

# Dietary profiles and associated metabolic risk factors in India from the ICMR–INDIAB survey-21

---

Received: 24 October 2024

---

Accepted: 6 August 2025

---

Published online: 30 September 2025

---

 Check for updates

---

---

A list of authors and their affiliations appears at the end of the paper

---

Rapid dietary transitions in India have been associated with an alarming rise in cardiometabolic diseases. Using data from the national Indian Council of Medical Research–India Diabetes survey (18,090 adults), we examined India’s dietary profile and the effect of isocaloric substitution of carbohydrates with other macronutrients on metabolic risk. Indian diets are characterized by high intakes of low-quality carbohydrates (white rice, milled whole grains and added sugar), high levels of saturated fat and low intakes of protein. Compared to those with the least carbohydrate intakes, those with the highest intakes had higher risk of newly diagnosed type 2 diabetes (T2D; odds ratio (OR) = 1.30, 95% confidence interval (CI) = 1.14, 1.47), prediabetes (OR = 1.20, 95% CI = 1.06, 1.33), generalized obesity (OR = 1.22, 95% CI = 1.07, 1.37) and abdominal obesity (OR = 1.15, 95% CI = 1.01, 1.30). Replacing refined cereals with whole wheat or millet flour without decreasing overall carbohydrate quantity was not associated with lower risk for T2D (OR = 0.94, 95% CI = 0.57, 1.56) or abdominal obesity (OR = 1.08, 95% CI = 0.66, 1.76). Modeled isocaloric substitution of carbohydrates for plant, dairy, egg or fish protein was associated with lower likelihood of T2D (ranging from OR = 0.89, 95% CI = 0.83, 0.95—for dairy to OR = 0.91, 95% CI = 0.82, 0.99—egg) and prediabetes (ranging from OR = 0.82, 95% CI = 0.72, 0.92—for dairy to OR = 0.94, 95% CI = 0.89, 0.99—for fish). Public health strategies that reduce overall carbohydrates and saturated fat while increasing intake of plant and dairy proteins could mitigate the risk of metabolic diseases in India.

Noncommunicable diseases (NCDs) are major contributors to morbidity and mortality worldwide, particularly in low-income and middle-income countries. India, home to nearly 20% of the global population, has undergone rapid nutrition and epidemiological transitions over the past two decades, resulting in an alarming increase in NCD prevalence. The recent Indian Council of Medical Research–India Diabetes (ICMR–INDIAB) study<sup>1</sup>, a national cross-sectional population-based survey conducted from November 2008 to December 2020, reported weighted (for region, age and sex) prevalence of type 2 diabetes (T2D) and prediabetes of 11.4% and 15.3%, respectively. The prevalence of

generalized obesity and abdominal obesity was also high at 28.6% and 39.5%, respectively. NCDs account for 6.3 million (68%) of all deaths in India<sup>2</sup>. By 2060, the economic costs of overweight and obesity are projected to reach \$839 billion (2.47% of the nation’s GDP)<sup>3</sup>. Given the enormous public health burden of NCDs and their associated economic costs, it is crucial to identify cost-effective, practical strategies for reducing NCD risk.

Unhealthy diets contribute substantially to the NCD burden. Improving diet and physical activity can prevent nearly 50% of incident T2D<sup>4</sup>. Global per capita calorie availability has increased from

---

✉ e-mail: [dranjana@drmohans.com](mailto:dranjana@drmohans.com)

**Table 1 | Policy implications of Indian diets on metabolic risk**

Section	Summary
Background	India is experiencing a dramatic rise in NCDs due to a rapid nutrition and epidemiologic transition, with unhealthy diets being a major contributing factor. Recent studies reveal that NCDs account for 68% of all deaths in India, emphasizing the urgent need for effective dietary strategies to mitigate this public health burden. Aggregated national data indicate that nearly 65–75% of calories in Indian diets are derived from carbohydrates. Given the diversity in Indian diets, no national survey has systematically examined differences in intakes of major macronutrients and their food sources, by region and macronutrient quality. Likewise, the effect of replacing carbohydrates with protein or fat subtypes on metabolic risk is unknown.
Main findings and limitations	Although it is known that regional preferences for main cereal staples in Indian diets vary, they are unified in their high consumption of carbohydrates. We report high intake (62.3%E) of low-quality carbohydrates, such as refined cereals like white rice, milled whole grains and added sugar. Millets were consumed by over one-third of the population in only three states in the country. Intakes of saturated fat were above national dietary recommendations of <10%E for the general population in ten states, whereas if the cutoff for metabolic risk (<7%E) was considered, all but four states were above recommended limits. Overall protein intake was low, with plant protein being the biggest contributor to intake (8.9%E), followed by dairy protein (2.1%E), and animal protein contributing only 1%E. States in the Northeastern region reported the highest intake of total protein (13.6%E) and animal protein (2.9%E). Irrespective of the grain type, higher intakes of total carbohydrates were associated with a 14–30% higher metabolic risk. Replacing 50 g of refined cereals with 50 g of either whole wheat or millet flour was not associated with lower metabolic risk. Replacing 5%E from carbohydrates with 5%E from plant, dairy, egg or fish protein was associated with lower likelihood of T2D and prediabetes but replacing with fat or its subtypes was not. Our study has a few limitations; given the cross-sectional nature of the study, observed associations do not imply a causal relationship and residual confounding remains a possibility. We were unable to capture intakes of added sugar from processed foods like street foods, sugar-sweetened beverages and <i>trans</i> -fat content of commercially available products from the unregulated food sector.
Policy implications	Public health strategies should focus on practical and sustainable approaches to reduce carbohydrate intake while gradually increasing consumption of protein from foods such as pulses, legumes and dairy to achieve a desirable macronutrient composition for preventing NCDs. Such dietary shifts cannot only reverse the nutrition transition but can also help address protein inadequacies and improve overall diet quality. These changes could be facilitated by government subsidies from the PDS to prioritize healthier protein sources like pulses and legumes over highly refined grains like white rice. In addition, encouraging the use of healthier edible oils—low in saturated fats—can further lower the risk of diet-related NCDs. Increasing the availability of protein sources under the minimum support prices scheme could enhance both the availability and intake of quality protein among the population. Because healthcare is a state government issue, the study’s findings on macronutrient intake from major sources are vital for helping states enhance their PDS plans and promote subsidies that prevent NCDs within their jurisdictions. A multisectoral approach involving healthcare, agriculture and socioeconomic policy reforms is crucial for promoting healthier diets to lower NCD risk.

the 1980s to 2013, with carbohydrates contributing the most (70.5%) and protein the least (10.5%)<sup>5</sup>. In India, data from epidemiological studies<sup>6</sup> and national nutrition surveys, such as those by the National Nutrition Monitoring Bureau (NNMB)<sup>7</sup> and National Sample Survey Organization<sup>8</sup>, indicate that Indian adults currently consume ~65% to 75% of calories (%E) from carbohydrates (one of the highest in the world), 9–11%E from proteins and 14–23%E from fats. Previous ICMR–INDIAB analyses reported macronutrient intakes stratified by glycemic status<sup>9</sup>. However, examining macronutrients by their quality (refined carbohydrates versus whole grains, animal versus plant protein and saturated fat versus unsaturated fat) is essential for effective national dietary recommendations, as they are differentially associated with NCDs<sup>10–12</sup>. Due to India’s dietary diversity, understanding inter-regional and intraregional differences in macronutrient intake can inform targeted policy recommendations. While evidence suggests that reducing low-quality carbohydrates and increasing intake of plant protein and healthy fats improves cardiometabolic health<sup>13,14</sup>, there is limited data on the effect of replacing carbohydrates with other macronutrients among Asian Indians.

The ICMR–INDIAB survey is a nationally representative cross-sectional survey of NCDs in Indian adults from urban and rural areas of 30 states, union territories and the National Capital Territory (NCT) of Delhi<sup>1</sup>. Dietary data from every fifth participant were used to characterize India’s dietary profile across regions and examine inter-regional and intraregional differences in macronutrient intakes. We also investigated the associations between dietary factors and metabolic risk, and evaluated the effects of isocaloric substitution of macronutrients on glycaemia and adiposity. Study findings and policy implications are summarized in Table 1.

## Results

### Study characteristics

The demographic and metabolic risk factors overall and by region are given in Table 2. The median age of participants was 40 years and

49% were male. Tobacco use (in both smoke and smokeless forms) was reported by 15% of the overall population, ranging from 8% in the West to 24% in the Northeast. Nationally, current alcohol use that included toddy (fermented Asian palm sap), country liquor and other alcoholic beverages was 15% with the highest prevalence in the Northeastern region (23%). A total of 61% were physically inactive, defined as a physical activity level (PAL) between 1.40 and 1.69 (ref. 15). The median body mass index (BMI) was 22.2 kg m<sup>-2</sup>, 43% were overweight (BMI ≥ 23 to <25 kg m<sup>-2</sup>), 26% were obese (BMI ≥ 25.0 kg m<sup>-2</sup>) and 36% had abdominal obesity (waist circumference ≥ 90 cm for men and ≥80 cm for women). The Northern region reported the highest rates of overweight (54%), obesity (37%) and abdominal obesity (48%). Conversely, Eastern region states had the lowest prevalence of overweight (31%), obesity (16%) and abdominal obesity (29%). Nearly one-third of the population had hypertension (27%) with little inter-regional differences. The prevalence of newly diagnosed T2D was 9% overall, with the highest prevalence in the West (12%) and lowest in the East and Northeast (6%). Newly diagnosed prediabetes was prevalent in 41% with the highest prevalence in the Central region (49%). Dyslipidaemia, defined using the National Cholesterol Education Program Adult Treatment Panel III guidelines<sup>16</sup>, was prevalent in half of the population. Overall, 83% of participants had at least one metabolic risk factor defined as the presence of either newly diagnosed T2D, prediabetes, dyslipidaemia, general obesity, abdominal obesity or hypertension.

Compared to rural participants, urban participants were less likely to use tobacco (14% versus 16%) and alcohol (14% versus 15%), but were more likely to be physically inactive (70% versus 57%), and had higher weighted median BMI (23.5 kg m<sup>-2</sup> versus 21.6 kg m<sup>-2</sup>), waist circumference (83.5 cm versus 78.5 cm), fasting capillary blood glucose (96 mg dl<sup>-1</sup> versus 95 mg dl<sup>-1</sup>), serum triglycerides (115 mg dl<sup>-1</sup> versus 108 mg dl<sup>-1</sup>) and low-density lipoprotein cholesterol (100 mg dl<sup>-1</sup> versus 90 mg dl<sup>-1</sup>). They also had a higher weighted prevalence of overweight (56% versus 37%), generalized obesity (37% versus 22%), abdominal obesity (48% versus 31%), hypertension (29% versus 27%) and metabolic

**Table 2 | Demographic characteristics and metabolic risk factors of the study population stratified by regions (n=18,090)**

Variables	Overall (n=18,090)	Northern (n=3,714)	Southern (n=3,198)	Eastern (n=2,045)	Western (n=1,730)	Central (n=2,470)	Northeastern (n=4,933)
Age (years)	40 (30, 52)	40 (31, 50) <sup>a</sup>	41 (31, 53)	40 (30, 53) <sup>a</sup>	42 (32, 53)	40 (30, 53) <sup>a</sup>	39 (30, 51)
Sex (male) (%)	49	49 <sup>a</sup>	45	50 <sup>a</sup>	48 <sup>a</sup>	51 <sup>a</sup>	50 <sup>a</sup>
Current tobacco use (%)	15	13	13	13	8	10	24
Current alcohol use (%)	15	9	16	16	7	10	23
Family history of T2D (%)	9	10	16	7	9	5	6
Physically inactive <sup>b</sup> (%)	61	67	64	55	71	61	54
Metabolic risk factors							
BMI (kg m <sup>-2</sup> )	22.2 (19.5, 25.1)	23 (20, 26)	23 (20, 26)	21 (18, 24)	22 (19, 25) <sup>a</sup>	22 (19, 25)	22 (19, 24)
Waist circumference (cm)	80.0 (72.0, 88.5)	84.9 (76.3, 92.4)	81.0 (72.7, 89.1)	76.9 (69.0, 86.0)	79.1 (70.0, 87.6)	79.7 (72.1, 87.4) <sup>a</sup>	78.0 (71.1, 86.0)
Male (cm)	81.1 (73.2, 89.4)	85.3 (78.1, 92.9)	82.1 (73.6, 90.7)	78.2 (71.0, 88.1)	81.2 (73.1, 90.0)	81.0 (73.9, 88.2) <sup>a</sup>	78.4 (72.0, 86.3)
Female (cm)	79.0 (70.4, 87.6)	84.0 (75.0, 92.1)	80.1 (72.1, 88.0)	75.0 (66.5, 83.5)	76.3 (67.0, 85.1)	78.5 (70.7, 86.1) <sup>a</sup>	77.9 (70.0, 86.0)
Overweight <sup>c</sup> (%)	43	54	49 <sup>a</sup>	31	43	40	37
Generalized obesity <sup>d</sup> (%)	26	37	33	16	26	23	21
Abdominal obesity <sup>e</sup> (%)	36	48	41	29	34	32	29
Systolic blood pressure (mm Hg)	126 (116, 137)	127 (118, 137)	125 (115, 136)	127 (117, 138)	126 (116, 137) <sup>a</sup>	125 (116, 134)	126 (116, 137) <sup>a</sup>
Diastolic blood pressure (mm Hg)	81 (74, 87)	82 (76, 88)	79 (71, 86)	80 (73, 87) <sup>a</sup>	79 (73, 86)	82 (76, 88)	80 (73, 87) <sup>a</sup>
Hypertension <sup>f</sup> (%)	27	28	26	29	27	26	28
Fasting CBG (mg dl <sup>-1</sup> )	96 (87, 105)	97 (88, 107)	96 (88, 106)	95 (88, 104) <sup>a</sup>	97 (90, 106)	98 (90, 107)	92 (84, 101)
HbA1c (%)	5.4 (5.0, 5.6)	5.4 (5.1, 5.7)	5.5 (5.2, 5.8)	5.3 (4.9, 5.5)	5.4 (5.0, 5.6) <sup>a</sup>	5.4 (5.1, 5.6)	5.2 (4.8, 5.5)
Newly diagnosed T2D <sup>g</sup> n (%)	9	11	11	6	12	9	6
Newly diagnosed prediabetes <sup>h</sup> (%)	41	43	46	39	42	49	34
Serum triglyceride (mg dl <sup>-1</sup> )	111 (79, 160)	114 (82, 169)	114 (82, 164)	101 (75, 143)	101 (74, 147)	104 (72, 148)	118 (84, 168)
LDL (mg dl <sup>-1</sup> )	93 (69, 118)	95 (67, 121)	102 (79, 128)	90 (70, 114)	107 (83, 136)	83 (52, 113)	87 (66, 108)
HDL (mg dl <sup>-1</sup> )	39 (33, 47)	41 (35, 48)	39 (32, 46)	38 (32, 46)	40 (34, 49)	39 (33, 46) <sup>a</sup>	39 (33, 46) <sup>a</sup>
Male (mg dl <sup>-1</sup> )	38 (32, 45)	38 (32, 44) <sup>a</sup>	37 (31, 44) <sup>a</sup>	37 (32, 44) <sup>a</sup>	38 (32, 46) <sup>a</sup>	37 (32, 44) <sup>a</sup>	38 (32, 46) <sup>a</sup>
Female (mg dl <sup>-1</sup> )	41 (35, 48)	43 (37, 50)	40 (34, 47) <sup>a</sup>	40 (34, 47) <sup>a</sup>	43 (36, 51)	41 (34, 47) <sup>a</sup>	40 (34, 47) <sup>a</sup>
Dyslipidaemia <sup>i</sup> (%)	50	56	53	43	53	44	50 <sup>a</sup>
Metabolic risk <sup>j</sup> (%)	83	89	86	77	85	84	79

Data are presented as weighted median (IQR) for continuous variables and weighted (%) for categorical variables. *P* < 0.05 was considered as significant. The *P* value was tested using median and quantile test for continuous variables and chi-square tests for categorical variables. These tests were used to compare either the weighted median (continuous variables) or the proportions (categorical variables) for each region with overall. PAL was calculated as TEE for 24 h expressed as a multiple of basal metabolic rates, and calculated as TEE/basal metabolic rate for 24 h. PAL values are as follows: sedentary, 1.40–1.69; moderate, 1.70–1.99 and vigorous, 2.0–2.40. <sup>a</sup>Nonsignificant variables. <sup>b</sup>Physically inactive was defined as participants with sedentary PALs (1.40–1.69). <sup>c</sup>Overweight was defined as BMI ≥ 23 to < 25 kg m<sup>-2</sup>. <sup>d</sup>Generalized obesity was defined as BMI ≥ 25 kg m<sup>-2</sup>. <sup>e</sup>Abdominal obesity was defined as waist circumference of ≥ 90 cm for males and ≥ 80 cm for females. <sup>f</sup>Hypertension was defined as systolic blood pressure ≥ 140 mm Hg and diastolic blood pressure ≥ 90 mm Hg. <sup>g</sup>Newly diagnosed T2D was defined as HbA1c ≥ 6.5% and/or if fasting CBG ≥ 126 mg dl<sup>-1</sup> (≥ 7 mmol l<sup>-1</sup>) or 2-h postglucose load CBG ≥ 220 mg dl<sup>-1</sup> (≥ 12.2 mmol l<sup>-1</sup>). <sup>h</sup>Newly diagnosed prediabetes was defined as HbA1c 5.7–6.4% or if fasting CBG was ≥ 100–125 mg dl<sup>-1</sup> (≥ 5.6–6.9 mmol l<sup>-1</sup>) and/or if 2-h postload CBG value was ≥ 160 and < 220 mg dl<sup>-1</sup> (≥ 8.9 to < 12.2 mmol l<sup>-1</sup>). <sup>i</sup>Dyslipidaemia was defined as serum cholesterol concentrations of ≥ 200 mg dl<sup>-1</sup> (≥ 5.2 mmol l<sup>-1</sup>); serum triglyceride concentrations of ≥ 150 mg dl<sup>-1</sup> (≥ 1.7 mmol l<sup>-1</sup>) or low HDL-C (male < 40 mg dl<sup>-1</sup> (1.04 mmol l<sup>-1</sup>) and female < 50 mg dl<sup>-1</sup> (1.3 mmol l<sup>-1</sup>)) and LDL-C concentrations of ≥ 130 mg dl<sup>-1</sup> (3.4 mmol l<sup>-1</sup>). <sup>j</sup>Metabolic risk was defined as the presence of any one of the metabolic risk factors, including newly diagnosed T2D, newly diagnosed prediabetes, dyslipidaemia, hypertension, generalized obesity or abdominal obesity. LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol; CBG, capillary blood glucose.

risk (87% versus 82%). Compared to men, women were younger (39 years versus 41 years) and were less likely to use tobacco (2% versus 28%) or alcohol (3% versus 27%), but more likely to be physically inactive (65% versus 56%), overweight (43% versus 40%), have general obesity (30% versus 23%) and abdominal obesity (47% versus 24%); however, they were less likely to have hypertension (24% versus 31%) or dyslipidaemia (42% versus 60%; Supplementary Table 1).

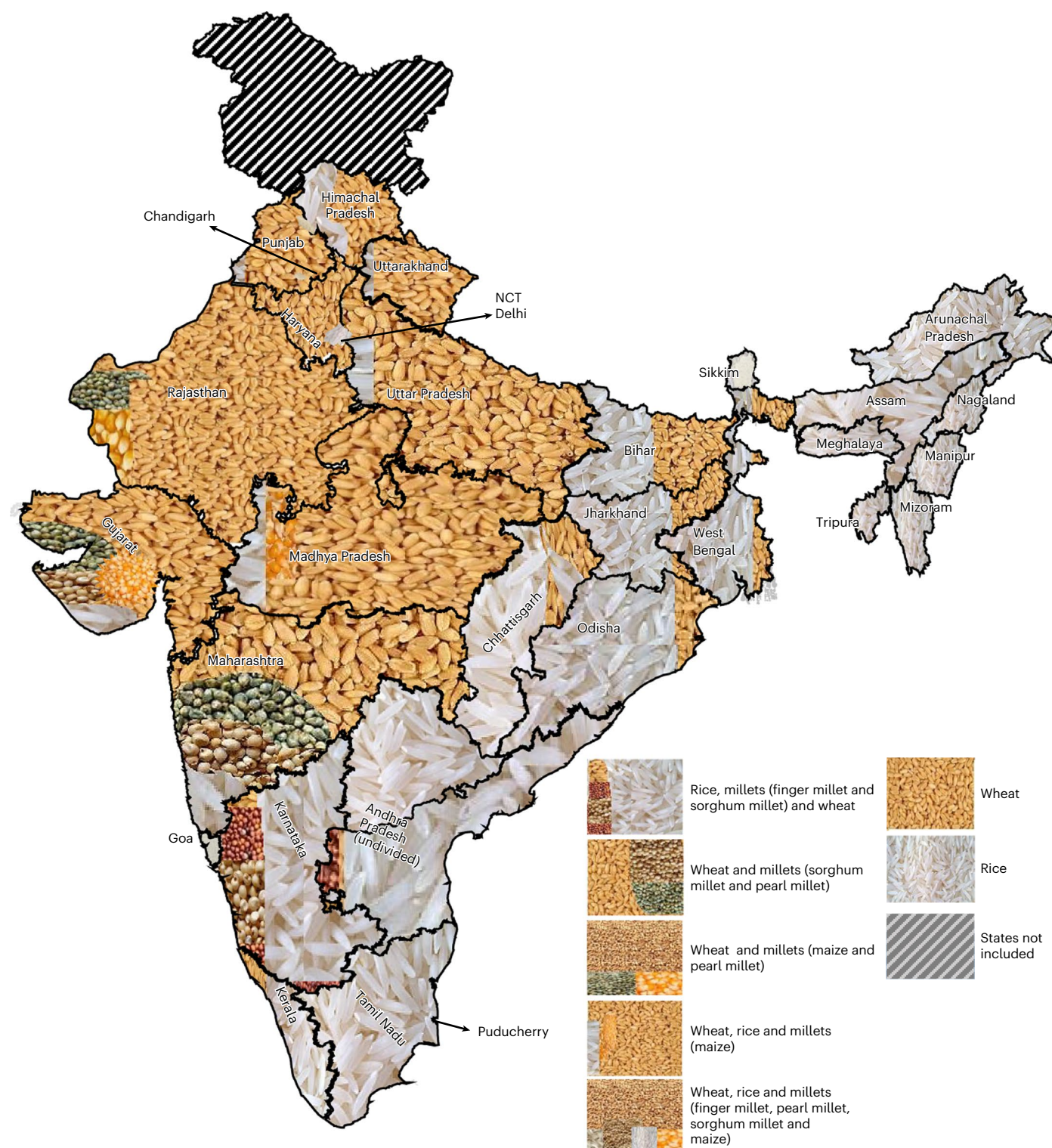
Sensitivity analyses showed participants in the diet sub-study were more likely to be male, current tobacco or alcohol users, and less likely to have a family history of T2D (Supplementary Table 2)

with no significant differences in BMI, blood pressure or fasting capillary glucose concentrations compared to those not included in this study.

**National nutrition profile**

Cereal staple choices, the primary constituents of Indian diets, varied widely by region (Fig. 1 and Extended Data Table 1). Refined cereal (white rice) was the primary choice in the Northeast (99%), South (87%) and East (78%), and whole wheat (milled as flour) in the North (90%) and Central (70%) regions. Intraregional variations included a lower preference for white rice in Karnataka (Southern region, 47.6%)

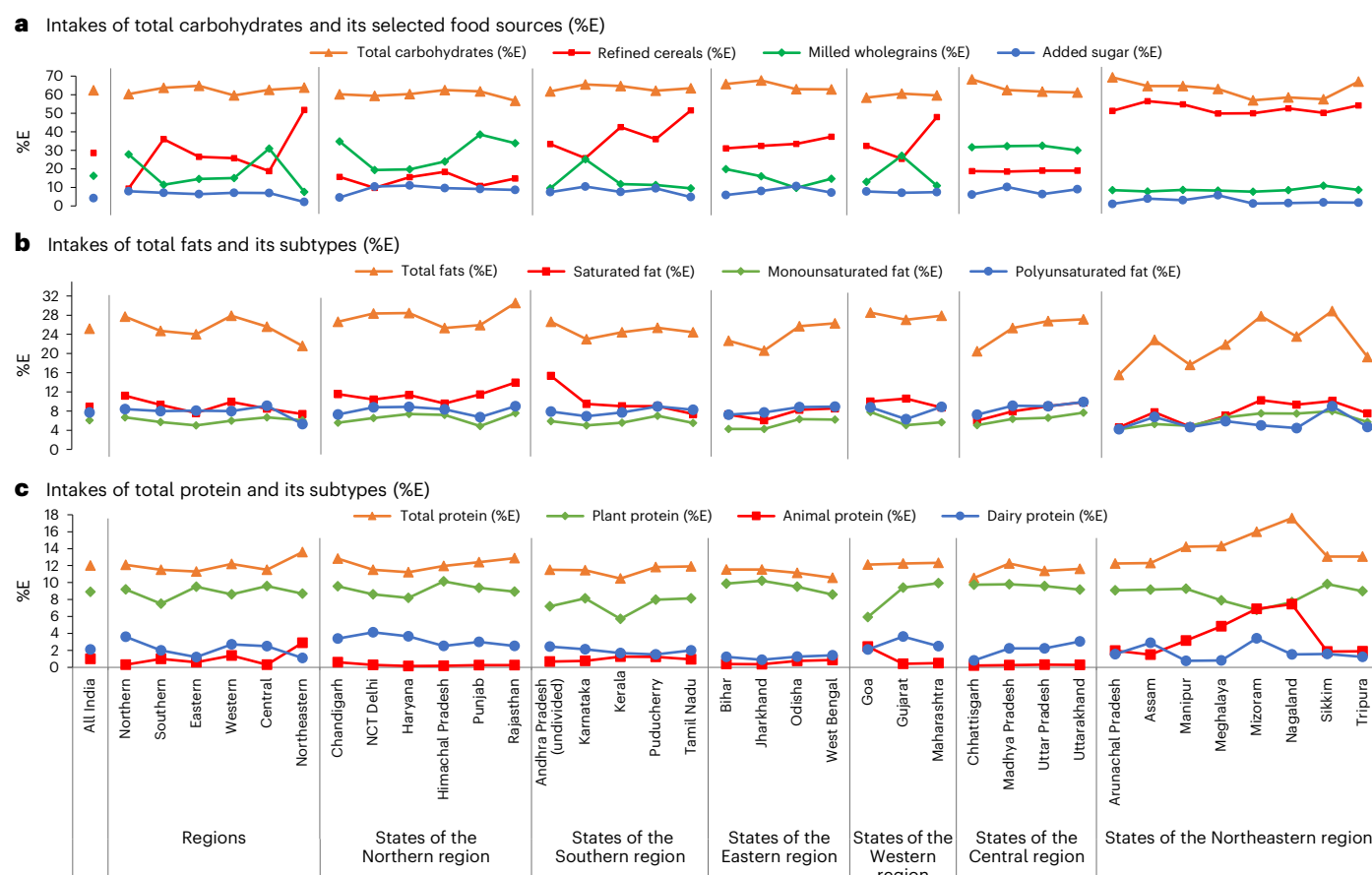




**Fig. 1 | Top three cereal grain staple choices reported by the study population by states and union territories in India ( $n = 18,090$ ).** Map represents the top three cereal staples reported by the weighted percentage of the study population in each state.

and Bihar (Eastern region, 53%). Millets were consumed alongside rice and wheat in 16 states, while whole millet flour was consumed only in Karnataka (47.7%) and Gujarat and Maharashtra (both 35%). Among the millet choices, sorghum and finger millet were commonly consumed in Karnataka, while pearl millet and maize consumption were greater in Gujarat, and pearl millet and sorghum were also commonly consumed in Maharashtra.

Macronutrient profiles expressed as a %E for the nation overall and by geographic regions are shown in Fig. 2a–c and Extended Data Table 2. Nationwide, carbohydrates contributed 62.3% of total daily energy intake, mainly from refined cereals (28.5%E) and milled whole grains (16.2%E). Total fat contributed 25.2%E, while protein intake was low at 12%E (Extended Data Table 2), below the National Dietary Guideline recommendation of 15%E<sup>17</sup>.



**Fig. 2 | Inter- and intraregional intakes of macronutrients among the study population ( $n = 18,090$ ).** Data are presented as weighted median. Median and IQR values are provided in Extended Data Tables 2–5. **a**, Intakes of total carbohydrates and their selected food sources (%E). Refined cereals include white rice, refined wheat flour (maida), refined millet grains/flour (minor millets like little millet and foxtail millet) and wheat semolina (rava). Milled whole grains include whole wheat, whole millets and whole grains milled into flour. Added sugar includes white sugar, honey, jaggery and palm sugar added during cooking and on the table. **b**, Intakes of total fats and their subtypes. **c**, Intakes of total protein and

their subtypes. Plant protein includes cereals, pulses, legumes, fruits, vegetables (including roots and tubers), nuts and oilseeds. Dairy protein includes fermented dairy protein (Indian style yogurt (curd), buttermilk, lassi) and nonfermented dairy protein (all types of milk (cow/buffalo/yak/goat), milk added to coffee and tea, paneer (curdled milk solids)). Animal protein includes red meat, poultry, fish and egg, and does not include dairy protein. For refined cereals and milled whole grains, the analysis only included participants who consumed  $\geq 50$  g. For added sugar, the analysis only included participants who consumed  $\geq 1$  teaspoon ( $\geq 5$  g  $d^{-1}$ ). The connected lines for each region do not imply a trend.

## Carbohydrate intake patterns

Carbohydrate intake was high across regions, with little inter-regional variability, 59.6% in the West to 64.8% in the East. However, sources of carbohydrate varied, with refined cereal intake highest in the Northeast (51.7%E), followed by the South (36%E) and East (31.5%E), largely from white rice (95% of refined cereals; Fig. 2a and Extended Data Table 2). The Central and Northern regions reported the highest intakes of milled whole grains (30.9%E and 27.8%E, respectively, versus 16.2%E nationally), mainly from wheat flour (15%E of 16.2%E). Millet flour contributed minimally (1.4%E). There was also a substantial inter-regional variability in added sugar intake, with median intakes ranging from 2.1%E in the Northeast to 7.9%E in the North.

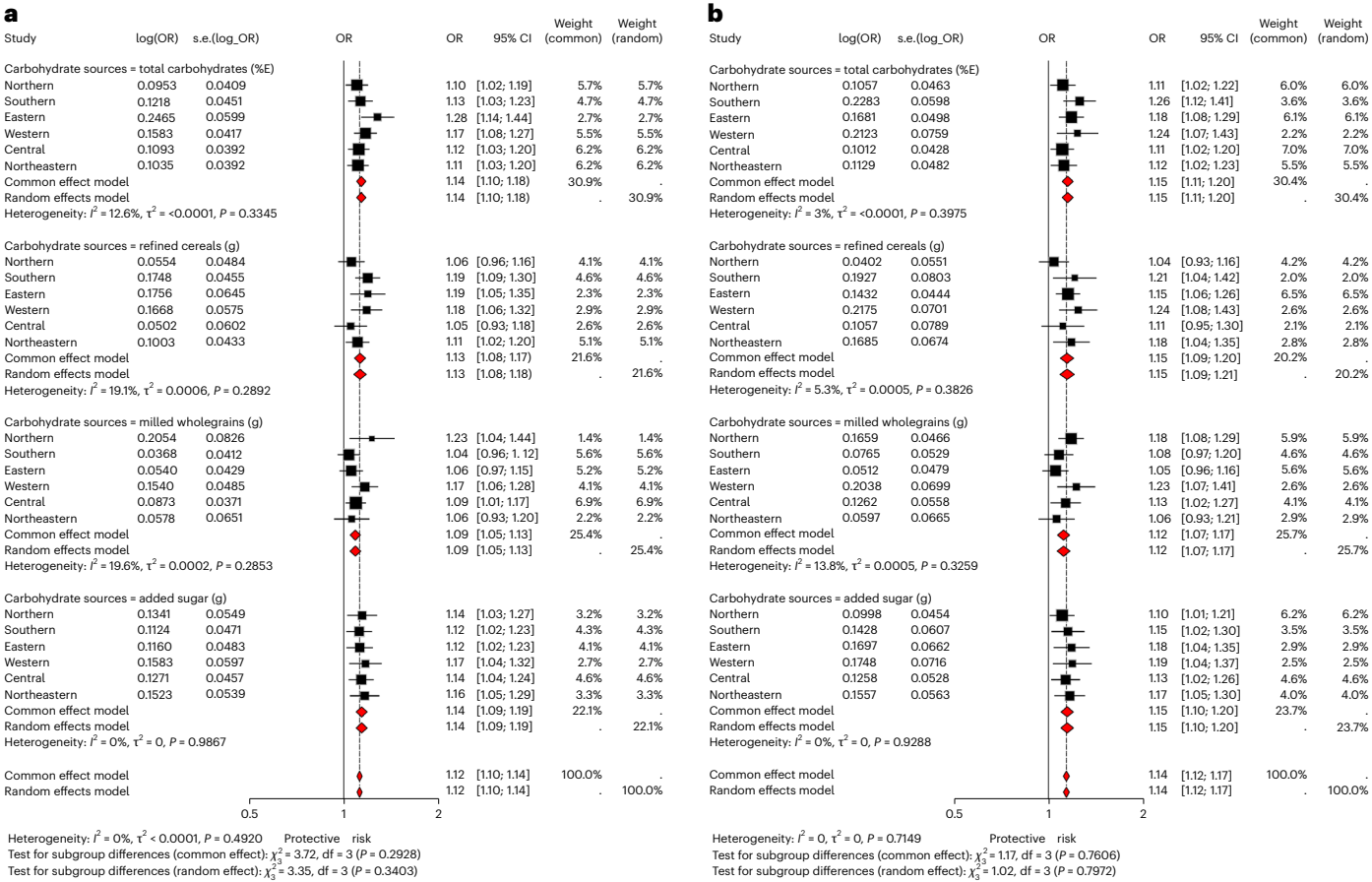
Substantial variability in refined cereal intake was seen among states within regions (Fig. 2a and Extended Data Table 3). For example, within the Southern region, the consumption of refined cereals in Tamil Nadu was 51.5%E, of which white rice contributed to 47.3%E, as compared to the neighboring state of Karnataka, where the intake was only 25.8%E (23.8%E from white rice). Karnataka had the highest whole grain consumption in the region (25.1%E) with millets contributing 10%E. In the North, refined cereal consumption varied from 9.9%E (7.4%E from white rice) in NCT of Delhi to 18.4%E (14.6%E from white rice) in Himachal Pradesh. The least variability in refined cereal intake was in

the Northeastern states, ranging from 49.8%E in Meghalaya to 56.5%E in Assam. Milled whole grain intake varied widely in the West, ranging from 9.3%E in Goa to 27.1%E in Gujarat. In the North, milled whole grain intakes ranged from 19.3%E in Delhi to 38.5%E in Punjab. Karnataka in the South and Gujarat in the West derived an equal proportion of energy from refined cereals and milled whole grains.

Consumption of added sugars (which only included sugar added during cooking, at the table and in processed foods that label for added sugar, such as aerated drinks) was above the recommended 5%E<sup>17</sup> in 19 states, the NCT of Delhi and the union territory of Puducherry (Extended Data Table 3). Haryana and NCT of Delhi in the North, Karnataka in the South, Odisha in the East and Madhya Pradesh in the Central region had the highest added sugar intakes ( $>10$ %E). Added sugar consumption varied widely in the South (4.8%E in Tamil Nadu to 10.4%E in Karnataka) with little intraregion variability ( $\sim 1$ %) in the West. In the Northeast, added sugar intake was within recommended limits ( $<5$ %) except in Meghalaya (5.8%E; Extended Data Table 3).

## Total fat and fat subtype patterns

Total fat intake ranged from 21.6%E in the Northeast to 27.9%E in the West (Fig. 2b and Extended Data Table 2). Saturated fats were the most consumed nationally (8.9%E), with the highest intake in the North



**Fig. 3 | Pooled odds ratios (95% CI) for the association of total carbohydrates (%E) with likelihood of newly diagnosed T2D and newly diagnosed prediabetes in the ICMR-INDIAB study (n = 18,090). a, Newly diagnosed T2D. b, Newly diagnosed prediabetes. A fixed-effects model was used to synthesize data from the regions. Two-sided P values were used in calculating ORs and 95% CIs through survey logistic regression. All dietary exposure variables (refined cereals (g d<sup>-1</sup>), milled whole grains (g d<sup>-1</sup>), added sugar (g d<sup>-1</sup>)) were adjusted for total energy using the residual model. Models were adjusted for PSU, state code, population weight, age (in years), sex (male/female), family history of T2D (yes/no), tobacco use (never/current/past), alcohol use (never/current/past), PAL (sedentary/moderate/vigorous), time period the study was conducted, education qualification (primary school/high school/college), BMI (kg m<sup>-2</sup>),**

(11.2%E) and lowest in the Northeast (7.4%E). Monounsaturated fatty acid (MUFA) intake was low overall (6.1%), ranging from 5% in the East to 6.7% in the North and Central regions. Polyunsaturated fatty acids (PUFA) intake showed more inter-regional variability, from 5.3%E in the Northeast to 9.1%E in the Central region. The Northeastern states had the lowest intakes of n6 polyunsaturated fat (4.6%E), and they had the highest intakes of n3 polyunsaturated fat (0.4%E) compared to intake nationally and in other regions.

Intraregional variation in total fat intake was minimal except in the Northeast, where it ranged from 15.6%E in Arunachal Pradesh to 28.9%E in Sikkim (Extended Data Table 4). All Northern states (except Himachal Pradesh), Andhra Pradesh (undivided) in the South, Goa and Gujarat in the West and Mizoram and Sikkim in the Northeast exceeded the general recommended saturated fat intake of <10%E<sup>18</sup>. Except for Jharkhand in the East, Chhattisgarh in the Central region and Arunachal Pradesh and Manipur in the Northeast, no state met the recommended saturated fat intake of <7%E<sup>19</sup> for populations at risk of cardiovascular disease. MUFA and PUFA intakes also varied widely within regions. Within the Northeast, MUFA and PUFA intakes

place of residence: urban versus rural, intakes of total energy (kcal d<sup>-1</sup>), fat (%E), fruits (g d<sup>-1</sup>), vegetables (g d<sup>-1</sup>), tubers (g d<sup>-1</sup>), milk and milk products (g d<sup>-1</sup>), alcohol (g d<sup>-1</sup>), pulses and legumes (g d<sup>-1</sup>), added salt (g d<sup>-1</sup>) and edible oils and fats (g d<sup>-1</sup>), hypertension, dyslipidaemia and abdominal obesity. Models with food sources as the main exposure were adjusted for the remaining food sources while models with total carbohydrates (%E) were adjusted for total fat (%E). Newly diagnosed T2D was defined as HbA1c ≥6.5% and/or if fasting CBG ≥126 mg dl<sup>-1</sup> (≥7 mmol l<sup>-1</sup>) or 2-h postglucose load CBG ≥220 mg dl<sup>-1</sup> (≥12.2 mmol l<sup>-1</sup>). Newly diagnosed prediabetes was defined as HbA1c 5.7–6.4% or if fasting CBG was ≥100–125 mg dl<sup>-1</sup> (≥5.6–6.9 mmol l<sup>-1</sup>) and/or if 2-h postload CBG value was ≥160 and <220 mg dl<sup>-1</sup> (≥8.9 to <12.2 mmol l<sup>-1</sup>). CBG, capillary blood glucose.

ranged at 4.2%E each in Arunachal Pradesh to 8%E and 9%E in Sikkim. Manipur and Nagaland (0.9%E each) had the highest intakes of n3 polyunsaturated fat in the nation, followed by Mizoram (0.5%E; Extended Data Table 4).

### Total protein and protein subtype patterns

Overall, total protein intake in India was low (12%E), with the Northeast having the highest intake (13.6%E; Fig. 2c and Extended Data Table 2). Most protein came from plant sources (8.9%E) with little inter-regional differences. The North had the highest dairy protein intake (3.6%E), while the East (1.2%E) and Northeast (1.1%E) had the lowest. Fermented dairy contributed 0.2%E nationally, with intakes ranging from 0.02%E in the Northeast to 0.5%E in the North. Nonfermented dairy intake was 1.7%E nationwide, ranging from 0.9%E in the Northeast to 3%E in the North. Animal protein intake varied widely from a low of 0.3%E in the North and Central regions to 2.9%E in the Northeast, with poultry as the main contributor (0.3%E).

Intraregional differences in protein intake were minimal except in the Northeast, where it ranged from 12.2%E in Arunachal Pradesh to



17.6%E in Nagaland (also the nation's highest). With regard to protein subtypes, Himachal Pradesh (North) and Jharkhand (East) had the highest plant protein intakes (10.2%E each), while Kerala in the South (5.7%E) and Goa in the West (5.9%E) had the lowest (Extended Data Table 5). The NCT of Delhi (4.1%E) and Haryana (3.7%E) in the North and Gujarat (3.6%E) in the West had the highest dairy protein intakes; Manipur and Meghalaya in the Northeast and Chhattisgarh in the Central region had the lowest intakes at 0.8%E. Haryana and Rajasthan in the North and Gujarat in the West reported the highest fermented dairy protein intakes at 0.7%E each. Rajasthan had the highest median intake of non-fermented dairy protein (4.7%E) in the nation, while Arunachal Pradesh in the Northeast reported the lowest at 0.4%E (Extended Data Table 5).

The highest animal protein intake was seen in the Northeastern states of Meghalaya (4.9%E), Mizoram (6.9%E) and Nagaland (7.4%E). They also had the highest protein intakes from red meat—Nagaland (4.2%E), Mizoram (2.3%E) and Meghalaya (2.0%E). Mizoram also reported a high intake of poultry protein (2.3%E). Egg protein intake was low and showed little variation nationwide. States in the Northeastern region like Manipur (1.5%E), Meghalaya (1.1%E) and Nagaland (1.0%E) reported the highest intakes of fish protein, along with Goa in the West (1.2%E; Extended Data Table 5).

### Nutrient profile by place of residence and sex

Compared to urban participants, rural participants consumed more energy from total carbohydrates and refined cereals, but less from added sugars and all major fat subtypes, including saturated fat, MUFA and PUFA. They had higher intakes of plant and animal protein but lower intakes of dairy protein. Compared to men, women consumed fewer calories (%E) and protein, but more added sugar and saturated fat (%E; Supplementary Table 3).

### Carbohydrate intake and metabolic risk

Higher calorie intake from carbohydrates was associated with a 14% higher likelihood of newly diagnosed T2D (odds ratio (OR) = 1.14, 95% confidence interval (CI) = 1.10–1.18). Pooled meta-analyses of the six major regions in India showed that refined cereals, milled whole grains and added sugars were associated with higher odds of newly diagnosed T2D by 13% (95% CI = 1.08–1.18), 9% (95% CI = 1.05–1.13) and 14% (95% CI = 1.09–1.19), respectively (Fig. 3a). There was no evidence of inter-regional heterogeneity in risk estimates. Higher intakes of total carbohydrates (OR = 1.15, 95% CI = 1.11–1.20), refined cereals (OR = 1.15, 95% CI = 1.09–1.21) and added sugars (OR = 1.15, 95% CI = 1.10–1.20) were each associated with a 15% higher likelihood of newly diagnosed prediabetes. Although the risk estimates for milled whole grains were attenuated, their consumption was still associated with higher odds of newly diagnosed prediabetes (OR = 1.12, 95% CI = 1.07–1.17; Fig. 3b).

Higher intakes of total carbohydrate calories (OR = 1.15, 95% CI = 1.10–1.20) and refined cereals (OR = 1.15, 95% CI = 1.10–1.21) were each associated with 15% higher odds of general obesity (Extended Data Fig. 1a). Milled whole grains (OR = 1.11, 95% CI = 1.06–1.16) and added sugars (OR = 1.12, 95% CI = 1.07–1.17) were associated with a weaker, but still significant, odds of general obesity. Likewise, higher intakes of total carbohydrates (OR = 1.18, 95% CI = 1.11–1.25), refined cereals (OR = 1.08, 95% CI = 1.03–1.14), milled whole grains (OR = 1.14, 95% CI = 1.08–1.20) and added sugars (OR = 1.10, 95% CI = 1.06–1.15) were all associated with higher odds of abdominal obesity (Extended Data Fig. 1b).

Compared to the lowest carbohydrate intakes, those with the highest intakes (69%E (66.7, 72.7)) had a 15–30% higher likelihood of newly diagnosed T2D, prediabetes, general obesity and abdominal obesity (Extended Data Fig. 2a). When examining carbohydrate subtypes, those with the highest consumption of refined cereals (median (interquartile range (IQR)) = 256.5 (218.7, 316.9) g d<sup>-1</sup>) had an 18–25% higher likelihood of being diagnosed with T2D, prediabetes, general obesity and abdominal obesity compared to the lowest intake group (median (IQR) = 93.2 (75.8, 109.8) g d<sup>-1</sup>; Extended Data Fig. 2b). Those who

consumed the highest quantities of white rice (median (IQR) = 198.4 (169.8–252.4) g d<sup>-1</sup>) had a 19–26% higher likelihood of being newly diagnosed with T2D, prediabetes, general obesity and abdominal obesity compared to those who consumed the least (median (IQR) = 81.5 (66–88.2) g d<sup>-1</sup>; Extended Data Fig. 2c).

Compared to those with the lowest consumption of milled whole grains (median (IQR) = 80.1 (64.9–93.6) g d<sup>-1</sup>), those with the highest intakes (median (IQR) = 262.2 (219.8–324.1) g d<sup>-1</sup>) had 11–29% higher risks of T2D, prediabetes, general obesity and abdominal obesity (Extended Data Fig. 2d). Similarly, a higher risk of 14–31% for T2D, prediabetes and general obesity was observed for the highest tertiles of whole wheat (median (IQR) = 249.9 (211.1–312.7) g d<sup>-1</sup>) and 21–29% for milled whole millets (median (IQR) = 172.6 (144.3–218.1) g d<sup>-1</sup>) consumption (Extended Data Fig. 2e,f). Those with the highest intakes of added sugar (median (IQR) = 46.5 (37.3–62.7) g d<sup>-1</sup>) had a 14–26% higher likelihood of newly diagnosed T2D, prediabetes, general obesity and abdominal obesity compared to those with the least intakes (median (IQR) = 10 (7.2–12.6) g d<sup>-1</sup>; Extended Data Fig. 2g).

### Carbohydrate substitution and metabolic risk

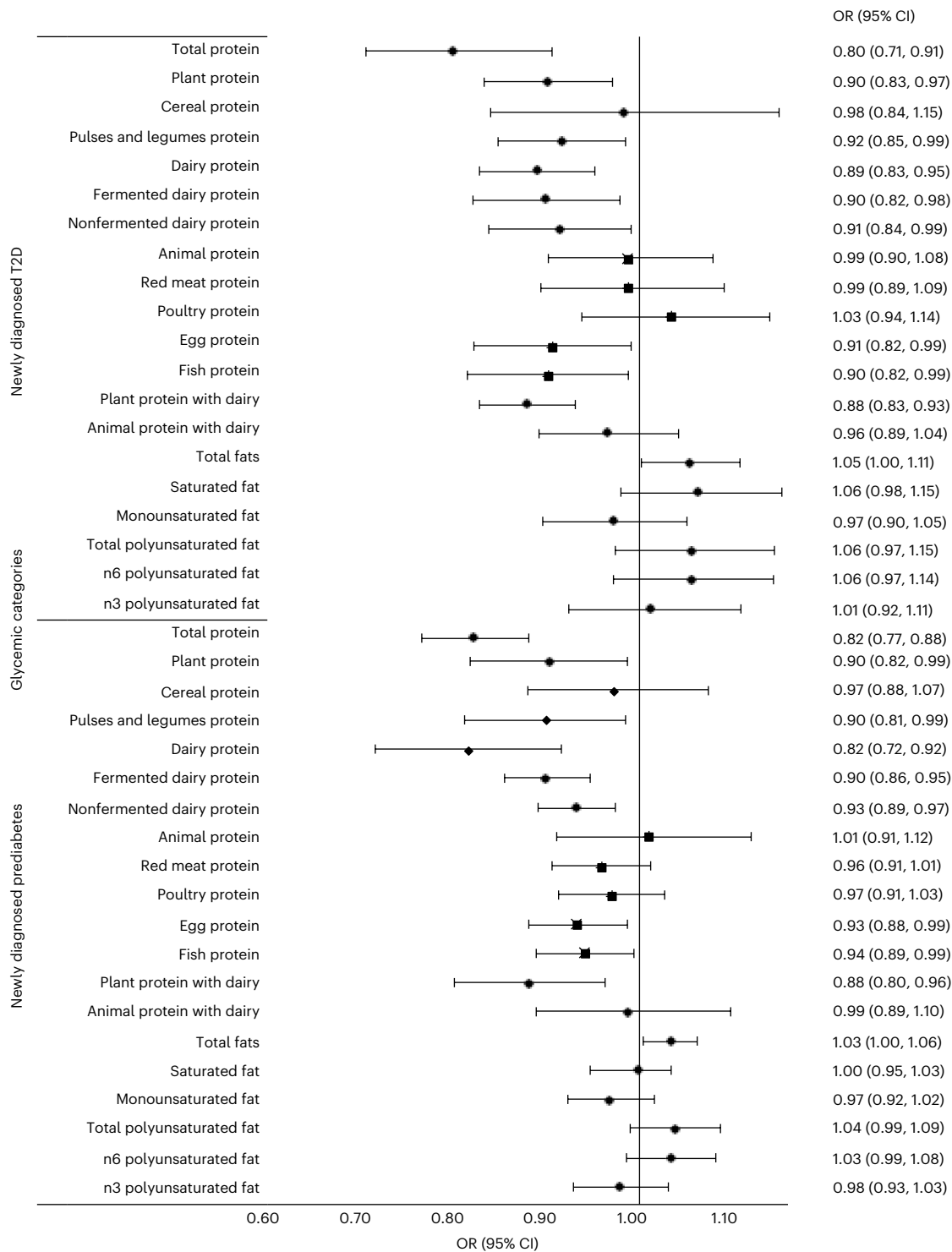
We mathematically modeled the effect of an isocaloric substitution of 5%E from carbohydrates with 5%E from other macronutrients on glycemic risk (Fig. 4). Replacing 5%E from carbohydrates with 5%E from protein, specifically from plants, pulses and legumes or dairy (both fermented and nonfermented), was associated with a lower likelihood of newly diagnosed T2D and prediabetes (Fig. 4). On the other hand, the replacement of carbohydrate with animal protein was not associated with a lower risk of newly diagnosed T2D or prediabetes except when the source of animal protein was fish or egg, where the risk was lower by 6–10%. Replacing 5%E from carbohydrates with 5%E from total fat or fat subtypes, including saturated fat, monounsaturated or n3 polyunsaturated fat did not alter the risk of newly diagnosed T2D or prediabetes (Fig. 4).

The replacement of carbohydrates with total protein or protein from various food sources was not associated with risk of general or abdominal obesity (Extended Data Table 6). However, replacing 5%E from carbohydrates with 5%E from total fat was associated with a 5% higher likelihood of general obesity (OR = 1.05, 95% CI = 1.02, 1.09; Extended Data Table 6). At the same time, replacing 50 g of refined cereals with 50 g milled whole grains or its subtypes (wheat and millet flour) was not associated with dysglycaemia or obesity (Extended Data Table 7). However, as the population's intake of intact whole grains (brown rice, whole wheat and millets) was limited, they could not be evaluated in this substitution model.

### Discussion

Using nationally representative data, we report high intakes (62.3%E) of low-quality carbohydrates (refined cereals like white rice, milled whole grains and added sugar) and saturated fat, and low protein intake. Millet consumption was high only in three states. Saturated fat intake exceeded national dietary recommendations (<10%E)<sup>18</sup> in ten states, while if metabolic risk threshold (<7%E) was used<sup>19</sup>, all but four states exceeded recommended limits. Total protein intake was low nationwide with plant protein contributing 8.9%E, dairy 2%E and animal protein 1%E. Carbohydrate intake (>56%E), regardless of grain type, was associated with a 14–30% higher metabolic risk. Replacing refined cereals with whole wheat or millet flour was not associated with lower risk. However, replacing 5%E from carbohydrates with 5%E from plant, dairy, egg or fish protein was associated with lower risk of newly diagnosed T2D and prediabetes, while replacement with fat of any subtype did not change risk.

Data from the current study along with previous surveys<sup>7,20</sup> confirm that Asian Indians consume high amounts of carbohydrates. Our results show that white rice is the most popular primary cereal staple (61% of population) followed by whole wheat flour (34%), with only a



**Fig. 4 | Odds ratios (95% CI) for glycemic risk in isocaloric substitution analysis for the replacement of 5% energy from carbohydrates for 5% energy from other macronutrients in the study population ( $n = 18,090$ ).** Two-sided  $P$  values were used in calculating ORs and 95% CIs through survey logistic regression. Models were adjusted for PSU, state code, population weight, age (in years), sex (male/female), family history of T2D (yes/no), tobacco use (never/current/past), alcohol use (never/current/past), PAL (sedentary/moderate/vigorous), time period the study was conducted, education qualification (primary school/high school/college), BMI ( $\text{kg m}^{-2}$ ), place of residence: urban versus rural, intakes of total energy ( $\text{kcal d}^{-1}$ ). It was further adjusted for hypertension, dyslipidaemia and abdominal obesity. Models were

additionally adjusted for all nutrients except the nutrient being replaced. Plant protein includes cereals, pulses, legumes, fruits, vegetables (including roots and tubers), nuts and oilseeds. Animal protein (%E) includes red meat, poultry, egg and fish, and does not include dairy protein. Dairy protein includes fermented dairy protein (Indian style yogurt (curd), buttermilk, lassi) and nonfermented dairy protein (all types of milk (cow/buffalo/yak/goat), milk added to coffee and tea, paneer (curdled milk solids)). Newly diagnosed T2D was defined as HbA1c  $\geq 6.5\%$  and/or if fasting CBG  $\geq 126 \text{ mg dl}^{-1}$  ( $\geq 7 \text{ mmol l}^{-1}$ ) or 2-h postglucose load CBG  $\geq 220 \text{ mg dl}^{-1}$  ( $\geq 12.2 \text{ mmol l}^{-1}$ ). Newly diagnosed prediabetes was defined as HbA1c  $5.7\text{--}6.4\%$  or if fasting CBG was  $\geq 100\text{--}125 \text{ mg dl}^{-1}$  ( $\geq 5.6\text{--}6.9 \text{ mmol l}^{-1}$ ) and/or if 2-h postload CBG value was  $\geq 160$  and  $< 220 \text{ mg dl}^{-1}$  ( $\geq 8.9$  to  $< 12.2 \text{ mmol l}^{-1}$ ).



small percentage using millet flour. Because of the Green Revolution (introduction of high-yielding varieties of wheat and rice in developing countries, leading to major increases in the production of these foodgrains) in the 1960s, there has been a decrease in millet consumption, and an increase in rice and wheat consumption<sup>21</sup>. Consistent with findings from the National Expenditure Survey 2011–2012 (ref. 8), we report that rural residents derive a greater proportion of total calories from cereals than urban residents. While refined cereals are the major source of carbohydrate calories (28.5%E) followed by milled whole grains (16.2%E), actual intakes may be higher, as processed foods, which often contain substantial amounts of refined cereals, were not included in our database.

India accounts for 15% of the global sugar consumption<sup>22</sup>. Our survey found that added sugar consumption was higher than the recommended 5%E in most states and exceeded levels reported in the NNMB survey ( $\approx 15 \text{ g d}^{-1}$ )<sup>23</sup>. The definition of added sugars in this survey included only sugar added at the table, during cooking and from products with added sugar on the product label. Therefore, actual intakes are likely much higher. Despite Food Safety and Standards Authority of India regulations<sup>24</sup> to label added sugars in the Nutrition Facts panel, many processed foods in the unregulated food sector (such as fast-food eateries, street vendors, local bakeries, fruit juices and traditional Indian sweets or mithais) do not label for added sugars.

Total fat intake aligned with the NNMB survey<sup>17</sup> and remained within the recommended  $\sim 30\%$ E<sup>19</sup>. However, most states exceeded the saturated fat recommendation ( $<7\%$ E)<sup>18</sup>, highlighting an urgent need for targeted public health interventions to lower intakes. While total PUFA intake was comparable with global levels<sup>25,26</sup>, MUFA intake was much lower in the United States and Europe<sup>26,27</sup>, likely due to higher consumption of nuts and red meat in Western diets. Total protein intake was low nationwide when compared to the United States<sup>25</sup>. Plant protein was higher in India compared to the West<sup>25</sup>, while animal protein was low due to low meat<sup>28</sup> and fish intake<sup>29</sup>. Conversely, dairy protein intake met National 'My plate for the day' recommendations<sup>17</sup> of  $\approx 2\%$ E. These low intakes of animal protein align with the recent National Family Health Survey-5 (ref. 30) data, which show that half of adults consumed pulses/beans and milk/curd daily, but fewer consumed animal foods daily.

The Global Burden of Disease Study demonstrated that, for South Asia,  $\sim 23\%$  of T2D burden was due to excess intake of white rice<sup>31</sup>. Although refined and whole grains differ in their associations with NCD risk<sup>32</sup>, higher intakes of milled whole grains (wheat and millets) were still associated with adverse metabolic profiles, showing no benefit for milled whole grains. In India, most whole grains are consumed as milled flour for flatbreads such as chapati and roti. Milling lowers the particle size of whole wheat and increases its glycemic index to the extent that the glycemic response becomes similar to that of refined wheat products and white rice<sup>32,33</sup>. In several randomized crossover trials<sup>34,35</sup> that examined the effects of whole grain processing, specifically milling, on glycemic control, finely milled whole wheat flour increased postprandial glycemic response and increased body weight. Given minimal intake of intact whole grains (brown rice, whole wheat and whole millets), public health recommendations should focus on lowering total carbohydrate intake from both refined and milled whole grains, which are the country's primary cereal staples.

Given the higher risk of T2D associated with refined cereal consumption<sup>6,36</sup>, identifying optimal macronutrient replacements is imperative. Replacing carbohydrates with protein from plant sources, dairy, eggs or fish was associated with lower T2D risk, unlike western studies where animal protein was associated with a higher T2D risk<sup>37,38</sup>. This likely reflects the predominance of plant protein in Indian diets. However, the quality of plant protein is equally important<sup>39</sup> because, in our study, protein from pulses and legumes, but not cereals, was associated with a lower risk of T2D. Replacing rapidly digested carbohydrates (refined grains) with legume protein may improve glycemic control and lower the overall glycemic index of the diet<sup>40</sup>, while also

addressing protein deficiencies<sup>39</sup>. Although literature surrounding dairy and egg consumption and metabolic risk has been inconsistent<sup>41–43</sup>, replacing low-quality carbohydrates with dairy, egg or fish can mitigate metabolic risk in the Asian Indian population due to high-carbohydrate intakes and low protein intakes. In fact, low yogurt intake alone accounts for 13% of South Asia's T2D burden<sup>31</sup>. The International Lipid Expert Panel recommends increase in protein intake from poultry and fish for cardiometabolic risk reduction<sup>44</sup>. Another multi-country study has shown benefits with weekly two servings of fish (175 g per week) for individuals with cardiometabolic risk<sup>45</sup>. While red meat has been previously associated with higher NCD risk<sup>46</sup>, we found that red meat protein was not associated with likelihood of T2D or prediabetes possibly due to low consumption.

While isocaloric substitutions are useful for understanding the effects of replacing one nutrient with another while keeping total energy intake constant, it is important to recognize that this approach involves statistical modeling rather than actual replacements by participants. Diet is a complex matrix of various foods and nutrients that interact synergistically or antagonistically, and simple isocaloric substitutions may not capture these interactions. We were unable to evaluate the effect of replacing carbohydrates with foods other than plant protein, such as intact whole grains, fruits and vegetables, which are also high in fiber and have been shown to decrease the risk of NCDs<sup>47</sup>. Further studies should investigate this in the Asian Indian population to inform more comprehensive nutritional policy recommendations.

The current nationwide high-carbohydrate intake represents a major risk factor for NCDs and should prompt policy action to lower consumption. Widespread reliance on subsidized rice and wheat likely drives high cereal intake, highlighting the need to prioritize pulses and legumes in subsidy programs over refined cereals<sup>48</sup>. Excess added sugar further endangers public health, and current tax rates on carbonated drinks have not reduced consumption<sup>49</sup>, suggesting the need for alternative strategies. Saturated fat intake also exceeds recommendations in many states, likely due to high intakes of ghee (clarified butter) in the North and the use of subsidized palm oil and coconut oil in the South. In addition, palmolein oil, due to its lower cost compared to other vegetable oils, is widely used in the food industry and in processed foods<sup>50</sup>. Policies should shift toward provision of healthier oils through the public distribution system (PDS), and raising consumer awareness to support healthier choices.

Our study has several strengths. This is a nationally representative study that highlights similarities and differences in the usual intakes of macronutrients between and within various regions of India. We used a validated food frequency questionnaire (FFQ)<sup>51</sup> to assess diet across regions. We also explored a strategy to reduce metabolic risk by modeling isocaloric substitution of protein and its subtypes for carbohydrates. The findings from this study can be extrapolated to other countries in South and Southeast Asia with similar dietary patterns and where high-carbohydrate diets from cereal staples are prevalent.

Our findings should be interpreted considering several limitations. First, this was a cross-sectional study, and therefore, reverse causation is a possibility. However, because our primary outcomes were newly diagnosed T2D and prediabetes, participants were unaware of their status when dietary information was collected. Second, a single measure of glycaemia (glycated hemoglobin (HbA1c) or blood glucose) was used to define T2D and prediabetes, although two measurements are recommended for making a clinical diagnosis. However, for epidemiological studies, a single measure is considered acceptable and often the only feasible option. The use of the American Diabetes Association's<sup>52</sup> less stringent cutoffs for fasting blood glucose ( $100 \text{ g d}^{-1}$ ) for prediabetes could have selected more individuals prone to reversion to normoglycemia compared to the higher cut-point advocated by the WHO ( $110 \text{ mg dl}^{-1}$ )<sup>53</sup>. Third, we were unable to capture intakes of processed foods and *trans*-fat content of commercial products due to lack of a *trans*-fat database and labeling regulations. Fourth, although we

adjusted for potential confounders, residual confounding is possible due to the observational design. Fifth, we did not include medication use in defining dyslipidaemia and hypertension; therefore, actual prevalence of these risk factors might be higher than reported. Sixth, because our analyses were prespecified, findings were not adjusted for multiple comparisons, potentially increasing the likelihood of a type I error. However, given the mechanistic links among our outcomes, a type I error is unlikely. Seventh, although the FFQ used in INDIAB was validated, measurement errors are likely. Because these errors are nondifferential with respect to the outcome, the associations are likely to be attenuated. Eighth, as the dietary data were collected from every fifth participant, the included population may not fully represent the larger population. However, demographic differences between included and excluded participants were minimal and not clinically relevant, suggesting a good external validity.

In conclusion, based on our findings, national guidelines and policy changes should emphasize reduced consumption of refined carbohydrates and saturated fat, and recommend increased consumption of protein, primarily from plant and dairy sources. Public health messaging should highlight practical strategies such as gradually increasing the ratio of pulses, legumes and dairy in the diet to achieve a desirable macronutrient composition for NCD prevention. Such dietary shifts can reverse the nutrition transition, address protein inadequacies and improve overall diet quality. Subsidies through the PDS should emphasize healthier protein sources (pulses and legumes) and healthier edible oils (low in saturated fat) over refined grains like white rice. Minimum support prices for cereals, if extended to pulses, legumes and dairy can improve both the quality and the quantity of protein consumed. Because healthcare is a state government responsibility in India, the study's findings are vital for helping states to enhance their PDS plans and to promote subsidies that prevent NCDs. A multisectoral approach involving healthcare, agriculture and socioeconomic policy is crucial for fostering healthier diets and lowering NCD risk in India.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-025-03949-4>.

## References











- Anjana, R. M. et al. Metabolic non-communicable disease health report of India: the ICMR-INDIAB national cross-sectional study (ICMR-INDIAB-17). *Lancet Diabetes Endocrinol.* **11**, 474–489 (2023).
- World Health Organization. Noncommunicable diseases data portal. [ncdportal.org/CountryProfile/GHE110/IND](http://ncdportal.org/CountryProfile/GHE110/IND) (accessed 24 March 2025).
- Okunogbe, A. et al. Economic impacts of overweight and obesity: current and future estimates for 161 countries. *BMJ Glob. Health* **7**, e009773 (2022).
- Anjana, R. M. et al. Diabetes in Asian Indians—how much is preventable? Ten-year follow-up of the Chennai Urban Rural Epidemiology Study (CURES-142). *Diabetes Res. Clin. Pract.* **109**, 253–261 (2015).
- Schmidhuber, J. et al. The Global Nutrient Database: availability of macronutrients and micronutrients in 195 countries from 1980 to 2013. *Lancet Planet. Health* **2**, e353–e368 (2018).
- Mohan, V. et al. Dietary carbohydrates, glycaemic load, food groups and newly detected type 2 diabetes among urban Asian Indian population in Chennai, India (Chennai Urban Rural Epidemiology Study 59). *Br. J. Nutr.* **102**, 1498–1506 (2009).
- Hemalatha, R., Laxmaiah, A., Sriswan, M. R., Boiraju, N. V. & Radhakrishna, K. V. What India eats. ICMR-NIN Report. [https://www.millet.res.in/pdf/what\\_india\\_eats.pdf](https://www.millet.res.in/pdf/what_india_eats.pdf) (2020).
- National Sample Survey Organisation. Nutritional Intake in India, 2011–12—NSS 68th round (July 2011–June 2012): household consumer expenditure. NSS Report 560. [https://www.mospi.gov.in/sites/default/files/publication\\_reports/nss\\_report\\_560\\_19dec14.pdf](https://www.mospi.gov.in/sites/default/files/publication_reports/nss_report_560_19dec14.pdf) (2014).
- Anjana, R. M. et al. Macronutrient recommendations for remission and prevention of diabetes in Asian Indians based on a data-driven optimization model: the ICMR-INDIAB national study. *Diabetes Care* **45**, 2883–2891 (2022).
- Aune, D., Norat, T., Romundstad, P. & Vatten, L. J. Whole grain and refined grain consumption and the risk of type 2 diabetes: a systematic review and dose–response meta-analysis of cohort studies. *Eur. J. Epidemiol.* **28**, 845–858 (2013).
- Hu, F. B., Van Dam, R. M. & Liu, S. Diet and risk of type II diabetes: the role of types of fat and carbohydrate. *Diabetologia* **44**, 805–817 (2001).
- Ardakani, A. F. et al. Association between total, animal, and plant protein intake and type 2 diabetes risk in adults: a systematic review and dose–response meta-analysis of prospective cohort studies. *Clin. Nutr.* **43**, 1941–1955 (2024).
- Inan-Eroglu, E. et al. Association between protein intake and diabetes complications risk following incident type 2 diabetes: the EPIC-Potsdam study. *Metabolites* **14**, 172 (2024).
- Estruch, R. et al. Primary prevention of cardiovascular disease with a Mediterranean diet supplemented with extra-virgin olive oil or nuts. *N. Engl. J. Med.* **378**, e34 (2018).
- Human energy requirements: report of a joint FAO/WHO/UNU expert consultation. *Food Nutr. Bull.* **26**, 166 (2005).
- Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults Executive summary of the third report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (adult treatment panel III). *JAMA* **285**, 2486–2497 (2001).
- ICMR-National Institute of Nutrition. Dietary guidelines for Indians. <https://www.nin.res.in/dietaryguidelines/pdfs/locale/DGI07052024P.pdf> (2024).
- ICMR-National Institute of Nutrition. Nutrient requirement for Indians, recommended dietary allowances (RDA) and estimated average requirements (EAR). <https://www.scribd.com/document/526073382/RDA-2020-Full-Ver> (2020).
- Virani, S. S. et al. AHA/ACC/ACCP/ASPC/NLA/PCNA guideline for the management of patients with chronic coronary disease: a report of the American Heart Association/American College of Cardiology Joint Committee on clinical practice guidelines. *Am. Coll. Cardiol.* **82**, 833–955 (2023).
- Radhika, G. et al. Dietary profile of urban adult population in South India in the context of chronic disease epidemiology (CURES-68). *Public Health Nutr.* **14**, 591–598 (2011).
- Singh, O. P. & Singh, P. K. Trends in millets production, consumption and export from India. *J. Res. ANGRAU* **52**, 121–133 (2024).
- Ministry of Consumer Affairs, Food & Public Distribution. India becomes Chair of International Sugar Organisation (ISO) for 2024 to lead global sugar sector. <https://www.pib.gov.in/PressReleasePage.aspx?PRID=1979507> (2023).
- ICMR-National Institute of Nutrition. Sugar intake and its sources in India: the first national survey report. [https://www.nin.res.in/survey\\_reports/sugar\\_study\\_report\\_part-1.pdf](https://www.nin.res.in/survey_reports/sugar_study_report_part-1.pdf) (2019).
- Food Safety and Standards Authority of India (FSSAI). Compendium of labelling requirements under FSS Act, 2006. [https://www.fssai.gov.in/upload/uploadfiles/files/Compendium\\_Labelling\\_Display\\_23\\_09\\_2021.pdf](https://www.fssai.gov.in/upload/uploadfiles/files/Compendium_Labelling_Display_23_09_2021.pdf) (2021).
- Shan, Z. et al. Trends in dietary carbohydrate, protein, and fat intake and diet quality among US adults, 1999–2016. *JAMA* **322**, 1178–1187 (2019).

26. Harika, R. K. et al. Intake of fatty acids in general populations worldwide does not meet dietary recommendations to prevent coronary heart disease: a systematic review of data from 40 countries. *Ann. Nutr. Metab.* **63**, 229–238 (2013).
27. Linseisen, J. et al. Dietary fat intake in the European prospective investigation into cancer and nutrition: results from the 24-h dietary recalls. *Eur. J. Clin. Nutr.* **63**, S61–S80 (2009).
28. Bruinsma, J. *World Agriculture: Towards 2015/2030—An FAO Study* (Routledge, 2017).
29. Padiyar, A. P. et al. *Fish Consumption in India: Patterns and Trends* (WorldFish, 2024).
30. IIPS & ICF. *National Family Health Survey (NFHS-5), 2019–21*, Vol. II (IIPS, 2021).
31. O’Hearn, M. et al. Incident type 2 diabetes attributable to suboptimal diet in 184 countries. *Nat. Med.* **29**, 982–995 (2023).
32. Åberg, S. et al. Whole-grain processing and glycemic control in type 2 diabetes: a randomized crossover trial. *Diabetes Care* **43**, 1717–1723 (2020).
33. Shobana, S. et al. Glycaemic index of three Indian rice varieties. *Int. J. Food Sci. Nutr.* **63**, 178–183 (2012).
34. Elbalshy, M. M. et al. Gelatinisation and milling whole-wheat increases postprandial blood glucose: randomised crossover study of adults with type 2 diabetes. *Diabetologia* **64**, 1385–1388 (2021).
35. Reynolds, A. N. et al. Wholegrain particle size influences postprandial glycemia in type 2 diabetes: a randomized crossover study comparing four wholegrain breads. *Diabetes Care* **43**, 476–479 (2020).
36. Bhavadharini, B. et al. White rice intake and incident diabetes: a study of 132,373 participants in 21 countries. *Diabetes Care* **43**, 2643–2650 (2020).
37. Li, J. et al. Dietary protein sources, mediating biomarkers, and incidence of type 2 diabetes: findings from the women’s health initiative and the UK Biobank. *Diabetes Care* **45**, 1742–1753 (2022).
38. Malik, V. S. et al. Dietary protein intake and risk of type 2 diabetes in US men and women. *Am. J. Epidemiol.* **183**, 715–728 (2016).
39. Swaminathan, S., Vaz, M. & Kurpad, A. V. Protein intakes in India. *Br. J. Nutr.* **108**, S50–S58 (2012).
40. Venn, B. J. & Mann, J. I. Cereal grains, legumes and diabetes. *Eur. J. Clin. Nutr.* **58**, 1443–1461 (2004).
41. Feng, Y. et al. Consumption of dairy products and the risk of overweight or obesity, hypertension, and type 2 diabetes mellitus: a dose–response meta-analysis and systematic review of cohort studies. *Adv. Nutr.* **13**, 2165–2179 (2022).
42. O’Connor, L. M. et al. Dietary dairy product intake and incident type 2 diabetes: a prospective study using dietary data from a 7-day food diary. *Diabetologia* **57**, 909–917 (2014).
43. Drouin-Chartier, J. P. et al. Egg consumption and risk of cardiovascular disease: three large prospective US cohort studies, systematic review, and updated meta-analysis. *Brit. Med. J.* **368**, m513 (2020).
44. Zhubi-Bakija, F. et al. The impact of type of dietary protein, animal versus vegetable, in modifying cardiometabolic risk factors: a position paper from the International Lipid Expert Panel (ILEP). *Clin. Nutr.* **40**, 255–276 (2021).
45. Mohan, D. et al. Associations of fish consumption with risk of cardiovascular disease and mortality among individuals with or without vascular disease from 58 countries. *JAMA Intern. Med.* **181**, 631–649 (2021).
46. Grosso, G. et al. Total, red and processed meat consumption and human health: an umbrella review of observational studies. *Int. J. Food Sci. Nutr.* **73**, 726–737 (2022).
47. Kelly, R. K. et al. Associations between types and sources of dietary carbohydrates and cardiovascular disease risk: a prospective cohort study of UK Biobank participants. *BMC Med.* **21**, 34 (2023).
48. Kishore, A. & Chakrabarti, S. Is more inclusive more effective? The ‘New Style’ public distribution system in India. *Food Policy* **55**, 117–130 (2015).
49. Law, C. et al. Changes in take-home aerated soft drink purchases in urban India after the implementation of Goods and Services Tax (GST): an interrupted time series analysis. *SSM Popul. Health* **14**, 100794 (2021).
50. Gulati, S. & Misra, A. Abdominal obesity and type 2 diabetes in Asian Indians: dietary strategies including edible oils, cooking practices and sugar intake. *Eur. J. Clin. Nutr.* **71**, 850–857 (2017).
51. Sudha, V. et al. Reproducibility and construct validity of a food frequency questionnaire for assessing dietary intake in rural and urban. *Asia Pac. J. Clin. Nutr.* **29**, 192–204 (2020).
52. American Diabetes Association. 2. Classification and diagnosis of diabetes: Standards of Medical Care in Diabetes—2021. *Diabetes Care* **44**, S15–S33 (2021).
53. World Health Organization. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation. <https://www.who.int/publications/i/item/definition-and-diagnosis-of-diabetes-mellitus-and-intermediate-hyperglycaemia> (2006).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

**Ranjit Mohan Anjana** <sup>1,2†</sup>, **Vasudevan Sudha**<sup>2,21</sup>, **Kuzhandaivelu Abirami**<sup>2</sup>, **Rajagopal Gayathri** <sup>2</sup>, **Valangaiman Sriram Manasa** <sup>2</sup>, **Mohan Deepa**<sup>3</sup>, **Rajendra Pradeepa** <sup>4</sup>, **Ranjit Unnikrishnan** <sup>1</sup>, **Shashank Joshi**<sup>5</sup>, **Banshi Saboo**<sup>6</sup>, **Arvind Gupta** <sup>7</sup>, **Prashant P. Joshi**<sup>8</sup>, **Prabha Adhikari**<sup>9</sup>, **Puthiyaveetil Kottayam Jabbar**<sup>10</sup>, **Sunil M. Jain**<sup>11</sup>, **Subhankar Chowdhury**<sup>12</sup>, **Anil J. Purty**<sup>13</sup>, **Saroj Kumar Tripathy**<sup>14</sup>, **Sarita Behera**<sup>15</sup>, **Jagdish Mahanta**<sup>16</sup>, **Avula Laxmaiah**<sup>17</sup>, **Kamala Krishnaswamy**<sup>2</sup>, **Anura Viswanath Kurpad** <sup>18</sup>, **Ashok Kumar Das**<sup>19</sup>, **Rupinder Singh Dhaliwal** <sup>20</sup>, **Tanvir Kaur**<sup>20</sup>, **Shilpa N. Bhupathiraju** <sup>2,22</sup> & **Viswanathan Mohan** <sup>1,22</sup>



<sup>1</sup>Department of Diabetology, Madras Diabetes Research Foundation, Chennai, India. <sup>2</sup>Department of Foods, Nutrition and Dietetics Research, Madras Diabetes Research Foundation, Chennai, India. <sup>3</sup>Department of Epidemiology, Madras Diabetes Research Foundation, Chennai, India. <sup>4</sup>Department of Research Operations and Diabetes Complications, Madras Diabetes Research Foundation, Chennai, India. <sup>5</sup>Department of Diabetology and Endocrinology, Lilavati Hospital and Research Centre, Mumbai, India. <sup>6</sup>Dia Care-Diabetes Care and Hormone Clinic, Ahmedabad, India. <sup>7</sup>Department of Diabetes, Obesity and Metabolic Disorders, Rajasthan Hospital, Jaipur, India. <sup>8</sup>Department of General Medicine, All India Institute of Medical Sciences, Nagpur, India. <sup>9</sup>Department of Epidemiology, Yenepoya Medical College, Yenepoya University Campus, Deralakatte, India. <sup>10</sup>Department of Endocrinology, Government Medical College, Trivandrum, India. <sup>11</sup>TOTALL Diabetes Thyroid Hormone Research Institute Pvt. Ltd., Indore, India. <sup>12</sup>Department of Endocrinology and Metabolism, Institute of Post Graduate Medical Education and Research (IPGMER) and SSKM Hospital, Kolkata, India. <sup>13</sup>Pondicherry Institute of Medical Sciences, Puducherry, India. <sup>14</sup>S.C.B. Medical College and Hospital, Cuttack, India. <sup>15</sup>Department of Medicine, Bhima Bhoi Medical College and Hospital, Balangir, India. <sup>16</sup>Regional Medical Research Centre, Dibrugarh, India. <sup>17</sup>Division of Public Health Nutrition, National Institute of Nutrition, Indian Council of Medical Research, Hyderabad, India. <sup>18</sup>Department of Physiology, St. John's Medical College, Bengaluru, India. <sup>19</sup>Department of Medicine, Mahatma Gandhi Medical College and Research Institute, Puducherry, India. <sup>20</sup>Non-Communicable Disease Division, Indian Council of Medical Research, New Delhi, India. <sup>21</sup>These authors contributed equally: Ranjit Mohan Anjana, Vasudevan Sudha. <sup>22</sup>These authors jointly supervised this work: Shilpa N. Bhupathiraju, Viswanathan Mohan. ✉ e-mail: [dranjana@drmhans.com](mailto:dranjana@drmhans.com)

## Methods

### Sampling and study population

The ICMR–INDIAB national study is a cross-sectional door-to-door population-based survey that includes a representative sample of individuals aged  $\geq 20$  years drawn from 30 states and union territories, and the NCT of Delhi. To be nationally representative, the survey was conducted in multiple phases using a stratified multistage sampling design using three-level stratification based on geography, population size and socioeconomic status of each state. In phase 1 (2008–2010), individuals from Tamil Nadu, Chandigarh, Jharkhand and Maharashtra were recruited. Phase 2 was conducted from 2012 to 2013, and included undivided Andhra Pradesh, Bihar, Gujarat, Karnataka and Punjab. Phase 3 (2017–2018) included Delhi, Madhya Pradesh, Rajasthan and Uttar Pradesh, while phase 4 (2018–2019) included Kerala, Goa, Puducherry, Haryana and Chhattisgarh. The Northeast phase was conducted between 2011 and 2017 and included states of Assam, Arunachal Pradesh, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim and Tripura. The final phase (phase 5) was conducted between 2019 and 2020 and covered Himachal Pradesh, Uttarakhand, Odisha and West Bengal. Additional methodological details of the ICMR–INDIAB survey have been published earlier<sup>1</sup>. The primary sampling units (PSUs) were villages in rural areas and census enumeration blocks in urban areas. Using a systematic sampling method described elsewhere<sup>1</sup>, from each PSU, 24 households from urban areas and 56 households from rural areas were selected. To avoid selection bias with respect to age and sex, one individual was selected from each household using the WHO Kish method<sup>54</sup>.

The total sample size, including all 30 states, union territories and the NCT of Delhi, was 113,043 adults, with 79,506 from rural areas and 33,537 from urban areas. Of the 113,043 individuals, every fifth individual ( $n = 24,800$ ) and all those with known and newly diagnosed diabetes were invited to provide information on diet. A total of 20,860 individuals provided information on diet (response rate 91.8%). The final dataset for this study comprised 18,090 adults (49% male, 27% urban), for whom all biochemical and dietary assessments were available for further analysis.

### Ethics

The Institutional Ethics Committee at Madras Diabetes Research Foundation (MDRF) granted approval for the study, and written informed consent was obtained from all study participants. The study was registered with the Clinical Trials Registry of India (CTRI/2019/03/018095). The study was conducted in accordance with the principles outlined in the Declaration of Helsinki for medical research.

### Study procedures

Information pertaining to demographic and socioeconomic characteristics of all participants was collected using a standardized, structured questionnaire. Weight, height and waist circumference were measured using standardized procedures. BMI was calculated as weight in kilograms divided by height in meters squared. Blood pressure was recorded in the sitting position twice with an interval of 5 min. Measurements were taken in the right arm to the nearest 1 mm Hg, using digital sphygmomanometers (Omron HEM-7101, Omron). The average of the two readings was used as the final reading. For quality assurance, equipment with the same specifications were used throughout the study. Interobserver and intra-observer coefficients of variation between the field technicians were documented to be less than 5% in all regions.

### Biochemical assessments

Biochemical assessments were performed using standardized methods by the same team of laboratory technicians throughout the study. In individuals without self-reported T2D, we conducted an oral glucose tolerance test using capillary blood glucose (One Touch Ultra, LifeScan, Johnson & Johnson). Fasting venous blood samples were drawn

for assessment of glycosylated hemoglobin (HbA1c) in every fifth participant. Venous samples were centrifuged within 1 h at the survey site and serum was transferred to labeled vials and temporarily stored in  $-20^{\circ}\text{C}$  freezers before being sent to the central laboratory. HbA1c was measured with high-pressure liquid chromatography using the Variant II Turbo machine (Bio-Rad, Hercules), which is certified by the National Glycohemoglobin Standardization Program as having documented traceability to the Diabetes Control and Complications Trial reference method<sup>3</sup>. Serum cholesterol (cholesterol esterase oxidase–peroxidase–amidopyrine method), serum triglycerides (glycerol phosphate oxidase–peroxidase–amidopyrine method) and high-density lipoprotein cholesterol (direct method–immune inhibition) were measured using the Olympus 2700/480 automated biochemistry analyzer (Fullerton) from 2008 to 2015, and the Beckman Coulter AU 680 clinical chemistry analyzer (Fullerton) from 2016 to 2021. LDL-C was calculated using the Friedwald equation<sup>55</sup>. The intra-assay and interassay coefficients of variation for biochemical assays conducted at the central laboratory ranged between 3.1% and 7.6%.

### Assessment of diet

Diet was assessed using an interviewer-administered questionnaire, MDRF–FFQ<sup>51</sup>. The MDRF–FFQ is a meal-based, quantitative FFQ that contains a list of 222 food items normally consumed by both urban and rural populations. Participants were instructed to indicate usual frequency of consumption for each food item (number of times per day, week, month, year or never) and their usual serving size, reflecting their dietary patterns over the past year, using one of the five following categories: ‘never’, ‘daily’, ‘weekly’, ‘monthly’ and ‘yearly’. To precisely estimate portion sizes and tools, participants were provided with a visual atlas of Indian foods. In addition, common household measures, such as household cups, bowls, ladles, spoons, wedges and circles of different diameters, were shown.

The MDRF–FFQ has been previously validated for adults in urban and rural areas across ten states in five regions of India<sup>51</sup>. The FFQ was administered to adults twice a year (at 0 and 12 months). Reproducibility, assessed using intraclass correlation coefficients, showed moderate to good correlations for nutrients and food groups (for example, 0.50–0.77 for saturated fat and energy in urban areas, and 0.61–0.72 for protein and saturated fat in rural areas). Intraclass correlation coefficients for foods such as whole grains (urban, 0.53; rural, 0.61) and fruits and vegetables ( $>0.7$ ) were also high, indicating good reproducibility. We also assessed the construct validity of carbohydrate and fat intake using serum lipids in regression analyses. The de-attenuated Pearson correlation for the energy-adjusted nutrients between FFQ and 24-h recalls ranged from 0.73 for carbohydrates to 0.35 for calcium.

The average daily intake of calories, macronutrients, dietary fiber, fatty acids and food groups (excluding nutrient supplements like protein powder) was computed using the ‘EpiNu’ Software (version 2.0). EpiNu consists of a wide range of recipes gathered from various places (ranging from households to commercial eating places such as fast-food outlets and popular restaurants) to be more realistic as to the actual eating habits of the population. The food groups were categorized based on the food items reported by the participants in the FFQ.

### Calculation of %E for macronutrients

The %E for each individual was calculated as follows: carbohydrates in  $\text{g} \times 4 \text{ calories} / \text{total daily calories} \times 100$ ; protein in  $\text{g} \times 4 \text{ calories} / \text{total daily calories} \times 100$ ; and fat in  $\text{g} \times 9 \text{ cal} / \text{total daily calories} \times 100$ .

### Calculation of %E food groups

- For the purpose of this analysis, %E from refined cereals included meals/snacks prepared with white rice, refined wheat flour (white flour/maida), refined millet grains/flour, wheat semolina (rava) and vermicelli.

- The %E from milled whole grains included calories from whole wheat and whole millets used in flour or grit form.
- The %E from added sugar included white sugar, honey, jaggery and palm sugar added during cooking or processing and on the table. Added sugar from processed foods was not included in this definition due to lack of adequate information.
- The %E from plant protein included protein calories derived from cereals (whole and refined), pulses, legumes, nuts and oilseeds.
- The %E from dairy protein included protein calories derived from milk and its derivatives.
- The %E from fermented dairy included Indian style yogurt (curd), buttermilk and lassi.
- The %E from nonfermented dairy included all types of milk (cow/buffalo/yak/goat), milk added to coffee and tea, paneer (curdled milk solids).
- The %E from animal protein included protein calories derived from meat, poultry, eggs, fish and seafoods.

### Physical activity assessment

The MDRF Physical Activity Questionnaire captures data from multiple activity domains over the period of a year from males and females of varying ages and has been validated in both urban and rural settings, covering ten states in different regions<sup>56</sup>. This interviewer-administered questionnaire documents the frequency and duration of various activities, encompassing regular, obligatory and discretionary tasks across all domains. These domains include occupation, general activities (like sleep, personal care and household chores), transportation and leisure activities. The calculation of physical activity for an ‘average’ day was derived from a sum of the activities in various domains for a 24-h period. Total energy expenditure (TEE) was calculated as the time spent on various activities in the multiple domains and the energy cost of these activities. Energy cost was reported as a multiple of basal metabolic rate (BMR) and called physical activity ratio. Total time spent on habitual activities was multiplied by the physical activity ratio to derive the TEE of 24 h. The PAL was then calculated as TEE/BMR for 24 h (ref. 15). Participants were categorized as physically active or inactive. Moderate and vigorous PAL were combined for the ‘active’ category based on PAL cutoffs for moderate (1.70–1.99) and vigorous (2.0–2.4) activities, while physical inactivity was defined as sedentary PAL of 1.40–1.69 (ref. 15).

### Assessment of metabolic risk

Newly diagnosed T2D was defined as HbA1c  $\geq 6.5\%$  and/or fasting (capillary) blood glucose  $\geq 126 \text{ mg dl}^{-1}$  ( $\geq 7 \text{ mmol l}^{-1}$ ) or 2-h postglucose load (capillary) blood glucose  $\geq 220 \text{ mg dl}^{-1}$  ( $\geq 12.2 \text{ mmol l}^{-1}$ )<sup>1,52,53,57</sup>. Newly diagnosed prediabetes was defined as an HbA1c 5.7–6.4% or if fasting glucose was  $\geq 100$ – $125 \text{ mg dl}^{-1}$  ( $\geq 5.6$ – $6.9 \text{ mmol l}^{-1}$ ) and/or if 2-h postload glucose (capillary) value was above  $160 \text{ mg dl}^{-1}$  but  $< 220 \text{ mg dl}^{-1}$  ( $\geq 8.9$  to  $< 12.2 \text{ mmol l}^{-1}$ )<sup>52,53,57</sup>. Generalized obesity was defined using Asia-specific cutoffs of BMI of  $25 \text{ kg m}^{-2}$  or higher. Abdominal obesity was defined based on the WHO Asia Pacific guidelines, with a waist circumference of 90 cm or greater for men and 80 cm or greater for women<sup>58</sup>.

Hypertension was defined as a systolic blood pressure of  $\geq 140 \text{ mm Hg}$ , or a diastolic blood pressure of  $\geq 90 \text{ mm Hg}$  (ref. 59). Dyslipidaemia was defined using the National Cholesterol Education Program Adult Treatment Panel III guidelines as follows: serum cholesterol concentrations of  $\geq 200 \text{ mg dl}^{-1}$  ( $\geq 5.2 \text{ mmol l}^{-1}$ ), serum triglyceride concentrations of  $\geq 150 \text{ mg dl}^{-1}$  ( $\geq 1.7 \text{ mmol l}^{-1}$ ), low high-density lipoprotein cholesterol (male,  $< 40 \text{ mg dl}^{-1}$  ( $1.04 \text{ mmol l}^{-1}$ ); female,  $< 50 \text{ mg dl}^{-1}$  ( $1.3 \text{ mmol l}^{-1}$ )) and high low-density lipoprotein cholesterol concentrations of  $\geq 130 \text{ mg dl}^{-1}$  ( $3.4 \text{ mmol l}^{-1}$ )<sup>16</sup>. Metabolic risk was defined as the presence of any one of newly diagnosed T2D, prediabetes, dyslipidaemia, generalized obesity, abdominal obesity or hypertension.

### Statistical analysis

For the purpose of this analysis, the Indian states were divided into the following six geographical zones: North (Chandigarh, Delhi, Haryana, Himachal Pradesh, Punjab and Rajasthan), South (Andhra Pradesh (undivided-subsequently bifurcated into Andhra Pradesh and Telangana), Karnataka, Kerala, Puducherry, and Tamil Nadu), East (Bihar, Jharkhand, Odisha and West Bengal), West (Goa, Gujarat and Maharashtra), Central (Chhattisgarh, Madhya Pradesh, Uttar Pradesh and Uttarakhand) and Northeast (Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim and Tripura). Sensitivity analysis on the dietary profile of undivided Andhra Pradesh did not differ from the separate states of Telangana and Andhra Pradesh (Supplementary Table 4), and hence in this study, undivided Andhra Pradesh (matching with the actual data collection period of 2012–2013) was considered. To account for the multistage complex survey design of the study, to ensure that our results are representative of the larger population and to compensate for the unequal probability of selection, our analyses adjusted for sampling weights using the PROC SURVEY (frequency/mean) procedure, where all key survey elements were used in statistical analyses. The PSU was accounted for as the cluster, the normalized weight was accounted for as the final study weight and the state was accounted for as the stratum to estimate population means, medians, variance and proportions.

We excluded participants with self-reported T2D ( $n = 1,404$ ), HbA1c of  $< 4\%$  ( $n = 26$ ) or if they had implausible dietary data defined as a reported energy intake ( $n = 1,340$  with values  $< 500$  and  $> 4,200 \text{ kcal d}^{-1}$ )<sup>60</sup>. After these exclusions, the final analytic sample consisted of 18,090 participants (Extended Data Fig. 3). Complete data were available for 90% of the study participants. Missing data were considered to be missing completely at random. To impute missing data, we used information on age, sex and BMI to identify the nearest neighbor and applied the  $k$ -nearest neighbor imputation method<sup>61</sup> using Python package (version 3.11.7). Given the pervasive presence of added sugar in foods and beverages, for analyses involving added sugar as an exposure, we excluded participants who consumed  $\leq 5 \text{ g d}^{-1}$  of added sugar. For analyses on saturated fat, we excluded implausible values (participants with  $> 30\%$ E coming from saturated fat,  $n = 80$ ).

Intakes of major macronutrients from different food sources were expressed as a percentage of total energy using the residual method<sup>60</sup>. We evaluated differences in intakes of carbohydrate and its major food sources (refined cereals including white rice, refined wheat flour (maida), refined millet grains/flour (minor millets like little millet and foxtail millet) and wheat semolina (rava)); milled whole grains including whole wheat, whole millets and whole grains milled into flour and added sugar including white sugar, honey, jaggery and palm sugar added during cooking and on the table), total protein and its subtypes (plant protein including cereals, pulses, legumes, fruits, vegetables (including roots and tubers), nuts and oilseeds; dairy protein including fermented dairy protein (Indian style yogurt (curd), buttermilk, lassi) and nonfermented dairy protein (all types of milk (cow/buffalo/yak/goat), milk added to coffee and tea, paneer (curdled milk solids)); animal protein including red meat, poultry, fish, egg but not dairy protein) and total fat and its subtypes (saturated fat, monounsaturated fat, polyunsaturated fat, n6 polyunsaturated fat, n3 polyunsaturated fat) by place of residence (urban versus rural), sex, geographical region (North, South, West, East, Central and Northeast) and within geographical regions). Because dietary data were skewed, we presented weighted medians and IQRs when describing intakes. All weighted descriptive statistics were tested for significance by regions, urban versus rural, male versus female using median and quantile test for continuous variables and chi-square test for categorical weighted variables. The map (Fig. 1) represents the top three cereal grains reported by the weighted percentage of adults in each state.

We used PROC SURVEY LOGISTIC REGRESSION to evaluate the association between carbohydrate (%E) and metabolic risk. We also



examined associations between refined cereals ( $\text{g d}^{-1}$ ), milled whole grains ( $\text{g d}^{-1}$ ), added sugar (%E) and metabolic risk. For this analysis, we excluded nonconsumers and those with intakes  $\leq 50$  g of refined cereals, white rice, milled whole grains, milled whole wheat and milled whole millets, and participants with intakes  $\leq 5$  g  $\text{d}^{-1}$  of added sugar. Models were adjusted for PSU, state code, population weight, age (in years), sex (male/female), family history of T2D (yes/no), tobacco use (never/current/past), alcohol use (never/current/past), PAL (sedentary/moderate/vigorous), time period the study was conducted, education qualification (primary school/high school/college), BMI ( $\text{kg m}^{-2}$ ; except for generalized obesity), place of residence: urban versus rural, intakes of total energy ( $\text{kcal d}^{-1}$ ), fat (%E), fruits ( $\text{g d}^{-1}$ ), vegetables ( $\text{g d}^{-1}$ ), tubers ( $\text{g d}^{-1}$ ), milk and milk products ( $\text{g d}^{-1}$ ), alcohol ( $\text{g d}^{-1}$ ), pulses and legumes ( $\text{g d}^{-1}$ ), added salt ( $\text{g d}^{-1}$ ) and edible oils and fats ( $\text{g d}^{-1}$ ). Models were further adjusted for newly diagnosed T2D (except in models with newly diagnosed T2D and newly diagnosed prediabetes as outcomes), newly diagnosed prediabetes (except in models with newly diagnosed T2D and newly diagnosed prediabetes as outcomes), hypertension, dyslipidaemia and abdominal obesity (except in models with abdominal and generalized obesity as outcomes). Models were additionally adjusted for all food sources and nutrients (carbohydrates, %E) except those that were the outcome. All analyses were conducted separately in each of the six regions. Heterogeneity in the association of the dietary exposure with metabolic risk was assessed by using the  $I^2$  statistic.  $I^2$  values of <25%, 50% and 75% were considered to indicate low, moderate and high heterogeneity, respectively.

We statistically modeled the isocaloric substitution of 5% of energy from carbohydrates with an equal amount from protein (from actual foods and not supplements like protein powders) and fat by simultaneously including %E derived from carbohydrates with %E from the replacement nutrient as continuous variables along with other covariates. The difference in the beta coefficients of the two macronutrients being compared, along with the variances and covariances, were used to estimate the OR and 95% confidence interval for the substitution effect<sup>62</sup>. Additionally, we statistically modeled the replacement of 50 g of refined cereals with 50 g of milled whole grains, whole wheat flour, or whole millet flour. The base model was adjusted for PSU, state code, population weight, age (in years), sex (male/female), family history of T2D (yes/no), tobacco use (never/current/past), alcohol use (never/current/past), PAL (sedentary/moderate/vigorous), time period the study was conducted, education qualification (primary school/high school/college), BMI ( $\text{kg m}^{-2}$ ; except for generalized obesity), place of residence (urban versus rural) and intakes of total energy ( $\text{kcal d}^{-1}$ ). It was further adjusted for newly diagnosed T2D (except in models with newly diagnosed T2D and newly diagnosed prediabetes as outcomes), newly diagnosed prediabetes (except in models with newly diagnosed T2D and newly diagnosed prediabetes as outcomes), hypertension, dyslipidaemia and abdominal obesity (except in models with abdominal obesity as an outcome). Models were additionally adjusted for all nutrients except the nutrient being replaced. In sensitivity analyses, we tested for potential effect modification of the association between nutrients and metabolic risk by place of residence (urban versus rural) and sex (male versus female). Since these tests were not significant ( $P_{\text{interaction}} > 0.05$ ), we present results without stratification. All statistical analyses were two sided and performed with R software (version 4.3.3) and SAS (version 9.4; SAS Institute).

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Data access (de-identified data) for the ICMR–INDIAB study cannot be made publicly available because of ethical and data protection constraints in India. Being a Government of India-funded study with

involvement of all states, multiple layers of permission are required for data access. However, if corresponding authors are contacted via email, they will get the required permissions and share the raw data without cost for noncommercial and research purposes within 3–4 months. However, the data use will be subject to the norms of the Research Ethics Board of the Madras Diabetes Research Foundation, the ICMR and Department of Health Research, Government of India. Data will be shared through an institutional data sharing agreement or arrangements will be made for analyses to be conducted remotely without the necessity for data transfer.

### Code availability

The underlying code for the results detailed in this study is available at GitHub—[https://github.com/MDRFFNDRstats/INDIAB\\_21\\_Substitution\\_model.git](https://github.com/MDRFFNDRstats/INDIAB_21_Substitution_model.git).

### References

- World Health Organization. STEPwise approach to surveillance (STEPS). <https://www.who.int/europe/tools-and-toolkits/who-stepwise-approach-to-surveillance#:~:text=Home/,information%20on%20NCD%20risk%20factors> (2020). Accessed on 10-09-2025
- Friedewald, W. T. et al. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin. Chem.* **18**, 499–502 (1972).
- Anjana, R. M. et al. Reliability and validity of a new physical activity questionnaire for India. *Int. J. Behav. Nutr. Phys. Act.* **12**, 40 (2015).
- Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care* **20**, 1183–1197 (1997).
- World Health Organization. The Asia-Pacific perspective: redefining obesity and its treatment. [iris.who.int/handle/10665/206936](https://iris.who.int/handle/10665/206936) (2000).
- James, P. A. et al. 2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). *JAMA* **311**, 507–520 (2014).
- Willett, W. *Nutritional Epidemiology* (eds Hofman, A. et al.) 260–286 (Oxford University Press, 2013).
- Li, J. et al. Comparison of the effects of imputation methods for missing data in predictive modelling of cohort study datasets. *BMC Med. Res. Methodol.* **24**, 41 (2024).
- Song, M. & Giovannucci, E. Substitution analysis in nutritional epidemiology: 919 proceed with caution. *Eur. J. Epidemiol.* **33**, 137–140 (2018).

### Acknowledgements

The authors would like to thank the ICMR, New Delhi, and the Department of Health Research, Ministry of Health and Family Welfare, Government of India, New Delhi, for providing financial support to the study and the ICMR–INDIAB Expert Group for their valuable suggestions and scientific input. The authors would also like to thank the ICMR–INDIAB quality managers, quality supervisors and the field team for smooth conduct of the study, and the participants for their cooperation. This is the 21st paper from the ICMR–INDIAB study (ICMR–INDIAB-21). A full list of ICMR–INDIAB collaborative study group members and their affiliations appears in the Supplementary Information.

### Author contributions

R.M.A. and V.M. conceived and designed the study, oversaw its implementation, trained the study team, developed quality assurance measures, interpreted the data and were involved in both the original drafting and revision of the manuscript. M.D. and R.P. were involved in the methodology and project administration, and in reviewing and

editing the manuscript. V.S. and R.U. were involved in the interpretation of the data and drafting. S.N.B. conceived and performed formal analysis and wrote the original draft of the manuscript. A.V.K. and K.K. provided critical revision of the manuscript. A.L. provided valuable comments on the manuscript. S.J., B.S., A.G., P.P.J., P.A., P.K.J., S.M.J., S.C., A.J.P., S.K.T., S.B. and J.M. were responsible for the supervision of the study in their respective states and reviewing the report, provided scientific inputs for the study, were involved in the quality control and helped to revise the report. R.G. helped in the field coordination of the study and methodology. K.A. and S.N.B. were responsible for data curation and statistical analyses. V.S., R.G. and V.S.M. coordinated investigations and contributed to the writing of original drafts. A.K.D., T.K. and R.S.D are members of ICMR–INDIAB Expert Group and helped to review the report. All authors contributed to the revision of the report and approved the final submitted version. R.M.A., S.N.B., V.S. and V.M. take full responsibility for the overall content of this work.

### Competing interests

The authors declare no competing interests.

### Additional information

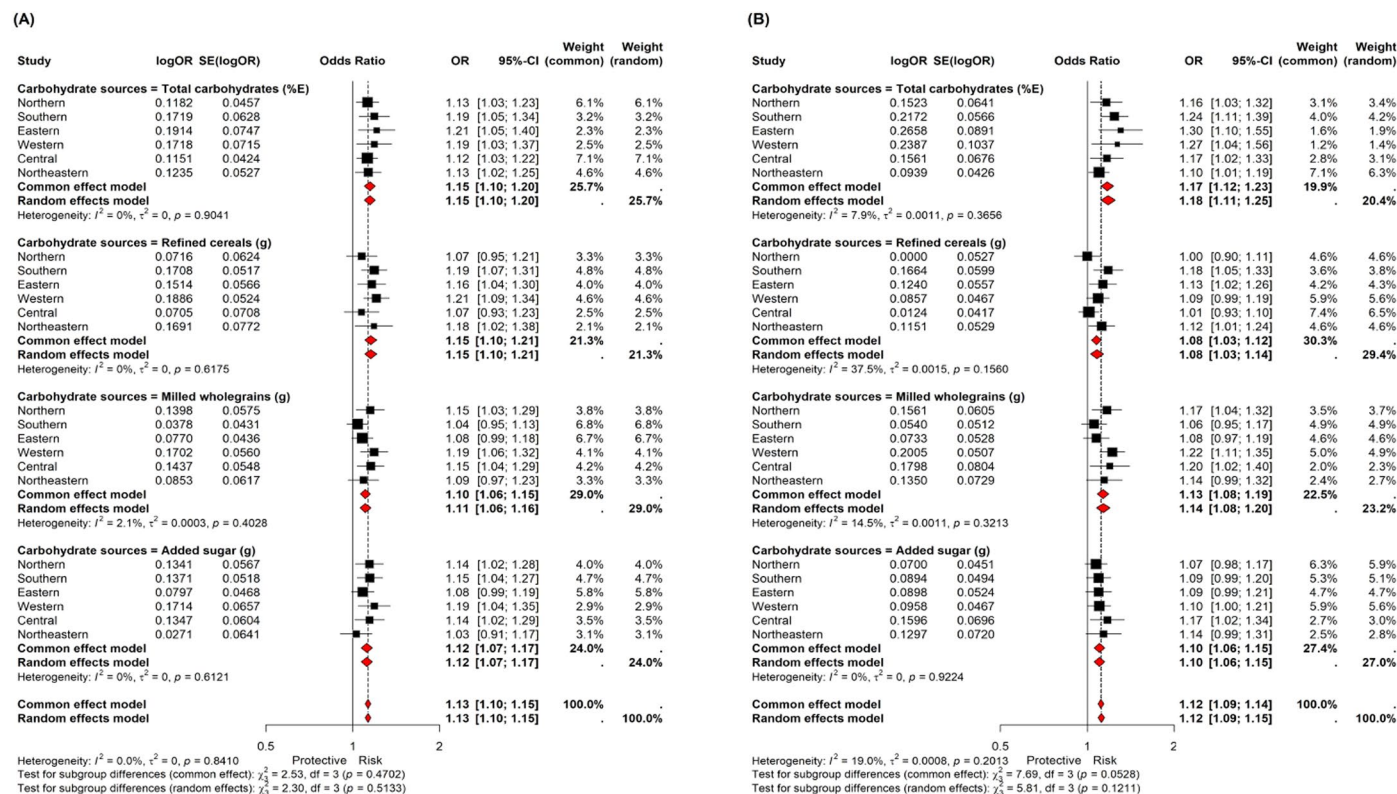
**Extended data** is available for this paper at <https://doi.org/10.1038/s41591-025-03949-4>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41591-025-03949-4>.

**Correspondence and requests for materials** should be addressed to Ranjit Mohan Anjana.

**Peer review information** *Nature Medicine* thanks Laura Lara-Castor, Monika Chaudhary, Jagmeet Madan and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Ming Yang, in collaboration with the *Nature Medicine* team.

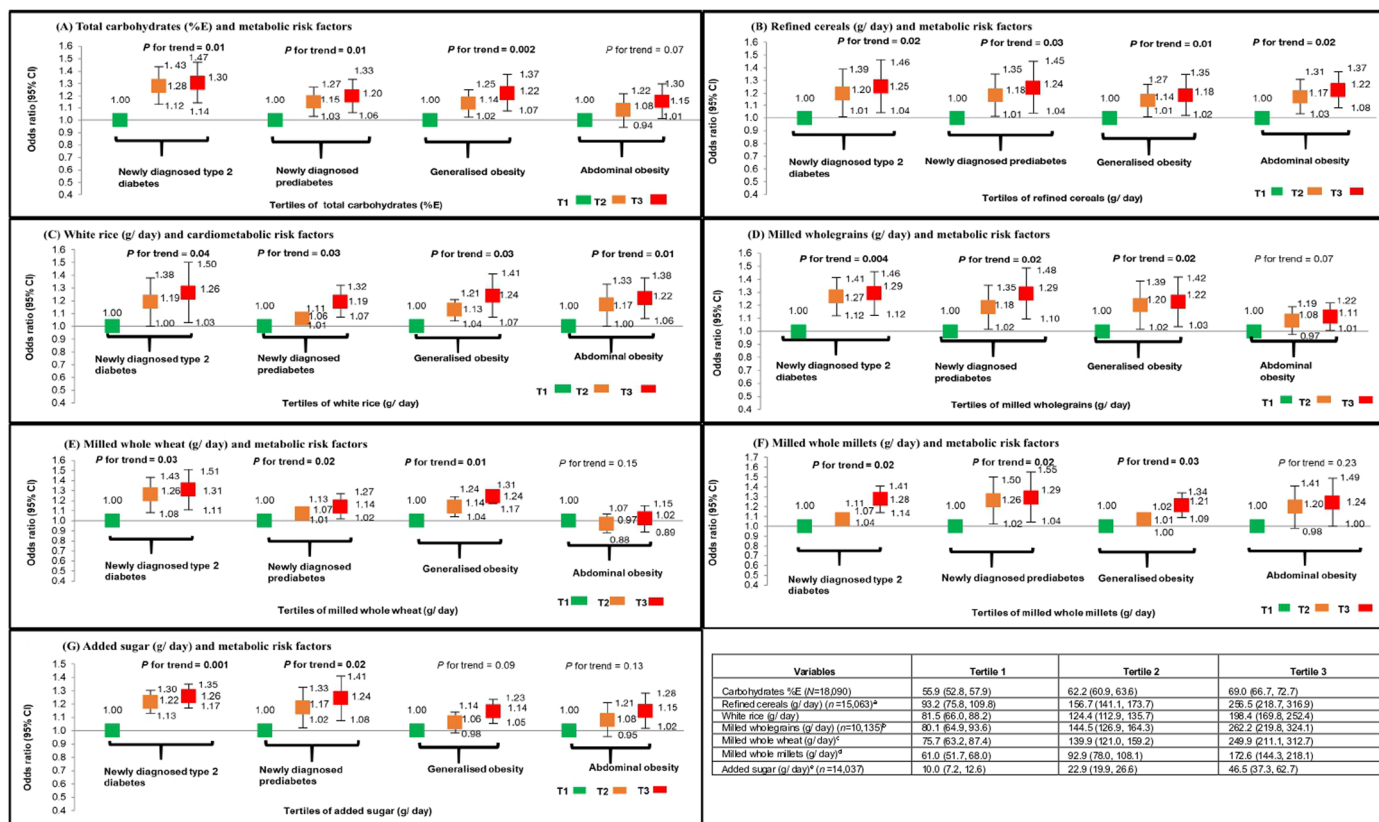
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



**Extended Data Fig. 1 | Pooled odds ratio (95% confidence interval) for the association of total carbohydrates (%E) and its selected food sources (g/day) with likelihood of generalized obesity and abdominal obesity in the ICMR-INDIAB study ( $n = 18,090$ ).** **a**, Forest plot for generalized obesity, which is defined as body mass index (BMI)  $\geq 25$  kg/m<sup>2</sup>. **b**, Forest plot for abdominal obesity, defined as waist circumference of  $\geq 90$  cm for males and  $\geq 80$  cm for females. A fixed-effects model was used to synthesize data from the regions. Two-sided p-values were used in calculating odds ratios and 95% confidence intervals via survey logistic regression. Energy adjusted was used for all the dietary exposure variables (refined cereals (g/day), milled wholegrains (g/day), added sugar (g/day), white rice (g/day), milled whole wheat (g/day) and milled whole millets (g/day)). Models were adjusted for primary sampling unit (PSU), state code, population weight, age (in years), sex (male/female), family history

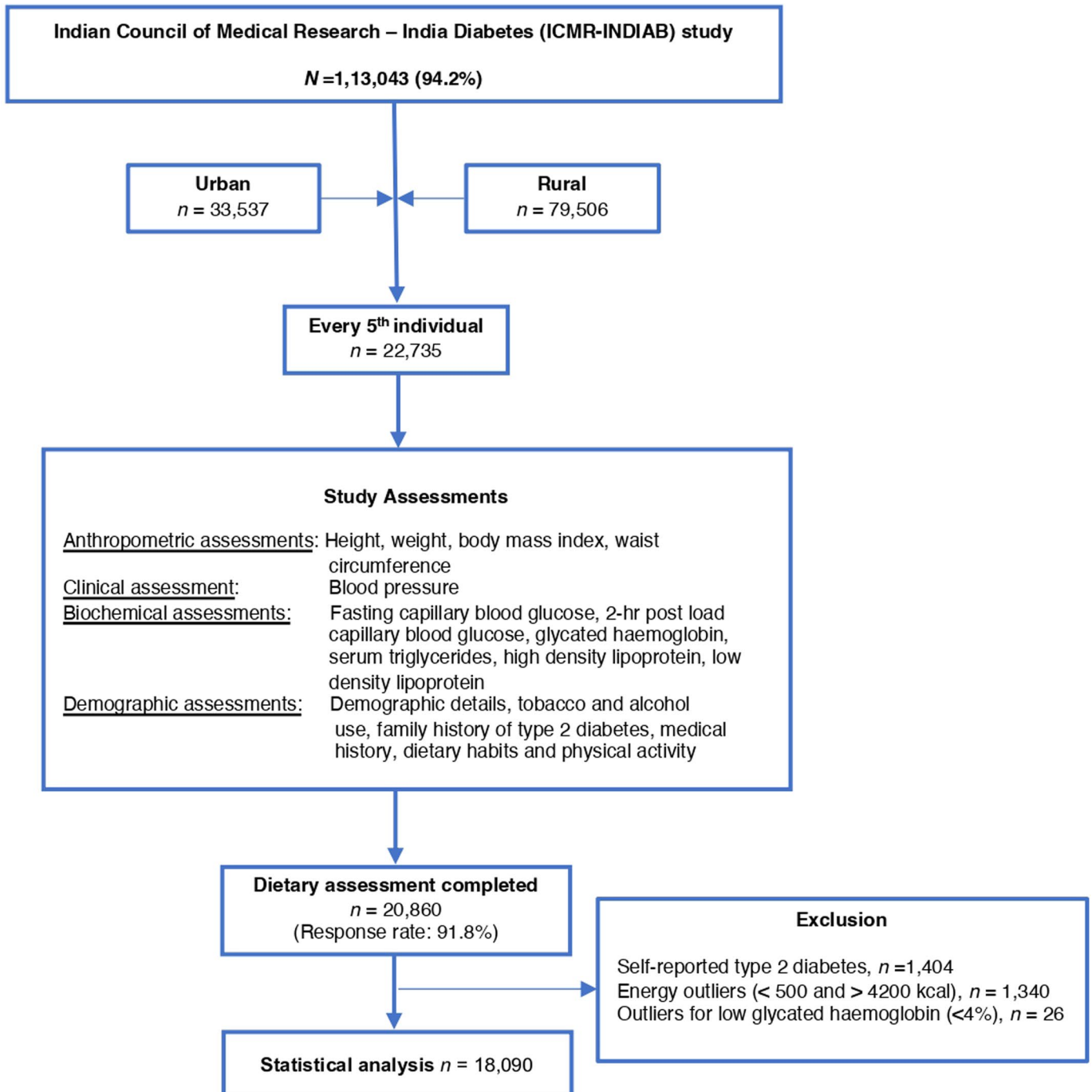
of type 2 diabetes (yes/no), tobacco use (never/current/past), alcohol use (never/current/past), physical activity level (sedentary/moderate/vigorous), time period the study was conducted, education qualification (primary school/high school/college), body mass index (BMI) (kg/m<sup>2</sup>) (except for generalized obesity), place of residence: urban vs rural, intakes of total energy (kcal/day). It was further adjusted for newly diagnosed type 2 diabetes, newly diagnosed prediabetes, hypertension, dyslipidaemia, and abdominal obesity (except in models with abdominal and generalized obesity are an outcome). Models with food sources as the main exposure were adjusted for the remaining food sources while models with total carbohydrates (%E) were adjusted for total fat (%E). For refined cereals and milled wholegrains, the analysis only included participants who consumed  $\geq 50$  g/day. For added sugar, the analysis only included participants who consumed  $\geq 1$  teaspoon ( $\geq 5$  g/day).





**Extended Data Fig. 2 | Odds ratios (95% confidence intervals) for the association between total carbohydrates and its selected food sources and cardiometabolic risk factors in the study population ( $n = 18,090$ ).** Data presented as weighted median (IQR). Energy adjusted by residual method was used for all the dietary exposure variables (refined cereals (g/day), milled wholegrains (g/day), added sugar (g/day), white rice (g/day), milled whole wheat (g/day) and milled whole millets (g/day)). Models were adjusted for primary sampling unit (PSU), state code, population weight, age (in years), sex (male/female), family history of type 2 diabetes (yes/no), tobacco use (never/current/past), alcohol use (never/current/past), physical activity level (sedentary/moderate/vigorous), time period the study was conducted, education qualification (primary school/high school/college), body mass index (BMI) ( $\text{kg}/\text{m}^2$ ) (except for generalized obesity), place of residence: urban vs rural, intakes of total energy (kcal/day), fat (%E), fruits (g/day), vegetables (g/day), tubers (g/day), milk and milk products (g/day), alcohol (g/day), pulses and legumes (g/day), added salt (g/day), and edible oils and fats (g/day). It was further adjusted for newly diagnosed type 2 diabetes (except in models with newly diagnosed type 2 diabetes and newly diagnosed prediabetes as outcomes), newly diagnosed prediabetes (except in models with newly diagnosed type 2 diabetes and newly diagnosed prediabetes as outcomes), hypertension, dyslipidaemia, and abdominal obesity (except in models with abdominal and generalized obesity as outcomes). Models with food sources as the main exposure were adjusted for the remaining food sources while models with total carbohydrates (%E) were adjusted for total fat (%E). Newly diagnosed type 2 diabetes was defined as glycated hemoglobin (HbA1c)  $\geq 6.5\%$  and/or if fasting capillary blood

glucose  $\geq 126$  mg/dL ( $\geq 7$  mmol/L) or 2-h post glucose load capillary blood glucose  $\geq 220$  mg/dL ( $\geq 12.2$  mmol/L). Newly diagnosed prediabetes was defined as glycated hemoglobin (HbA1c)  $5.7\% - 6.4\%$  or if fasting capillary blood glucose was  $\geq 100 - 125$  mg/dL ( $\geq 5.6 - 6.9$  mmol/L) and/or if 2-h post load capillary blood glucose value was  $\geq 160$  and  $< 220$  mg/dL ( $\geq 8.9$  to  $< 12.2$  mmol/L). Generalized obesity was defined as body mass index (BMI)  $\geq 25$   $\text{kg}/\text{m}^2$ . Abdominal obesity was defined as waist circumference of  $\geq 90$  cm for males and  $\geq 80$  cm for females. Superscript letter 'a' indicates refined cereals include white rice, refined wheat flour (maida), refined millet grains/flour (minor millets like little millet and foxtail millet) and wheat semolina (rava). Superscript letter 'b' indicates milled whole grains include milled whole wheat and milled whole millet as flours. White rice includes foods made from white rice flour and white rice flakes. Superscript letter 'c' indicates milled whole wheat includes foods made from whole wheat grits and flour. Superscript letter 'd' indicates milled whole millets include foods made from millet grits and flours. Superscript letter 'e' indicates added sugar from white sugar, honey, jaggery and palm sugar added during cooking and on the table. For refined cereals and milled whole grains, the analysis only included participants who consumed  $\geq 50$  g/day. For added sugar, the analysis only included participants who consumed  $\geq 1$  teaspoon ( $\geq 5$  g/day). **a**, Total carbohydrates (%E) and metabolic risk factors. **b**, Refined cereals (g/day) and metabolic risk factors. **c**, White rice (g/day) and cardiometabolic risk factors. **d**, Milled wholegrains (g/day) and metabolic risk factors. **e**, Milled whole wheat (g/day) and metabolic risk factors. **f**, Milled whole millets (g/day) and metabolic risk factors. **g**, Added sugar (g/day) and metabolic risk factors.



**Extended Data Fig. 3 | Study flow diagram.** Flow diagram outlining the number of participants assessed for eligibility, excluded, enrolled and analyzed in the final diet dataset.

**Extended Data Table 1 | Top three cereal staple choices reported by the weighted percentage of the study population stratified by states and union territories (n=18,090)**

Region	Region/ State	n	White rice n (%)	Whole wheat flour n (%)	Total millet flours n (%)	Finger millet flour n (%)	Pearl millet flour n (%)	Sorghum millet flour n (%)	Maize flour n (%)
All India		18,090	10,998 (60.8)	6,213 (34.3)	879 (4.9)	163 (0.9)	260 (1.4)	343 (1.9)	113 (0.6)
Region	Northern	3,714	298 (8)	3,349 (90)	67 (2)	4 (0.1)	34 (0.9)	-	29 (0.8)
	Southern	3,198	2,777 (87)	56 (2)	365 (11)	143 (4.4)	4 (0.1)	216 (6.8)	2 (0.1)
	Eastern	2,045	1,601 (78.2)	437 (21.3)	7 (0.5)	4 (0.2)	1 (0.1)	-	2 (0.1)
	Western	1,730	703 (41)	607 (35)	420 (24)	11 (0.6)	219 (12.7)	126 (7.2)	64 (3.7)
	Central	2,470	729 (29.5)	1,723 (69.8)	18 (0.7)	1 (0.02)	1 (0.02)	-	16 (0.7)
	Northeastern	4,933	4,890 (99)	41 (0.8)	2 (0.2)	-	1 (0.02)	1 (0.02)	-
Northern	Chandigarh	505	83 (12)	422 (88)	-	-	-	-	-
	NCT Delhi	616	48 (12)	568 (88)	-	-	-	-	-
	Haryana	675	9 (1)	666 (99)	-	-	-	-	-
	Himachal Pradesh	603	143 (25.0)	456 (74.5)	4 (0.5)	4 (0.5)	-	-	-
	Punjab	615	12 (2.0)	603 (98.0)	-	-	-	-	-
	Rajasthan	700	3 (0.4)	634 (90.4)	63 (9.2)	-	34 (5.3)	-	29 (3.9)
Southern	Andhra Pradesh <sup>a</sup>	630	613 (97.4)	7 (1.2)	10 (1.4)	10 (1.4)	-	-	-
	Karnataka	723	344 (47.6)	31 (4.7)	348 (47.7)	130 (17.4)	2 (0.3)	215 (29.8)	1 (0.2)
	Kerala	584	574 (97.7)	7 (1.7)	3 (0.6)	1 (0.2)	1 (0.2)	1 (0.2)	-
	Puducherry	625	618 (99.1)	7 (0.9)	-	-	-	-	-
	Tamil Nadu	636	628 (98.7)	4 (0.8)	4 (0.5)	2 (0.3)	1 (0.2)	-	1 (0.1)
	Bihar	559	290 (53.0)	264 (46.2)	5 (0.8)	2 (0.2)	1 (0.2)	-	2 (0.4)
Eastern	Jharkhand	520	404 (80.8)	114 (18.9)	2 (0.3)	2 (0.3)	-	-	-
	Odisha	505	480 (93.7)	25 (6.3)	-	-	-	-	-
	West Bengal	461	427 (92.0)	34 (8.0)	-	-	-	-	-
	Goa	628	573 (91.6)	50 (7.7)	5 (0.7)	-	-	5 (0.7)	-
Western	Gujarat	634	42 (5.8)	348 (58.9)	244 (35.3)	9 (1.2)	161 (23.4)	12 (1.5)	62 (9.2)
	Maharashtra	468	88 (17.4)	209 (47.5)	171 (34.6)	2 (0.5)	58 (11.6)	109 (22.5)	2 (0.3)
	Chhattisgarh	552	525 (95.8)	27 (4.2)	-	-	-	-	-
Central	Madhya Pradesh	673	30 (4.6)	627 (92.6)	16 (2.8)	-	-	-	16 (2.8)
	Uttar Pradesh	606	115 (19.3)	490 (80.6)	1 (0.2)	-	1 (0.2)	-	-
	Uttarakhand	639	59 (9.2)	579 (90.7)	1 (0.2)	1 (0.2)	-	-	-
	Arunachal Pradesh	662	654 (98.8)	8 (1.2)	-	-	-	-	-
Northeastern	Assam	682	670 (99.1)	12 (0.9)	-	-	-	-	-
	Manipur	688	687 (99.9)	-	1 (0.1)	-	1 (0.1)	-	-
	Meghalaya	623	621 (99.7)	2 (0.3)	-	-	-	-	-
	Mizoram	594	593 (99.8)	-	1 (0.2)	-	-	1 (0.2)	-
	Nagaland	550	543 (99.0)	7 (1)	-	-	-	-	-
	Sikkim	461	451 (98.8)	10 (1.2)	-	-	-	-	-
	Tripura	673	671 (99.7)	2 (0.3)	-	-	-	-	-

Data are presented as weighted percentages. <sup>a</sup>Undivided.



**Extended Data Table 2 | Inter-regional differences in intakes of macronutrients and their selected food sources and subtypes for the study population (n=18,090)**

Macronutrient	Region						
	All India	Northern	Southern	Eastern	Western	Central	Northeastern
<b>Carbohydrates (%E) and food sources</b>							
Carbohydrates (%E)	62.3 (60.0, 66.8)	60.3 (56.8, 63.3)	63.7 (59.6, 67.4)	64.8 (60.8, 69.0)	59.6 (55.6, 59.2)	62.6 (59.2, 66.0)	63.8 (57.1, 70.5)
Refined cereals (%E) <sup>a</sup>	28.5 (16.9, 48.6)	9.4 (3.4, 18.1)	36.0 (27.4, 45.6)	31.5 (19.1, 45.8)	25.7 (16.5, 50.5)	18.8 (11.4, 26.3)	51.7 (31.0, 69.0)
White rice (%E)	26.2 (15.5, 37.5)	8.1 (3.6, 13.9)	33.8 (22.5, 45.1)	31.1 (17.8, 41.9)	23.3 (11.5, 39.5)	17.1 (10.1, 23.9)	48.1 (28.1, 67.5)
Milled wholegrains (%E) <sup>b</sup>	16.2 (8.9, 25.1)	27.8 (19.5, 32.1)	11.4 (6.0, 21.9)	14.6 (6.5, 24.6)	15.0 (6.9, 25.9)	30.9 (22.4, 38.1)	7.7 (1.6, 15.6)
Milled whole wheat (%E)	15.0 (7.1, 23.5)	26.3 (13.5, 35.0)	10.9 (6.7, 15.1)	13.4 (6.5, 24.1)	10.0 (6.7, 23.5)	30.0 (21.5, 40.1)	7.6 (1.7, 13.6)
Milled whole millets (%E)	1.4 (0.3, 4.9)	0.2 (0.1, 0.7)	0.4 (0.1, 2.7)	0.3 (0.1, 0.9)	5.1 (1.1, 12.1)	0.3 (0.1, 1.2)	0.0 (0.0, 0.3)
Added sugar (%E) <sup>c</sup>	4.2 (1.5, 8.5)	7.9 (4.1, 12.9)	7.1 (3.9, 11.7)	6.3 (3.2, 12.7)	7.1 (3.4, 11.4)	7.0 (3.1, 10.9)	2.1 (0.4, 4.1)
<b>Fats and subtypes</b>							
Total fats (%E)	25.2 (20.8, 29.1)	27.7 (24.7, 30.7)	24.7 (21.1, 28.2)	24.0 (19.6, 28.0)	27.9 (24.2, 31.7)	25.6 (22.2, 29.0)	21.6 (16.3, 27.1)
Saturated fat (%E)	8.9 (6.6, 11.6)	11.2 (8.8, 14.1)	9.3 (7.1, 12.3)	7.6 (5.5, 9.8)	9.9 (8.0, 12.2)	8.5 (6.4, 10.9)	7.4 (4.8, 9.8)
Monounsaturated fat (%E)	6.1 (4.8, 7.5)	6.7 (5.6, 7.7)	5.7 (4.8, 6.8)	5.0 (4.0, 6.4)	6.0 (4.9, 7.5)	6.7 (5.4, 7.8)	6.1 (4.3, 7.9)
Total Polyunsaturated fat (%E)	7.7 (5.8, 9.3)	8.4 (7.2, 9.5)	8.0 (6.5, 9.4)	8.1 (6.7, 9.6)	8.0 (6.3, 9.5)	9.1 (7.6, 10.3)	5.3 (3.9, 7.1)
n6 Polyunsaturated fat (%E)	7.1 (5.2, 8.7)	7.8 (6.5, 9.0)	7.5 (6.1, 8.9)	7.6 (6.3, 9.2)	7.3 (5.8, 8.8)	8.4 (7.1, 9.6)	4.6 (3.2, 6.4)
n3 Polyunsaturated fat (%E)	0.2 (0.2, 0.3)	0.2 (0.2, 0.3)	0.2 (0.1, 0.2)	0.2 (0.2, 0.3)	0.2 (0.1, 0.2)	0.2 (0.2, 0.3)	0.4 (0.2, 0.7)
<b>Protein and subtypes</b>							
Total protein (%E)	12.0 (11.0, 13.3)	12.1 (11.2, 13.1)	11.5 (10.7, 12.3)	11.3 (10.4, 12.1)	12.2 (11.4, 13.1)	11.5 (10.6, 12.4)	13.6 (12.0, 16.1)
Plant protein <sup>d</sup> (%E)	8.9 (7.6, 10.0)	9.2 (8.2, 10.1)	7.5 (6.2, 8.6)	9.5 (8.7, 10.3)	8.6 (6.5, 9.9)	9.6 (8.7, 10.4)	8.7 (7.5, 9.9)
Dairy protein <sup>e</sup> (%E)	2.1 (1.1, 3.3)	3.6 (2.6, 4.8)	2.0 (1.3, 2.8)	1.2 (0.7, 2.0)	2.7 (1.8, 3.7)	2.5 (1.4, 3.5)	1.1 (0.5, 1.7)
Fermented dairy protein (%E)	0.2 (0.1, 0.5)	0.5 (0.3, 0.8)	0.1 (0.07, 0.3)	0.1 (0.0, 0.3)	0.4 (0.2, 0.7)	0.3 (0.2, 0.6)	0.02 (0.0, 0.1)
Non-fermented dairy protein (%E)	1.7 (0.8, 2.7)	3.0 (2.2, 4.1)	1.8 (1.1, 2.5)	1.0 (0.4, 1.6)	2.2 (1.4, 3.0)	2.1 (1.2, 2.9)	0.9 (0.3, 1.6)
Animal protein <sup>f</sup> (%E)	1.0 (0.1, 2.0)	0.3 (0.1, 0.4)	1.0 (0.4, 1.6)	0.6 (0.1, 1.0)	1.4 (0.5, 2.0)	0.3 (0.0, 0.5)	2.9 (1.5, 5.5)
Red meat protein (%E)	0.2 (0.1, 0.8)	0.1 (0.0, 0.2)	0.1 (0.0, 0.3)	0.1 (0.0, 0.1)	0.1 (0.0, 0.2)	0.1 (0.0, 0.3)	0.8 (0.3, 2.2)
Poultry protein (%E)	0.3 (0.1, 0.6)	0.2 (0.1, 0.4)	0.3 (0.1, 0.6)	0.2 (0.1, 0.4)	0.2 (0.1, 0.4)	0.1 (0.1, 0.3)	0.5 (0.2, 1.2)
Egg protein (%E)	0.1 (0.02, 0.2)	0.0 (0.0, 0.2)	0.1 (0.0, 0.2)	0.1 (0.0, 0.2)	0.1 (0.0, 0.2)	0.0 (0.0, 0.1)	0.1 (0.0, 0.2)
Fish protein (%E)	0.2 (0.0, 0.6)	0.0 (0.0, 0.1)	0.3 (0.1, 0.7)	0.1 (0.0, 0.4)	0.5 (0.2, 1.5)	0.1 (0.0, 0.2)	0.5 (0.1, 1.6)

Data are presented as weighted median (IQR). <sup>a</sup>Refined cereals include white rice, refined wheat flour (maida), refined millet grains/flour (minor millets like little millet and foxtail millet) and wheat semolina (rava). <sup>b</sup>Milled wholegrains included milled whole wheat and milled whole millet as flours. <sup>c</sup>Added sugar from white sugar, honey, jaggery and palm sugar added during cooking and on the table. For refined cereals and milled wholegrains the analysis only included participants who consumed  $\geq 50$  g/day. For added sugar, the analysis only included participants who consumed  $\geq 1$  teaspoon ( $\geq 5$  g/day). <sup>d</sup>Plant protein includes cereals, pulses, legumes, fruits, vegetables (including roots and tubers), nuts and oilseeds. <sup>e</sup>Dairy protein includes fermented dairy protein (Indian style yoghurt (curd), buttermilk, lassi) and non-fermented dairy protein (all types of milk (cow/buffalo/yak/goat), milk added to coffee and tea, paneer (curdled milk solids)). <sup>f</sup>Animal protein includes red meat, poultry, egg, and fish and does not include dairy protein.

**Extended Data Table 3 | Intra-regional differences in intakes of total carbohydrates and their selected food sources (%E) for the study population (n=18,090)**

Region	State	Carbohydrates	Refined cereals <sup>a</sup>	White rice	Milled wholegrains <sup>b</sup>	Milled whole wheat	Milled whole millets	Added sugar <sup>c</sup>
Northern	Chandigarh	60.2 (56.5, 63.4)	15.6 (6.9, 23.7)	12.4 (6.7, 18.4)	34.8 (31.0, 39.6)	33.7 (29.0, 39.6)	0.0 (0.0, 0.0)	4.6 (3.0, 6.4)
	NCT Delhi	59.3 (57.0, 62.6)	9.9 (2.2, 17.3)	7.4 (3.0, 12.5)	19.3 (13.3, 25.1)	18.2 (12.3, 24.1)	0.0 (0.0, 0.1)	10.4 (5.6, 15.1)
	Haryana	60.3 (57.0, 62.8)	15.5 (3.2, 27.4)	11.8 (6.9, 15.6)	19.7 (14.0, 27.1)	19.1 (10.9, 29.3)	0.0 (0.0, 0.1)	11.0 (5.9, 15.5)
	Himachal Pradesh	62.5 (60.0, 65.1)	18.4 (11.9, 24.8)	14.6 (9.9, 18.7)	23.9 (18.4, 31.7)	23.6 (15.1, 32.4)	0.3 (0.1, 0.5)	9.6 (5.4, 13.8)
	Punjab	61.8 (58.6, 65.1)	10.8 (4.4, 16.7)	8.9 (3.8, 13.1)	38.5 (35.0, 45.7)	37.3 (31.0, 43.7)	0.0 (0.0, 1.2)	9.1 (5.0, 13.7)
Southern	Rajasthan	56.7 (53.2, 59.3)	14.8 (5.1, 24.5)	11.9 (5.3, 17.5)	33.8 (23.1, 44.3)	31.1 (21.6, 43.1)	2.7 (0.7, 4.4)	8.7 (4.0, 12.7)
	Andhra Pradesh <sup>d</sup>	61.8 (57.8, 65.4)	38.4 (31.0, 47.9)	35.8 (28.1, 42.7)	10.1 (7.0, 17.1)	9.6 (6.1, 16.7)	0.0 (0.0, 0.0)	7.4 (4.4, 12.3)
	Karnataka	65.5 (62.2, 68.8)	25.8 (21.9, 33.5)	23.8 (19.7, 28.1)	25.1 (18.7, 31.8)	22.0 (13.0, 31.0)	10.0 (1.6, 19.4)	10.4 (5.2, 15.4)
	Kerala	64.7 (61.2, 68.3)	42.5 (35.0, 47.2)	39.5 (35.0, 46.5)	11.8 (4.9, 18.5)	8.6 (5.4, 12.1)	4.1 (2.7, 6.3)	7.5 (4.2, 12.6)
	Puducherry	62.1 (58.1, 66.0)	36.0 (25.7, 49.3)	33.6 (22.8, 45.8)	11.3 (3.6, 18.4)	10.9 (4.0, 17.7)	0.0 (0.0, 0.0)	9.5 (4.2, 14.6)
Eastern	Tamil Nadu	63.4 (58.8, 67.5)	51.5 (43.8, 57.8)	47.3 (38.9, 56.5)	9.5 (4.5, 15.4)	9.2 (3.9, 15.1)	0.1 (0.0, 0.2)	4.8 (1.9, 7.6)
	Bihar	65.7 (61.7, 69.7)	31.1 (18.2, 44.2)	28.9 (17.5, 39.6)	19.9 (9.8, 30.0)	19.5 (9.5, 29.7)	1.3 (0.2, 2.6)	5.9 (3.1, 8.9)
	Jharkhand	67.6 (63.5, 71.5)	32.4 (21.1, 46.1)	29.8 (17.9, 43.8)	16.0 (8.8, 24.0)	15.7 (9.3, 24.0)	0.9 (0.1, 1.6)	8.0 (4.1, 12.9)
	Odisha	63.0 (59.7, 66.0)	33.5 (19.1, 46.1)	30.7 (19.0, 41.4)	9.7 (5.5, 14.4)	9.3 (4.9, 13.6)	0.0 (0.0, 0.0)	10.7 (5.6, 15.4)
	West Bengal	62.8 (58.8, 67.0)	37.6 (19.4, 53.1)	34.9 (17.2, 50.9)	14.7 (5.1, 21.5)	13.9 (5.0, 21.3)	0.0 (0.0, 0.0)	7.2 (3.1, 10.9)
Western	Goa	58.4 (53.9, 63.2)	32.4 (17.4, 50.1)	29.8 (13.5, 45.5)	9.3 (4.1, 15.8)	8.3 (4.5, 12.2)	1.0 (0.1, 1.9)	7.8 (3.7, 11.4)
	Gujarat	60.6 (57.2, 63.7)	25.4 (11.0, 36.0)	22.6 (10.7, 33.6)	27.1 (18.6, 39.4)	19.7 (10.1, 30.3)	8.9 (0.7, 17.1)	7.1 (3.1, 10.8)
	Maharashtra	59.6 (55.7, 63.7)	47.9 (29.2, 60.4)	42.9 (27.5, 58.2)	10.9 (2.9, 17.8)	7.1 (1.2, 15.0)	5.5 (2.6, 9.2)	7.4 (3.1, 11.6)
Central	Chhattisgarh	68.2 (62.4, 73.6)	18.7 (13.5, 23.8)	17.5 (12.4, 22.1)	31.7 (22.2, 39.1)	30.7 (21.8, 37.6)	0.0 (0.0, 1.0)	6.1 (2.3, 10.8)
	Madhya Pradesh	62.5 (58.9, 65.3)	18.5 (5.2, 31.4)	16.6 (7.9, 25.2)	32.3 (23.5, 42.5)	32.0 (23.2, 42.1)	1.2 (0.1, 2.4)	10.2 (4.7, 15.6)
	Uttar Pradesh	61.6 (58.0, 64.4)	19.0 (11.2, 28.0)	18.0 (11.7, 24.6)	32.5 (25.1, 45.3)	32.0 (24.2, 44.1)	0.0 (0.0, 0.3)	6.4 (3.4, 9.5)
	Uttarakhand	61.2 (58.5, 63.7)	19.0 (15.7, 24.9)	17.4 (9.9, 24.5)	30.0 (21.0, 37.9)	29.5 (20.7, 36.9)	0.0 (0.0, 1.4)	9.0 (5.2, 13.5)
Northeastern	Arunachal Pradesh	69.3 (62.4, 77.9)	51.3 (35.2, 65.2)	47.7 (32.7, 62.8)	8.5 (1.3, 15.6)	8.5 (1.6, 14.5)	0.0 (0.0, 0.2)	1.1 (0.4, 3.5)
	Assam	64.7 (59.7, 69.3)	56.5 (38.6, 68.3)	52.8 (28.3, 72.1)	7.8 (1.3, 12.2)	7.5 (1.6, 12.0)	0.0 (0.0, 0.0)	3.5 (1.9, 5.1)
	Manipur	66.6 (59.8, 72.3)	54.8 (32.6, 69.1)	51.1 (27.3, 73.9)	8.7 (2.4, 15.0)	8.7 (1.7, 14.1)	0.0 (0.0, 0.1)	3.1 (2.0, 4.5)
	Meghalaya	63.1 (55.8, 70.2)	49.8 (26.9, 68.7)	45.2 (25.8, 65.8)	8.4 (1.2, 15.8)	8.4 (1.3, 15.8)	0.0 (0.0, 0.0)	5.8 (2.8, 6.9)
	Mizoram	55.9 (48.2, 63.6)	49.9 (34.1, 62.3)	45.1 (29.0, 61.6)	7.7 (1.9, 13.6)	7.5 (1.5, 13.5)	0.0 (0.0, 0.0)	1.3 (0.5, 3.1)
	Nagaland	58.5 (54.0, 63.1)	52.6 (28.5, 74.4)	49.5 (35.3, 64.3)	8.5 (1.1, 16.5)	8.4 (1.1, 15.8)	0.0 (0.0, 0.1)	1.5 (0.2, 3.5)
	Sikkim	57.5 (54.3, 60.8)	50.2 (23.6, 74.3)	47.1 (22.7, 71.8)	10.9 (2.5, 17.4)	9.9 (2.6, 15.5)	1.1 (0.1, 2.4)	1.9 (0.6, 3.3)
	Tripura	67.0 (62.7, 70.9)	54.2 (29.7, 75.3)	51.0 (26.4, 72.1)	8.7 (2.3, 14.7)	8.4 (1.6, 14.2)	0.0 (0.0, 0.0)	1.8 (0.3, 3.7)

Data are presented as weighted median (IQR). <sup>a</sup>Refined cereals include white rice, refined wheat flour (maida), refined millet grains/flour (minor millets like little millet and foxtail millet) and wheat semolina (rava). <sup>b</sup>Milled wholegrains included milled whole wheat and milled whole millet as flours. <sup>c</sup>Added sugar from white sugar, honey, jaggery and palm sugar added during cooking and on the table. For refined cereals and milled wholegrains the analysis only included participants who consumed  $\geq 50$  g/day. For added sugar, the analysis only included participants who consumed  $\geq 1$  teaspoon ( $\geq 5$  g/day). <sup>d</sup>Undivided.

**Extended Data Table 4 | Intra-regional differences in intakes of total fats and their subtypes (%E) for the study population (n=18,090)**

Region	State	Total fats	Saturated fat	Monounsaturated fat	Total Polyunsaturated fat	n6 Polyunsaturated fat	n3 Polyunsaturated fat
Northern	Chandigarh	26.6 (23.5, 29.8)	11.5 (8.6, 15.6)	5.6 (4.8, 6.5)	7.3 (6.2, 8.9)	6.7 (5.6, 8.3)	0.2 (0.2, 0.2)
	NCT Delhi	28.4 (25.9, 30.5)	10.4 (8.0, 12.9)	6.6 (5.9, 7.5)	8.8 (7.7, 9.9)	8.3 (7.2, 9.4)	0.2 (0.1, 0.3)
	Haryana	28.4 (26.1, 31.3)	11.3 (9.1, 13.8)	7.4 (6.6, 8.4)	8.9 (7.9, 9.9)	8.0 (7.0, 9.1)	0.2 (0.2, 0.3)
	Himachal Pradesh	25.3 (23.0, 27.4)	9.6 (8.2, 11.0)	7.2 (6.5, 7.9)	8.4 (7.7, 9.2)	7.7 (6.9, 8.5)	0.3 (0.2, 0.3)
	Punjab	25.9 (22.0, 29.1)	11.5 (8.9, 15.1)	4.9 (4.2, 5.6)	6.7 (5.3, 8.3)	6.3 (4.8, 7.9)	0.2 (0.2, 0.2)
	Rajasthan	30.6 (27.9, 33.8)	13.9 (10.9, 18.5)	7.6 (6.7, 8.7)	9.0 (8.0, 9.9)	8.6 (7.6, 9.7)	0.3 (0.2, 0.3)
Southern	Andhra Pradesh <sup>a</sup>	26.6 (23.2, 30.3)	15.3 (9.5, 24.6)	5.9 (5.0, 6.8)	7.9 (6.4, 9.5)	7.4 (5.9, 9.1)	0.2 (0.1, 0.2)
	Karnataka	23.0 (19.8, 26.0)	9.5 (7.2, 13.2)	5.0 (4.3, 5.8)	6.9 (5.8, 8.3)	6.4 (5.3, 7.8)	0.2 (0.1, 0.2)
	Kerala	24.4 (21.2, 27.4)	9.0 (7.2, 11.0)	5.6 (4.7, 6.5)	7.7 (6.5, 9.1)	7.4 (6.1, 8.8)	0.1 (0.1, 0.2)
	Puducherry	25.4 (21.9, 28.4)	9.0 (7.3, 10.8)	7.0 (6.1, 8.0)	9.0 (7.7, 10.4)	8.5 (7.2, 9.9)	0.2 (0.2, 0.2)
	Tamil Nadu	24.4 (20.6, 28.5)	7.4 (5.9, 9.3)	5.5 (4.7, 6.5)	8.3 (6.8, 9.5)	7.7 (6.3, 9.0)	0.2 (0.1, 0.2)
Eastern	Bihar	22.7 (18.4, 26.3)	7.2 (5.4, 8.9)	4.3 (3.5, 5.0)	7.3 (6.0, 8.8)	6.9 (5.7, 8.4)	0.2 (0.1, 0.2)
	Jharkhand	20.6 (16.8, 24.7)	6.1 (4.4, 8.5)	4.3 (3.5, 5.1)	7.7 (6.3, 9.2)	7.4 (6.0, 8.8)	0.2 (0.1, 0.2)
	Odisha	25.7 (22.7, 28.5)	8.3 (6.7, 10.0)	6.3 (5.3, 7.2)	8.8 (7.4, 10.0)	8.1 (6.7, 9.4)	0.2 (0.2, 0.3)
	West Bengal	26.3 (22.0, 29.7)	8.5 (6.1, 10.9)	6.2 (5.0, 7.5)	8.9 (7.3, 10.2)	8.3 (6.8, 9.8)	0.2 (0.2, 0.3)
Western	Goa	28.6 (24.8, 32.7)	10.0 (7.8, 12.2)	7.8 (6.5, 9.0)	8.8 (7.6, 10.1)	7.8 (6.6, 9.2)	0.2 (0.1, 0.2)
	Gujarat	27.1 (23.9, 30.3)	10.6 (9.1, 13.0)	5.1 (4.4, 5.8)	6.3 (5.4, 7.7)	5.9 (5.0, 7.3)	0.2 (0.2, 0.2)
	Maharashtra	27.9 (24.0, 31.9)	8.7 (7.1, 10.9)	5.7 (4.8, 6.6)	8.9 (7.1, 10.7)	8.4 (6.7, 10.1)	0.2 (0.1, 0.2)
Central	Chhattisgarh	20.5 (15.6, 25.8)	6.0 (4.3, 8.4)	5.1 (3.8, 6.3)	7.2 (5.7, 8.9)	6.7 (5.3, 8.1)	0.2 (0.1, 0.2)
	Madhya Pradesh	25.3 (22.4, 28.7)	8.0 (6.1, 10.6)	6.3 (5.4, 7.5)	9.1 (7.7, 10.3)	8.8 (7.4, 9.9)	0.2 (0.2, 0.3)
	Uttar Pradesh	26.8 (23.9, 30.6)	9.0 (7.2, 11.4)	6.6 (5.6, 7.9)	9.0 (7.7, 10.7)	8.4 (7.2, 10.1)	0.2 (0.2, 0.3)
	Uttarakhand	27.1 (24.8, 29.5)	9.9 (8.4, 11.5)	7.7 (7.1, 8.2)	9.9 (9.0, 10.7)	8.9 (8.0, 9.7)	0.2 (0.2, 0.3)
	Arunachal Pradesh	15.6 (9.7, 23.1)	4.6 (2.6, 6.9)	4.2 (2.6, 6.7)	4.2 (2.7, 6.3)	3.8 (2.3, 5.6)	0.3 (0.2, 0.4)
Northeastern	Assam	22.9 (19.2, 26.8)	7.7 (6.4, 9.5)	5.3 (4.3, 6.4)	6.8 (5.4, 8.3)	6.3 (5.0, 7.8)	0.3 (0.2, 0.4)
	Manipur	17.6 (13.5, 23.4)	4.7 (3.0, 6.8)	4.9 (3.4, 6.7)	4.6 (3.6, 6.0)	4.0 (3.0, 5.3)	0.9 (0.5, 1.5)
	Meghalaya	21.9 (16.8, 27.4)	7.0 (4.5, 9.5)	6.7 (4.9, 9.0)	5.9 (4.6, 7.5)	5.3 (4.3, 6.9)	0.4 (0.2, 0.7)
	Mizoram	27.8 (21.4, 33.0)	10.3 (7.7, 13.4)	7.5 (6.0, 9.2)	5.0 (4.0, 6.1)	4.0 (3.1, 5.0)	0.5 (0.3, 0.8)
	Nagaland	23.5 (19.5, 26.2)	9.3 (7.0, 11.3)	7.5 (6.1, 8.9)	4.5 (3.5, 5.5)	3.4 (2.6, 4.5)	0.9 (0.5, 1.9)
	Sikkim	28.9 (25.6, 31.8)	10.1 (8.5, 12.2)	8.0 (7.0, 9.1)	9.0 (7.9, 10.2)	7.8 (6.7, 8.9)	0.3 (0.3, 0.4)
	Tripura	19.3 (15.7, 22.7)	7.5 (5.6, 9.5)	5.7 (4.6, 6.9)	4.7 (3.5, 6.1)	4.0 (2.9, 5.2)	0.3 (0.2, 0.4)

Data are presented as weighted median (IQR). <sup>a</sup>Undivided.



**Extended Data Table 5 | Intra-regional differences in intakes of total protein and its subtypes (%E) for the study population (n=18,090)**

Region	State	Total protein	Plant protein <sup>a</sup>	Dairy protein <sup>b</sup>	Fermented dairy protein	Non-fermented dairy protein	24	Red meat protein	Poultry protein	Egg protein	Fish protein
Northern	Chandigarh	12.8 (12.0, 13.7)	9.6 (8.5, 10.5)	3.4 (2.5, 4.4)	0.3 (0.2, 0.5)	2.9 (2.2, 3.8)	0.6 (0.3, 1.1)	0.1 (0.0, 0.1)	0.3 (0.1, 0.5)	0.3 (0.1, 0.5)	0.1 (0.0, 0.2)
	NCT Delhi	11.5 (10.7, 12.1)	8.6 (8.0, 9.5)	4.1 (3.2, 4.7)	0.6 (0.5, 0.8)	3.4 (2.6, 4.1)	0.3 (0.1, 0.8)	0.1 (0.0, 0.2)	0.2 (0.1, 0.4)	0.0 (0.0, 0.0)	0.1 (0.0, 0.2)
	Haryana	11.2 (10.2, 12.1)	8.2 (7.0, 9.3)	3.7 (2.7, 4.7)	0.7 (0.5, 1.0)	2.9 (2.1, 3.8)	0.2 (0.1, 0.4)	0.1 (0.0, 0.2)	0.2 (0.1, 0.4)	0.0 (0.0, 0.1)	0.1 (0.0, 0.2)
	Himachal Pradesh	12.0 (11.0, 12.8)	10.2 (9.5, 11.0)	2.5 (1.8, 3.2)	0.3 (0.2, 0.5)	2.1 (1.4, 2.7)	0.2 (0.1, 0.3)	0.1 (0.0, 0.1)	0.1 (0.1, 0.2)	0.0 (0.0, 0.1)	0.02 (0.0, 0.04)
	Punjab	12.4 (11.6, 13.2)	9.4 (8.7, 10.1)	3.0 (2.4, 4.0)	0.3 (0.2, 0.5)	2.7 (2.1, 3.5)	0.3 (0.1, 0.8)	0.1 (0.0, 0.1)	0.3 (0.1, 0.5)	0.1 (0.0, 0.4)	0.02 (0.0, 0.04)
Southern	Rajasthan	12.9 (12.1, 13.6)	8.9 (7.7, 10.0)	2.7 (4.3, 6.6)	0.7 (0.6, 1.0)	4.7 (3.6, 5.7)	0.2 (0.1, 0.3)	0.1 (0.0, 0.2)	0.1 (0.1, 0.2)	0.0 (0.0, 0.1)	0.03 (0.0, 0.04)
	Andhra Pradesh <sup>d</sup>	11.5 (10.6, 12.5)	7.2 (6.4, 8.0)	2.4 (1.5, 3.7)	0.1 (0.1, 0.2)	2.2 (1.4, 3.4)	0.7 (0.3, 1.1)	0.1 (0.0, 0.2)	0.4 (0.2, 0.7)	0.2 (0.1, 0.4)	0.0 (0.0, 0.1)
	Karnataka	11.5 (11.0, 12.1)	8.1 (7.1, 8.9)	2.1 (1.6, 2.8)	0.2 (0.1, 0.4)	1.9 (1.5, 2.4)	0.8 (0.3, 1.3)	0.1 (0.0, 0.2)	0.4 (0.2, 0.7)	0.1 (0.0, 0.3)	0.2 (0.0, 0.3)
	Kerala	10.5 (9.8, 11.4)	5.7 (4.4, 7.0)	1.7 (0.9, 2.5)	0.2 (0.1, 0.3)	1.5 (0.7, 2.3)	1.3 (0.6, 1.9)	0.2 (0.1, 0.4)	0.1 (0.1, 0.3)	0.1 (0.0, 0.1)	0.7 (0.3, 1.4)
	Puducherry	11.8 (11.1, 12.7)	8.0 (6.5, 8.8)	1.5 (0.9, 2.4)	0.1 (0.1, 0.3)	1.3 (0.7, 2.0)	1.2 (0.7, 2.2)	0.2 (0.1, 0.4)	0.3 (0.1, 0.5)	0.0 (0.0, 0.1)	0.5 (0.2, 1.4)
Eastern	Tamil Nadu	11.9 (11.3, 12.8)	8.1 (6.9, 9.1)	2.0 (1.4, 2.7)	0.1 (0.0, 0.2)	1.8 (1.2, 2.5)	0.9 (0.4, 1.7)	0.1 (0.0, 0.4)	0.3 (0.1, 0.6)	0.3 (0.1, 0.6)	0.3 (0.0, 0.6)
	Bihar	11.5 (11.0, 12.2)	9.9 (9.2, 10.5)	1.2 (0.7, 1.9)	0.0 (0.0, 0.1)	1.1 (0.6, 1.8)	0.4 (0.1, 0.7)	0.0 (0.0, 0.3)	0.1 (0.1, 0.3)	0.1 (0.1, 0.2)	0.0 (0.0, 0.1)
	Jharkhand	11.5 (10.6, 12.4)	10.2 (9.5, 11.0)	0.9 (0.2, 1.6)	0.0 (0.0, 0.1)	0.5 (0.1, 1.3)	0.4 (0.1, 0.7)	0.0 (0.0, 0.2)	0.1 (0.1, 0.3)	0.2 (0.1, 0.5)	0.1 (0.0, 0.2)
	Odisha	11.1 (10.1, 12.1)	9.5 (8.9, 10.3)	1.3 (0.8, 1.8)	0.3 (0.1, 0.4)	0.9 (0.6, 1.5)	0.8 (0.3, 1.3)	0.1 (0.0, 0.2)	0.2 (0.1, 0.4)	0.1 (0.0, 0.1)	0.4 (0.1, 0.7)
	West Bengal	10.6 (9.9, 11.5)	8.6 (8.0, 9.4)	1.4 (0.8, 2.2)	0.3 (0.1, 0.4)	1.0 (0.5, 1.7)	0.9 (0.4, 1.4)	0.1 (0.0, 0.2)	0.2 (0.1, 0.5)	0.1 (0.0, 0.2)	0.3 (0.1, 0.6)
Western	Goa	12.1 (11.0, 13.2)	5.9 (4.5, 7.2)	2.1 (1.4, 2.9)	0.3 (0.2, 0.5)	1.7 (1.0, 2.5)	2.4 (1.6, 3.5)	0.0 (0.0, 0.2)	0.1 (0.1, 0.4)	0.1 (0.0, 0.2)	1.2 (1.1, 3.4)
	Gujarat	12.2 (11.6, 12.9)	9.4 (8.7, 10.1)	3.6 (2.6, 4.5)	0.7 (0.3, 1.0)	2.8 (2.1, 3.6)	0.4 (0.1, 0.7)	0.1 (0.0, 0.2)	0.3 (0.1, 0.5)	0.1 (0.0, 0.2)	0.0 (0.0, 0.0)
	Maharashtra	12.3 (11.6, 13.2)	9.9 (8.8, 10.9)	2.5 (1.8, 3.4)	0.4 (0.2, 0.6)	2.1 (1.4, 2.8)	0.5 (0.1, 0.7)	0.1 (0.0, 0.2)	0.3 (0.1, 0.5)	0.2 (0.1, 0.3)	0.1 (0.0, 0.2)
	Chhattisgarh	10.5 (9.7, 11.3)	9.7 (8.9, 10.5)	0.8 (0.4, 1.4)	0.2 (0.1, 0.3)	0.6 (0.3, 1.1)	0.2 (0.1, 0.5)	0.0 (0.0, 0.2)	0.1 (0.1, 0.2)	0.0 (0.0, 0.0)	0.1 (0.0, 0.2)
	Madhya Pradesh	12.3 (11.4, 13.0)	9.8 (8.3, 10.7)	2.2 (2.5, 4.4)	0.6 (0.4, 0.7)	2.6 (2.0, 3.6)	0.2 (0.1, 0.6)	0.1 (0.0, 0.2)	0.1 (0.1, 0.3)	0.0 (0.0, 0.1)	0.1 (0.0, 0.2)
Central	Uttar Pradesh	11.4 (10.7, 12.1)	9.2 (8.0, 10.2)	2.2 (1.6, 3.0)	0.3 (0.1, 0.4)	1.9 (1.4, 2.6)	0.3 (0.1, 0.5)	0.1 (0.0, 0.2)	0.1 (0.1, 0.3)	0.1 (0.0, 0.1)	0.0 (0.0, 0.0)
	Uttarakhand	11.6 (10.8, 12.4)	9.6 (9.0, 10.2)	3.1 (2.4, 3.9)	0.4 (0.3, 0.6)	2.6 (2.0, 3.3)	0.3 (0.1, 0.7)	0.1 (0.0, 0.2)	0.1 (0.1, 0.2)	0.1 (0.0, 0.2)	0.1 (0.0, 0.2)
	Arunachal Pradesh	12.2 (11.0, 14.1)	9.1 (8.1, 9.9)	1.6 (0.2, 1.3)	0.0 (0.0, 0.0)	0.4 (0.1, 1.1)	2.0 (1.0, 3.5)	0.6 (0.3, 1.3)	0.8 (0.3, 1.8)	0.1 (0.0, 0.1)	0.2 (0.03, 0.3)
	Assam	12.3 (11.0, 13.8)	9.2 (7.9, 10.4)	2.9 (0.5, 1.5)	0.0 (0.0, 0.0)	0.8 (0.4, 1.4)	1.5 (0.7, 2.6)	0.1 (0.1, 0.4)	0.3 (0.2, 0.6)	0.1 (0.0, 0.1)	0.5 (0.2, 1.3)
	Manipur	14.2 (12.8, 16.4)	9.3 (8.2, 10.9)	0.8 (0.3, 1.4)	0.0 (0.0, 0.0)	0.7 (0.2, 1.3)	3.1 (2.4, 4.5)	0.3 (0.1, 0.8)	0.4 (0.2, 1.1)	0.0 (0.0, 0.1)	1.5 (1.4, 2.6)
Northeastern	Meghalaya	14.3 (12.4, 16.8)	7.9 (6.6, 8.8)	0.8 (0.4, 2.1)	0.0 (0.0, 0.1)	0.5 (0.2, 1.6)	4.9 (3.1, 7.4)	2.0 (1.0, 3.8)	0.3 (0.2, 0.8)	0.1 (0.0, 0.2)	1.1 (1.0, 2.5)
	Mizoram	16.0 (13.8, 19.2)	6.8 (5.6, 8.0)	3.4 (1.0, 2.2)	0.0 (0.0, 0.1)	1.3 (0.6, 1.9)	6.9 (3.9, 10.6)	2.3 (1.2, 3.9)	2.3 (0.7, 3.2)	0.2 (0.1, 0.3)	0.6 (0.1, 1.0)
	Nagaland	17.6 (15.7, 19.6)	7.7 (6.9, 8.6)	1.5 (1.0, 1.9)	0.1 (0.0, 0.1)	1.5 (0.9, 1.8)	7.4 (5.0, 10.1)	4.2 (2.6, 5.9)	0.7 (0.4, 1.1)	0.0 (0.0, 0.1)	1.0 (0.7, 2.5)
	Sikkim	13.1 (12.2, 14.2)	9.8 (9.0, 10.7)	1.6 (1.2, 2.2)	0.1 (0.0, 0.2)	1.5 (1.1, 2.0)	1.9 (0.8, 2.8)	0.7 (0.3, 1.4)	0.7 (0.4, 1.3)	0.1 (0.0, 0.2)	0.1 (0.0, 0.2)
	Tripura	13.1 (11.8, 14.5)	9.0 (8.0, 10.0)	1.2 (0.7, 1.8)	0.0 (0.0, 0.0)	1.1 (0.4, 1.7)	1.9 (1.1, 3.0)	1.0 (0.5, 1.7)	0.3 (0.1, 0.6)	0.0 (0.0, 0.1)	0.3 (0.0, 0.5)

Data are presented as weighted median (IQR). <sup>a</sup>Plant protein includes cereals, pulses, legumes, fruits, vegetables (including roots and tubers), nuts and oilseeds. <sup>b</sup>Dairy protein includes fermented dairy protein (Indian style yoghurt (curd), buttermilk, lassi) and non-fermented dairy protein (all types of milk (cow/buffalo/yak/goat), milk added to coffee and tea, paneer (curdled milk solids)). <sup>c</sup>Animal protein includes red meat, poultry, egg, and fish and does not include dairy protein. <sup>d</sup>Undivided.

**Extended Data Table 6 | Odds ratios (95% confidence intervals) for generalized and abdominal obesity in isocaloric substitution analysis for the replacement of 5% energy from carbohydrates with 5% energy from other macronutrients in the study population (n=18,090)**

Replacement nutrient (%E)	5% Carbohydrate	
	Generalised obesity <sup>a</sup>	Abdominal obesity <sup>b</sup>
Total protein (%E)	0.97 (0.90, 1.04)	0.99 (0.91, 1.07)
Plant protein <sup>c</sup> (%E)	0.94 (0.84, 1.05)	1.06 (0.93, 1.20)
Cereal protein (%E)	0.93 (0.83, 1.04)	1.06 (0.94, 1.20)
Pulses and legumes protein (%E)	0.94 (0.84, 1.05)	1.06 (0.94, 1.21)
Dairy protein <sup>d</sup> (%E)	1.08 (0.98, 1.19)	1.10 (0.98, 1.23)
Fermented dairy protein (%E)	1.01 (0.90, 1.13)	1.03 (0.91, 1.16)
Non-fermented dairy protein (%E)	1.06 (0.96, 1.18)	1.11 (0.97, 1.26)
Animal protein <sup>e</sup> (%E)	0.97 (0.92, 1.04)	0.95 (0.89, 1.02)
Red meat protein %E	0.97 (0.90, 1.04)	0.94 (0.87, 1.02)
Poultry protein (%E)	0.99 (0.93, 1.06)	0.97 (0.89, 1.05)
Egg protein (%E)	0.98 (0.92, 1.06)	0.94 (0.87, 1.02)
Fish protein (%E)	0.99 (0.93, 1.06)	0.94 (0.87, 1.02)
Plant protein with dairy (%E)	1.00 (0.91, 1.10)	1.08 (0.98, 1.20)
Animal protein with dairy (%E)	1.00 (0.94, 1.05)	0.95 (0.90, 1.01)
Total fats (%E)	1.05 (1.02, 1.09)	1.00 (0.96, 1.04)
Saturated fat (%E)	1.04 (0.99, 1.11)	1.01 (0.94, 1.08)
Monounsaturated fat (%E)	1.02 (0.97, 1.08)	0.94 (0.88, 1.00)
Total polyunsaturated fat (%E)	1.01 (0.95, 1.07)	0.94 (0.88, 1.01)
n6 Polyunsaturated fat (%E)	1.02 (0.96, 1.07)	0.94 (0.88, 1.01)
n3 Polyunsaturated fat (%E)	1.01 (0.95, 1.08)	0.94 (0.87, 1.01)

Two-sided p-values were used in calculating odds ratios and 95% confidence intervals via survey logistic regression. Models were adjusted for primary sampling unit (PSU), state code, population weight, age (in years), sex (male/female), family history of type 2 diabetes (yes/no), tobacco use (never/current/past), alcohol use (never/current/past), physical activity level (sedentary/moderate/vigorous), time period the study was conducted, education qualification (primary school/high school/college), body mass index (BMI) (kg/m<sup>2</sup>) (except for generalized obesity), place of residence: urban vs rural, intakes of total energy (kcal/day). It was further adjusted for newly diagnosed type 2 diabetes, newly diagnosed prediabetes, hypertension, dyslipidaemia, and abdominal obesity (except in models with abdominal obesity as an outcome). Models were additionally adjusted for all nutrients except the nutrient being replaced.

<sup>a</sup>Generalized obesity was defined as body mass index (BMI)  $\geq 25$  kg/m<sup>2</sup>. <sup>b</sup>Abdominal obesity was defined as waist circumference (cm)  $\geq 90$  cm for male and  $\geq 80$  cm for female. <sup>c</sup>Plant protein includes cereals, pulses, legumes, fruits, vegetables (including roots and tubers), nuts and oilseeds. <sup>d</sup>Dairy protein includes fermented dairy protein (Indian style yoghurt (curd), buttermilk, lassi) and non-fermented dairy protein (all types of milk (cow/ buffalo/yak/goat), milk added to coffee and tea, paneer (curdled milk solids)). <sup>e</sup>Animal protein (%E) includes red meat, poultry, and egg and egg products and fish and does not include dairy protein.

**Extended Data Table 7 | Odds ratios (95% confidence intervals) for cardiometabolic risk factors in substitution analysis for the replacement of 50g refined cereals with 50g wholegrains’ subtypes in the study population (n=18,090)**

Food groups (g/ day)	50 g Refined cereals			
	Newly diagnosed type 2 diabetes <sup>a</sup>	Newly diagnosed prediabetes <sup>b</sup>	Generalised obesity <sup>c</sup>	Abdominal obesity <sup>d</sup>
Milled wholegrains	1.01 (0.66,1.56)	1 (0.72,1.38)	0.99 (0.71,1.39)	1.01 (0.69,1.46)
Milled whole wheat <sup>f</sup>	0.94 (0.57,1.56)	0.96 (0.65,1.42)	1.00 (0.66,1.52)	1.04 (0.65,1.66)
Milled whole millets <sup>g</sup>	0.96 (0.56,1.67)	1.04 (0.68,1.58)	0.96 (0.62,1.51)	1.08 (0.66,1.76)

Two-sided p-values were used in calculating odds ratios and 95% confidence intervals via survey logistic regression. Total energy adjusted by residual method was used for all the dietary exposure variables (refined cereals (g/day), milled wholegrains (g/day), milled whole wheat (g/day) and milled whole millets (g/day)). Models were adjusted for primary sampling unit (PSU), state code, population weight, age (in years), sex (male/female), family history of type 2 diabetes (yes/no), tobacco use (never/current/past), alcohol use (never/current/past), physical activity level (sedentary/moderate/vigorous), time period the study was conducted, education qualification (primary school/high school/college), body mass index (BMI) (kg/m<sup>2</sup>) (except for generalized obesity), place of residence: urban vs rural, intakes of total energy (kcal/day), fruits (g/day), vegetables (g/day), tubers (g/day), milk and milk products (g/day), pulses and legumes (g/day), animal foods (g/day), nuts and oilseeds (g/day), and edible oils and fats (g/day). It was further adjusted for newly diagnosed type 2 diabetes (except in models with newly diagnosed type 2 diabetes and newly diagnosed prediabetes as outcomes), newly diagnosed prediabetes (except in models with newly diagnosed type 2 diabetes and newly diagnosed prediabetes as outcomes), hypertension, dyslipidaemia, and abdominal obesity (except in models with generalized and abdominal obesity as an outcome). Models were additionally adjusted for all food groups except the food group being replaced. <sup>a</sup>Newly diagnosed type 2 diabetes was defined as glycated hemoglobin (HbA1c) ≥6.5% and/or if fasting capillary blood glucose ≥126 mg/dL (≥7 mmol/L) or 2-h post glucose load capillary blood glucose ≥220 mg/dL (≥12.2 mmol/L). <sup>b</sup>Newly diagnosed prediabetes was defined as glycated hemoglobin (HbA1c) 5.7%–6.4% or if fasting capillary blood glucose was ≥100–125 mg/dL (≥5.6–6.9 mmol/L) and/or if 2-h post load capillary blood glucose value was ≥160 and <220 mg/dL (≥8.9 to <12.2 mmol/L). <sup>c</sup>Generalized obesity was defined as body mass index (BMI) ≥25 kg/m<sup>2</sup>. <sup>d</sup>Abdominal obesity was defined as waist circumference of ≥90 cm for males and ≥80 cm for females. <sup>e</sup>Milled wholegrains include milled whole wheat and milled whole millet as flours. <sup>f</sup>Milled whole wheat includes foods made from whole wheat grits and flour. <sup>g</sup>Milled whole millets include foods made from millet grits and flours. For milled wholegrains, the analysis only included participants who consumed ≥50 g.



## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- |                                     |                                     |  |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A description of all covariates tested   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated   |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

### Software and code

Policy information about [availability of computer code](#)

Data collection No Software was Used

Data analysis

All statistical analyses were performed with R software, version 4.3.3 and SAS version 9.4 (SAS Institute Inc., Cary, NC, USA)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Provide your data availability statement here.

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

### Reporting on sex and gender

The term sex (biological attribute) and gender (shaped by social and cultural circumstances) were appropriately use. Findings of the study apply to both sexes; adults of either sex were included in the study; sex was determined based on self reporting. The information was collected after obtaining informed consent for sharing individual level data; 49 percent of the participants were male. Sensitivity analysis stratified by sex has been included as part of the findings of the study.

### Reporting on race, ethnicity, or other socially relevant groupings

All participants belonged to the Asian Indian race. Participants were categorized by geography (regions of India), sex and place of residence. The data was reported by the respondents themselves and were in accordance with the population census.

### Population characteristics

Same as above

### Recruitment

*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*

### Ethics oversight

Institutional Ethics committee at Madras Diabetes Research Foundation

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences

☐ Behavioural & social sciences

☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

### Sample size

Sample size calculations were performed based on the primary objective of determining the national prevalence of diabetes mellitus in India while maintaining enough power to demonstrate robust estimates when stratified by state. This was done primarily because health is a "state" subject in India. The sample size was calculated separately for urban and rural areas, as previous studies show large variations in urban and rural prevalence of type 2 diabetes mellitus (approximately 10% and 4%, respectively). Using a precision of 20% and allowing for a nonresponse rate of 20%, the sample size was calculated to be 2800 individuals in rural areas and 1200 individuals in urban areas, with a total sample size of 4000 individuals per state. Hence, the total sample size for the study, including all 28 states, NCT of Delhi, and two UTs, was 124,000 individuals (4000 × 31). The national prevalence will be calculated from the state prevalence using weight-age analysis. Of the 113,043 individuals, every fifth individual were invited to collect information on diet.

### Data exclusions

We excluded participants with self-reported type 2 diabetes (n=1,404), HbA1c of <4% (n=26) or if they had implausible dietary data defined as a reported energy intake (n=1,340 with values <500 and >4200 kcal/ day). After these exclusions, the final analytic sample was 18,090 adults (49% male and 27% urban) for whom biochemical and dietary assessments were available.

### Replication

Sudha V, Anjana RM, Vijayalakshmi P, Lakshmi Priya N, Kalpana N, Gayathri R, Priyadarshini RD, Malini HM, Anitha C, Deepa M, Raj SS, Parthiban K, Ramakrishnan R, Geetha G, Krishnaswamy K, Unnikrishnan R, Mohan V. Reproducibility and construct validity of a food frequency questionnaire for assessing dietary intake in rural and urban Asian Indian adults. Asia Pac J Clin Nutr. 2020;29(1):192-204. doi: 10.6133/apjcn.202003\_29(1).0025. PMID: 32229459.

### Randomization

Not relevant as it is a cross-sectional/observational epidemiological study

### Blinding

Not relevant as it is a cross-sectional/observational epidemiological study however the data analyst/biostatistician were blinded to data

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials &amp; experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

## Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration	The study was registered with the Clinical Trials Registry of India (CTRI/2019/03/018095).
Study protocol	Anjana, R. M. et al. The Indian Council of Medical Research-India Diabetes (ICMR-INDIAB) study: methodological details. J. Diabetes Sci. Technol. 5, 906-914 (2011)
Data collection	The Indian Council of Medical Research-India Diabetes (ICMR-INDIAB) national study is a cross-sectional population-based survey that includes a representative sample of individuals aged 20 years and older drawn from 30 states, union territories (UTs), and the National Capital Territory (NCT) of Delhi. To be nationally representative, the survey was conducted in multiple phases using a stratified multistage sampling design using three-level stratification based on geography, population size, and socioeconomic status of each state. In Phase 1 (2008 to 2010) individuals from Tamil Nadu, Chandigarh, Jharkhand and Maharashtra were recruited into the study. Phase 2 was conducted from 2012-2013 and included undivided Andhra Pradesh (subsequently divided into Andhra Pradesh and Telangana), Bihar, Gujarat, Karnataka and Punjab. Phase 3 (2017-2018) included Delhi, Madhya Pradesh, Rajasthan and Uttar Pradesh while phase 4 (2018-2019) included Kerala, Goa, Puducherry, Haryana and Chhattisgarh. The Northeast phase was conducted between 2011-2017 and included the states of Assam, Arunachal Pradesh, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim, and Tripura. The final phase (phase 5) was conducted between 2019-2020 and covered Himachal Pradesh, Uttarakhand, Odisha and West Bengal.
Outcomes	Leveraging dietary data collected among every fifth participant, we aimed to characterize the dietary profile of India across regions and examine intra and inter- regional similarities and differences in intakes of macronutrients. Given the high carbohydrate content of Indian diets, we evaluated the effects of isocaloric substitution of carbohydrates with other macronutrients on measures of glycemia and adiposity. In all study participants, a structured questionnaire was administered and anthropometric parameters and blood pressure were measured. Fasting capillary blood glucose was first determined using a glucose meter. An oral glucose load was administered to all participants except those with self-reported diabetes, and the 2 h post-load capillary blood glucose was estimated. Additionally, in every fifth subject, a fasting venous sample was collected for measurement of lipids and validated FFQ questionnaire was administered for dietary assessment questionnaire.

## Plants

Seed stocks	<i>Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.</i>
Novel plant genotypes	<i>Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.</i>
Authentication	<i>Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.</i>