district of Trivandrum, population-level sampling to include participants from both urban and rural domiciles make it one of the largest cohort studies of its kind.
- Extensive baseline data includes detailed cohort phenotyping (including food frequency and physical activity data) as well as biochemical profiles and whole blood aliquots for DNA analysis.
- Low rates of migration and robust cohort infrastructure will enable comprehensive longitudinal follow-up and form the basis for future interventional studies.

INTRODUCTION

Rising levels of obesity, type 2 diabetes (T2DM) and the metabolic syndrome (MetS) are a global health concern. The risk of T2DM is increasing worldwide, particularly in Indian populations where the risk is higher in both native, and migrant populations compared to other ethnicities,[1] and at a lower body mass index (BMI).[2]

One of the consequences of this metabolic pandemic, is increasing prevalence of non-alcoholic fatty liver disease (NAFLD) which is thought to be the hepatic manifestation of MetS. NAFLD encompasses a spectrum of disease from simple steatosis, non-alcoholic steatohepatitis, through increasing stages of fibrosis to cirrhosis – which may result in liver decompensation, hepatocellular carcinoma and death. Individuals with NAFLD, particularly those with MetS, have higher overall mortality, cardiovascular mortality and liver-related mortality compared to the general population.[3] NAFLD is particularly prevalent in Western populations,[4] but with the westernisation of Asia-Pacific culture, we are seeing increasing levels of obesity, MetS and NAFLD in this group.[5] The rise in MetS and its complications is variable according to region and degree of urbanisation – being induced by a variety of lifestyle factors such as diet and exercise habits.[6]

A recent meta-analysis showed that the pooled global prevalence of NAFLD is 25%.[7] The issue with this generalisation is that it is mainly based on large studies from the US, which include small numbers of people of Indian ethnicity. The documented prevalence of NAFLD also varies depending on the diagnostic tool used. In low-middle income countries (like India) it is difficult to perform large population-based studies, with those that are published being biased towards urban, tertiary centres where ultrasound and biopsy are more readily available.[8] Evidence has shown that within these urban areas – often of higher socioeconomic class – there are higher rates of NAFLD and as such these studies are likely to give an overestimate of NAFLD prevalence in India.[9]

The creation of the Trivandrum NAFLD cohort was funded and supported by the NIHR Nottingham Digestive Diseases Biomedical Research Unit, Nottingham University Hospitals NHS Trust and the University Of Nottingham, Nottingham UK with data collection and management run through the Population Health and Research Institute (PHRI), Trivandrum, Kerala, India. PHRI have specific expertise in population-based studies, having established and run the Trivandrum Tobacco Study (a study of tobacco habits of 0.4 million men in Trivandrum), which has undergone five yearly follow up since 2002 (data can be found through <https://www.ctsu.ox.ac.uk/research/the-trivandrum-tobacco-study>). The Trivandrum NAFLD cohort was created to examine how the interaction between genetics and lifestyle factors result in increased risk of NAFLD within this population. The cohort will combat the paucity of population-based studies in India and enable accurate estimation of NAFLD prevalence within an Indian population as a whole (including both urban and rural domiciles). It will also allow the development of a nested case-control study to look at the impact of different variables on NAFLD risk. Through large, population-based sampling of a relatively static populace, this cohort will also facilitate further disease-specific studies (including interventional studies), genetic studies and has the capability of generating longitudinal follow-up data.

COHORT DESCRIPTION

Cohort creation

Trivandrum (Thiruvananthapuram) is the southern-most district in the state of Kerala, located on the south-western coast of India (Fig. 1). At the time of the Indian census in 2011, it had a population of 3.3 million divided into urban and rural domiciles (urban 54%, rural 46% - census data captured through <https://www.census2011.co.in/>). The Trivandrum NAFLD cohort was created between February 2013 and July 2016, through population-based sampling of all inhabitants over the age of 25 years.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 1.

As described, Trivandrum has both urban and rural districts. As per the population census, the rural districts are divided into 12 blocks, which are themselves formed of 78 panchayats (originally denoting a village council). Panchayats are then subdivided further into wards (originally denoting a local authority area – usually indicating a neighbourhood). The urban districts are divided into wards only.

The enrolment of study participants was done through un-weighted multi-stage cluster sampling. Using random number tables, three out of 12 rural blocks were selected (made up of 22 panchayats), four rural panchayats were selected out of 22 (made up of 71 wards), and finally 40 out of 71 wards were selected. Using the same technique seven out of 81 urban wards were selected (Figure 2.).

Figure 2.

Using the Trivandrum electoral roll from 2011, households within these wards were grouped into clusters of seven. Using random number tables 112 of 2,420 rural clusters and 128 of 2,631 urban clusters were identified for sampling. Each household within the chosen clusters were sampled, and through house-to-house survey by local social workers all eligible inhabitants were invited to enrol. For some households this meant repeated visits to ensure all those who were eligible (anyone over the age of 25 years) were given the opportunity to participate. At the time of recruitment, within the household, baseline demographic, anthropometric and past medical history data were collected.

Originally the aim was to recruit around 2000 cases of NAFLD. Based on an estimated NAFLD prevalence of 31.8% from a previous pilot study,[10] this would mean recruiting a total cohort of 10,000 participants. Unfortunately due to time constraints, cohort recruitment at household-level was ceased at around 2000 participants (Figure 3), which based on estimated prevalence would result in 840 NAFLD cases.

Figure 3.

In order to collect pathological data (including bloods and abdominal ultrasound) study camps were held within the area local to the selected wards and those who had been recruited were invited to attend. Of the original 2222 population sample 2161 participants attended the camps – demonstrating a high attendance rate with only 3% drop-out. Three participants did not have an ultrasound so were excluded leaving the total number of participants as 2158. Analysis of the demographic/anthropometric data of those who did and did not have an ultrasound shows that a higher proportion of men (61% vs 40.6% $p < 0.001$) and a higher proportion of people within the rural setting (73.4% vs 46.3% $p < 0.001$) did not attend. This is likely due to the difficulty in getting to the rural camps (which covered a larger geographical area than the urban camps), and that men were often limited by their need to work.

Data collection

Clinical data were collected through house-to-house survey and pathological data were collected through the local study camps. This is outlined below:

Clinical data

Potential participants were approached and consented to the cohort through house-to-house survey by social workers within their wards. Initial clinical data were collected at the time of consent within the participant's home. Baseline demographic data included date of birth, gender, level of education, work status, marital status and religion. Information on past medical history was also documented along with history of tobacco smoking/chewing and alcohol intake. Anthropometric data collected during household visits included height measured in centimetres (cm) using a wall-

mounted vertical rule, weight measured in kilograms (kg) using mechanical standing-scales, waist circumference – taking an average of two measurements in cm using a non-expandable tape between the lower rib margin and iliac crest, and an average of three blood pressure readings in mmHg taken using a manual cuff in a sitting position. Participants undertook a regional food frequency questionnaire (FFQ), designed to incorporate variation in nutrient intake attributed to local recipes according to domicile, social class and religion. The validation of this FFQ is ongoing – utilising repeated 24-hour food diaries and blood/urine biomarkers. Participants also undertook the global physical activity questionnaire (GPAQ)[11] and Duke's anxiety and depression scale (HADS)[12].

Pathological data

Participants who attended the study camps had blood taken for assessment of liver function, platelet count, fasting glucose, insulin and lipids. Additional whole blood aliquots were taken for DNA extraction. Participants then underwent abdominal ultrasonography to assess presence and grade of fatty infiltration (Grade 1: Liver echogenicity is increased. Grade 2: Echogenic liver obscures the echogenic walls of portal vein branches. Grade 3: Echogenic liver obscures the diaphragmatic outline).[13] Those who had evidence of NAFLD subsequently underwent Transient Elastography (Fibroscan®) as a surrogate marker to stage degree of fibrosis (≥ 8 kPa indicating presence of significant fibrosis).

Statistical analyses

Categorical data are presented as numbers (percentage) and continuous data as mean (SD) for normally distributed data and medians (range) for non-parametric data. Univariate analysis to compare characteristics of the cohort with/without NAFLD was undertaken using Students t test (normal continuous), Mann-Whitney U test (non-normal continuous) and Chi squared test (categorical).

Baseline characteristics

Demographic data from the population sample (n=2222) and NAFLD cohort (n=2158) compared to the census data of Trivandrum in 2011 is outlined in Table 1. The NAFLD cohort has a larger proportion of women to men and a larger proportion of Hindu population than the national average. Importantly, the proportion of rural and urban population within the cohort matches the census, implying adequate population-level sampling.

Demography	Census (n=3,301,427)	Population sample (n=2222)	<i>p value</i>	NAFLD Cohort (n=2158)	<i>p value</i>
Gender					
- Male n (%)	1,581,678 (47.9)	915 (41.2)	*	876 (40.6)	*
- Female n (%)	1,719,749 (52.1)	1307 (58.8)		1282 (59.4)	
Domicile					
- Rural n (%)	1,529,831 (46.3)	1046 (47.1)	<i>p=0.49</i>	999 (46.3)	<i>p=1.04</i>
- Urban n (%)	1,771,596 (53.7)	1176 (52.9)		1159 (53.7)	
Literacy (%)	93.0	98.0	<i>p=0.09</i>	98.0	<i>p=0.09</i>
Religion					
- Hindu n (%)	2,194,057 (66.5)	1821 (81.9)	*	1765 (81.8)	*
- Muslim n (%)	452,915 (13.7)	170 (7.7)	*	167 (7.7)	*
- Christian n (%)	630,573 (19.1)	231 (10.4)	*	226 (10.5)	*
- Other n (%)	23,882 (<1)	0		0	

Table 1. Comparison of the population sample demographics to the Trivandrum population census and comparison of the NAFLD cohort demographics to the population sample. * p<0.001

FINDINGS TO DATE

A total of 2158 participants were included in the final cohort. The prevalence of NAFLD within this population-based cohort is 49.8% (n=1075).

Within this cohort, a nested NAFLD case-control study has been done to look at the difference in aetiological factors between those with NAFLD and those without. In order to do this, data from the original NAFLD cohort (2158) was cleaned to remove other causes of non-NAFLD related liver disease such as alcohol intake >21 units per week and alternate causes of an echo-bright liver (such as history of viral hepatitis or other liver disease), leaving a sub-cohort of 2089 (NAFLD 49.8%).

Baseline demographic data from this sub-cohort is presented in Table 2.

Demography		NAFLD n = 1040	Control n = 1049	P value
Age	Mean (SD)	48.34 (10.74)	45.98 (12.31)	*
Gender				
- Male	n (%)	473 (45.5)	339 (32.3)	*
- Female	n (%)	567 (54.5)	710 (67.7)	
Domicile				
- Rural	n (%)	417 (40.1)	543 (51.8)	*
- Urban	n (%)	623 (59.9)	506 (48.2)	
Literacy	n (%)	1018 (97.9)	1027 (97.9)	0.978
BMI	Mean (SD)	26.90 (4.16)	24.30 (4.08)	*
- Underweight <18.5kg/m ²	n (%)	4 (0.4)	68 (6.5)	
- Normal weight 18.5-22.9kg/m ²	n (%)	165 (15.9)	336 (32)	
- Overweight 23-24.9kg/m ²	n (%)	195 (18.8)	222 (21.2)	
- Obese >25kg/m ²	n (%)	675 (64.9)	423 (40.3)	
Diabetes	n %	355 (34.13)	195 (18.59)	*
HOMA-IR	Median (IQR)	7.9 (7.48)	4.7 (4.5)	*
Metabolic Syndrome	n (%)	417 (40.1)	240 (22.88)	*
Cholesterol	Mean (SD)	219.19 (44.05)	212.26 (44.34)	*
HDL	Mean (SD)	47.65 (11.25)	50.35 (11.83)	*
Triglycerides	Median (IQR)	133 (85)	94.5 (59)	*
LDL-C	Mean (SD)	140.71 (39.04)	140.03 (38.87)	0.689

Table 2. Nested NAFLD case-control demographics. Diabetes defined as formal diagnosis, treatment or fasting glucose>126mg/dl. HOMA-IR ((fasting insulin*fasting glucose)/405) calculated for those not on treatment for diabetes. LDL-C = (total cholesterol-HDL cholesterol-(triglycerides/5)) *p<0.001

Overall, females outnumbered males and constituted 1277 of the total 2089 participants. 55.2% of the urban population had NAFLD compared to 43.4% of the rural population (an unadjusted OR of 1.60, 1.35-1.91) – a finding which is consistent with the literature on increased NAFLD prevalence in areas of higher socioeconomic class.[14] As expected, mean BMI was higher in the NAFLD group (OR 1.79 if overweight, OR 3.25 if obese) as was rate of diabetes (OR 2.27, 1.85-2.79) and the metabolic syndrome (OR 2.26, 1.86, 2.74). This data is presented in Table 3. Under logistic regression, for every unit increase in BMI, the odds of having NAFLD increases by 1.17 (17%) (1.15, 1.20 p<0.001).

Variable	Unadjusted OR (95% CI, p value)	Adjusted OR (95% CI, p value)
Male sex	1.75 (1.46, 2.09 p<0.001)	2.22 (1.82, 2.72 p<0.001)
Urban domicile	1.60 (1.35, 1.91 p<0.001)	1.21 (1.00, 1.46 p=0.048)
BMI		
- Underweight <18.5kg/m ²	0.12 (0.04, 0.34 p<0.001)	0.14 (0.05, 0.41 p<0.001)
- Overweight 23-24.9kg/m ²	1.79 (1.36, 2.34 p<0.001)	1.78 (1.34, 2.35 p<0.001)
- Obese >25kg/m ²	3.25 (2.58, 4.10 p<0.001)	3.27 (2.57, 4.16 p<0.001)
Diabetes	2.27 (1.85, 2.79 p<0.001)	1.63 (1.28, 2.07 p<0.001)
Metabolic Syndrome	2.26 (1.86, 2.74 p<0.001)	1.62 (1.29, 2.07 p<0.001)

Table 3. NAFLD odds ratios adjusting for sex, domicile, BMI category (with normal weight as baseline), diabetes and metabolic syndrome. Variables were mutually adjusted for each other with baseline categories being those variables presented.

This data confirms commonality of NAFLD risk factors between India and the Western world. Increasing levels of obesity and the dramatic rise in diabetes prevalence in India[1] is reflected in the World Health Organisation (WHO) Global Burden of Disease study, which shows that hypertension and raised fasting blood glucose are the leading cause of non-communicable disease in India, unlike other low-middle income countries where childhood malnutrition and unsafe water predominate.[15] This is reflected within this cohort, where we demonstrate high rates of diabetes - well above the 2011 estimated population prevalence of 8.3%[1] - even within the control group. The main reason for this is the fact that the inception cohort from which the case and control groups were identified included only people of 25 years of age and older, enriching the cohort with people with an age group where prevalence of type 2 diabetes is expected to be substantially higher. In addition, collection of fasting blood glucose levels at enrolment enabled us to identify those who meet the IDF criteria for diabetes without previous diagnosis.

Interestingly, presence of obesity conveys the strongest risk of NAFLD within this group (above diabetes) which highlights the “Asian paradox” – whereby cardiovascular risk increases with BMI in Asians, as in other ethnicities, but at a lower BMI. This data supports the significant impact of BMI on risk of disease in this group and the reclassification of BMI cut-offs that was implemented in 2000.[2]

Recruitment to the Trivandrum NAFLD cohort was completed in July 2016. As yet, it has not undergone a complete follow up. The aim of the project is to collect longitudinal data on risk factors for NAFLD progression and to make an assessment of long-term outcomes such as incident diabetes, cardiovascular events and all-cause mortality rates using verbal autopsy – a validated instrument used to ascertain time and cause of death.[16] Further lifestyle data that was collected from the cohort is under review with the aim to look at the impact of diet and physical activity on risk of NAFLD, both in native Indians and on those who now live in the UK. Data on liver stiffness within the NAFLD sub-group will be combined with a larger piece of work looking at the identification and stratification of liver disease in Trivandrum as a whole.

The cohort also forms the basis of a Global Challenges Research Fund (GCRF) recently awarded to the NIHR Nottingham BRC in collaboration with PHRI to develop lifestyle interventions for improved liver and metabolic health in South India. As a part of GCRF, a pilot randomised control trial of 16 months dietary intervention with low glycaemic index food with yoga sessions vs yoga sessions alone has been initiated in the NAFLD cohort developed and described here (ClinicalTrials.gov NCT03844165).

1
2
3 Finally, DNA has been extracted from whole blood aliquots from this cohort with the aim to
4 complete the first genome-wide association study of NAFLD within an Indian population.
5

6 **STRENGTHS AND LIMITATIONS**

7
8 The biggest strength of the Trivandrum NAFLD cohort is the fact that it was designed to collect
9 detailed anthropometric, clinical and pathological data at a population-level, representing the full
10 spectrum of demography of a large Indian state. Through population-level sampling we have
11 enabled accurate estimation of NAFLD prevalence, which given its significant departure from the
12 estimated global prevalence has filled an important gap in the knowledge-base. NAFLD prevalence of
13 31.8% was estimated from a previous pilot study,[10] which was reliant on raised liver enzymes to
14 identify patients for further investigations, hence, underestimated the true prevalence in this
15 population. Through the utilisation of local social workers within the community, PHRI have created
16 a cohort with excellent uptake and retention of participants, have methods in place to document
17 rates of migration and accurate mortality data as well as having the infrastructure and man-power in
18 place to facilitate longitudinal and interventional studies.
19

20
21 The main limitation of the study from a clinical perspective is the use of ultrasound as the diagnostic
22 tool for NAFLD. Ultrasound has been shown to have a poor yield in the diagnosis of microvesicular
23 fat within the liver, with an overall sensitivity of 43% and specificity of 73%.[17] This means that our
24 NAFLD prevalence of 49.8% may in fact be an under-estimate. Liver biopsy is generally considered to
25 be the gold standard for NAFLD diagnosis, but its invasive nature combined with its inherent risk of
26 sampling error make it unsuitable for population-based studies. Despite this, the message that
27 NAFLD prevalence is significantly higher in this Indian population compared to the global estimates is
28 an important one. Finally, although representative of a large Indian population, the data collected
29 from this cohort is not necessarily representative of the India as a whole – whose diverse population
30 includes thousands of distinct and unique cultures.
31

32 **COLLABORATION**

33
34 Both the NIHR Nottingham Biomedical Research Centre, Gastrointestinal and Liver disorder theme
35 and Population Health and Research Institute (PHRI) welcome interest and collaboration with
36 national and international colleagues. For further information please contact the corresponding
37 author.
38

39 **PATIENT INVOLVEMENT**

40
41 Participants of the Trivandrum NAFLD cohort were not involved in setting the research question or
42 the outcome measures of this project, but during data collection the patient and public involvement
43 (PPI) team from the NIHR Nottingham Biomedical Research Centre ran a 3 day workshop alongside
44 PHRI in Trivandrum. During this workshop, participants were consulted on their opinions of having
45 lifestyle information collected, and the best way to do this. This qualitative feedback has helped
46 shape the way we approach data collection from a cultural standpoint in this ethnic group.
47
48

49 **CONTRIBUTORS**

50
51 KTS and GA contributed the original concept and design of the study. KTS, KBL and JG were involved
52 in data collection and sample management. JC and LB did the data analysis and interpretation whilst
53 JC, LB and KE drafted the work. All the authors have critically revised this article and approved the
54 final version to be published.
55

56 **FUNDING**

57
58 The Nottingham Biomedical Research Centre (digestive diseases theme) supports PHRI and is funded
59 by the National Institute for Health Research (NIHR).
60

COMPETING INTERESTS

None declared

ETHICAL APPROVAL

Ethical approval for the study was granted by the Sree Gokulam Medical College and Research Foundation, Venjaramoodu, Trivandrum, India ethics committee. Written informed consent was obtained from all participants at the time of recruitment.

DATA SHARING STATEMENT

Researchers may request access to data and biomaterial by contacting the corresponding author

For peer review only

1. Whiting, D.R., et al., *IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030*. Diabetes Res Clin Pract, 2011. **94**(3): p. 311-21.
2. WHO, E.C., *Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies*. Lancet, 2004. **363**(9403): p. 157-63.
3. Ong, J.P., A. Pitts, and Z.M. Younossi, *Increased overall mortality and liver-related mortality in non-alcoholic fatty liver disease*. J Hepatol, 2008. **49**(4): p. 608-12.
4. Farrell, G.C. and C.Z. Larter, *Nonalcoholic fatty liver disease: from steatosis to cirrhosis*. Hepatology, 2006. **43**(2 Suppl 1): p. S99-s112.
5. Mohan, V., et al., *Prevalence of non-alcoholic fatty liver disease in urban south Indians in relation to different grades of glucose intolerance and metabolic syndrome*. Diabetes Res Clin Pract, 2009. **84**(1): p. 84-91.
6. Mohan, V., et al., *Intra-urban differences in the prevalence of the metabolic syndrome in southern India -- the Chennai Urban Population Study (CUPS No. 4)*. Diabet Med, 2001. **18**(4): p. 280-7.
7. Younossi, Z.M., et al., *Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes*. Hepatology, 2016. **64**(1): p. 73-84.
8. Prashanth, M., et al., *Prevalence of nonalcoholic fatty liver disease in patients with type 2 diabetes mellitus*. J Assoc Physicians India, 2009. **57**: p. 205-10.
9. Singh, S., G.N. Kuftinec, and S. Sarkar, *Non-alcoholic Fatty Liver Disease in South Asians: A Review of the Literature*. Journal of Clinical and Translational Hepatology, 2017. **5**(1): p. 76-81.
10. K T Shenoy, e.a. *Prevalence and risk factors for non-alcoholic fatty liver in a transition society with high incidence of diabetes mellitus*. in *The 15th Annual Conference of the INASAL*. 2007. Vellore.
11. Cleland, C.L., et al., *Validity of the Global Physical Activity Questionnaire (GPAQ) in assessing levels and change in moderate-vigorous physical activity and sedentary behaviour*. BMC Public Health, 2014. **14**(1): p. 1255.
12. Parkerson, G.R., Jr. and W.E. Broadhead, *Screening for anxiety and depression in primary care with the Duke Anxiety-Depression Scale*. Fam Med, 1997. **29**(3): p. 177-81.
13. Saadeh, S., et al., *The utility of radiological imaging in nonalcoholic fatty liver disease*. Gastroenterology, 2002. **123**(3): p. 745-50.
14. Petersen, K.F., et al., *Increased prevalence of insulin resistance and nonalcoholic fatty liver disease in Asian-Indian men*. Proc Natl Acad Sci U S A, 2006. **103**(48): p. 18273-7.
15. Forouzanfar, M.H., et al., *Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015*. The Lancet, 2016. **388**(10053): p. 1659-1724.
16. Jha, P., et al., *Prospective Study of One Million Deaths in India: Rationale, Design, and Validation Results*. PLOS Medicine, 2005. **3**(2): p. e18.
17. Dasarathy, S., et al., *Validity of real time ultrasound in the diagnosis of hepatic steatosis: a prospective study*. J Hepatol, 2009. **51**(6): p. 1061-7.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 1. Map of Kerala

Figure 2. Trivandrum NAFLD cohort sampling framework

Figure 3. Trivandrum NAFLD cohort recruitment diagram

For peer review only



Figure 1. Map of Kerala

282x463mm (300 x 300 DPI)

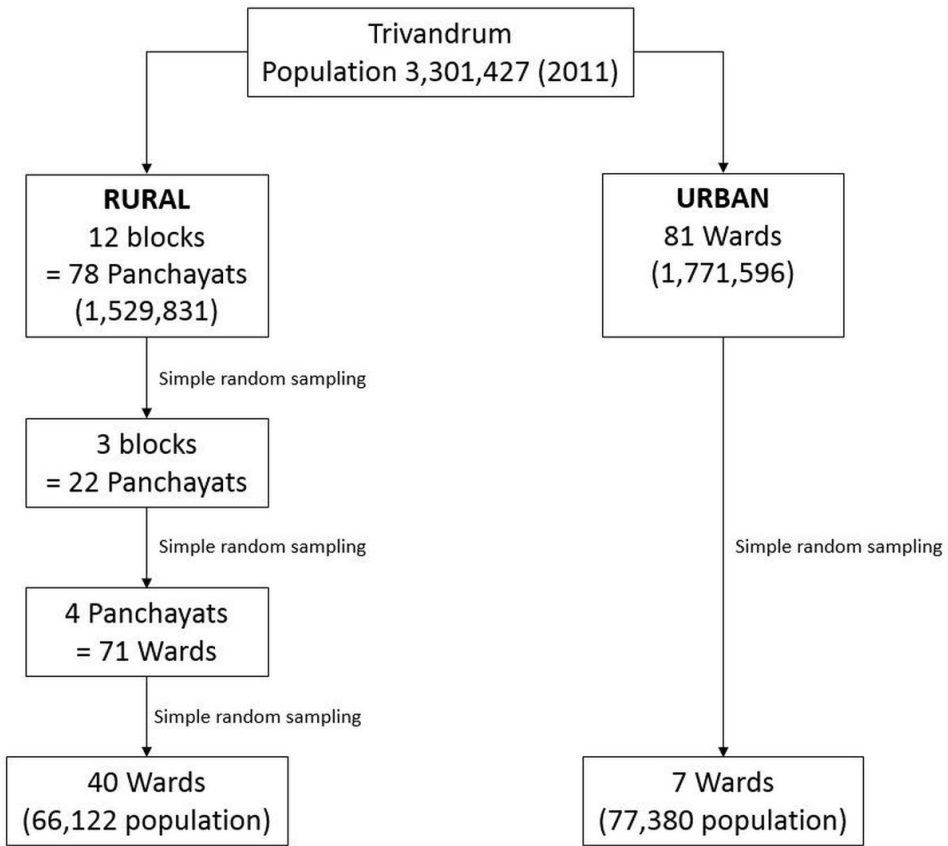


Figure 2. Trivandrum NAFLD cohort sampling framework

106x90mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

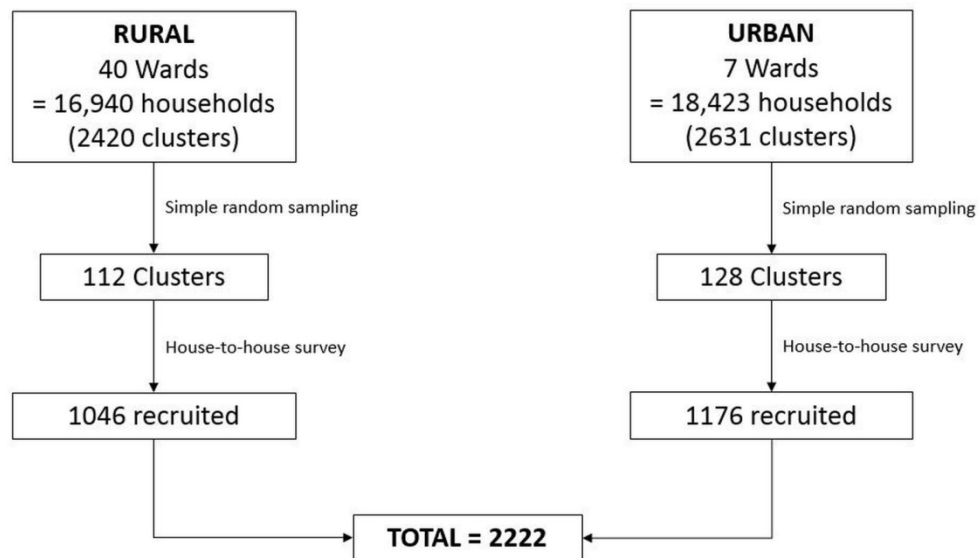


Figure 3. Trivandrum NAFLD cohort recruitment diagram

155x90mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60