scientific reports

Check for updates

OPEN

# Healthcare utilisation and economic burden of cancer on Indian households

Md. Azharuddin Akhtar[1], Indrani Roy Chowdhury[2]✉ & Bhawna Taneja[2]

**India is experiencing a rising incidence of cancer with a high mortality rate at a younger age. With high catastrophic healthcare expenditures and inadequate social security, it is imperative to investigate the health-seeking behaviour and corresponding economic burden on cancer-affected households. Using a nationally representative database, we use the matching methods to analyse healthcare utilization, healthcare expenditure, and financial distress of cancer-affected households. We find that the cancer-affected households report a longer hospital stay, more surgery, medicine uptake, and diagnostic tests per member compared to non-cancer households. The out-of-pocket healthcare expenditure borne by the affected households are significantly higher and mostly inflated through inpatient care. Further, we find some adverse spill-over effect in terms of lower per-member inpatient and outpatient visits and lower healthcare expenditure for non-cancer members of affected households when they are afflicted with any diseases. The catastrophic expenditures inflict income loss and distress financing to the cancer affected households. Further, we find per-member work force participation and non-medical consumption expenditures significantly low. We also observe heterogeneous effect in terms of lower inpatient visits, lower medicine uptake, lower capability to pay for treatment, and significantly higher borrowing and selling off assets among various cancer affected socioeconomically disadvantaged groups.**

**Keywords** Healthcare utilization, Out-of-pocket healthcare expenditure, Distress financing, Cancer affected households, Non-cancer affected households, Matching methods, Heterogeneity analysis

Cancer is the uncontrolled exponential growth of abnormal malignant cells (cancerous or neo-plasma cells) in any part of the body. Most of the Cancer cases are reported in vital organs like the lung, stomach, colorectal, liver, and breast[1,2]. It is the second leading cause of mortality, accounting for 13% of total global deaths (9.6 million deaths) in 2018 alone[3]. Moreover, the low-income countries share a significantly high Cancer mortality (around 70% of all Cancer deaths across the globe[4]) than high-income countries[5–7]. India has the third highest prevalence of cancer in the world after China and the United States of America[8]. The incidence of cancer in India as given by different sources is presented in Table 1. The National Cancer Registry Programme (NCRP) reported an incidence of cancer of 100.4 per 100,000 population. It is around 86.5 and 145.47 (per 100000 population) as reported by NSS and NFHS respectively. Among the total cancer cases, breast cancer (47.9 per 100,000 population) and prostate cancer (30.5 per 100,000 population) have the highest prevalence[9]. These estimates are expected to rise by around 13% in the year 2025[10]. In addition, almost 1,300 people die every day, with greater prevalence in urban regions[3,11].

According to a Lancet study[4], close to 70% of Cancer patients die either because of delayed treatment-seeking (more than 60% of Cancer-affected individuals visit healthcare when the cancerous cell spreads across local regions; see Supplementary Appendix: Fig. A1) due to lack of awareness, financial constraints, or inaccessibility to the diagnostic facilities. The incidence and mortality due to cancer are spatially heterogeneous with some of the southern Indian states along with Punjab and Mizoram having the same levels (see Fig. 1). According to India's latest National Family and Health Survey (NFHS 2019-21), only 81.4% of the reported Cancer patients managed to continue treatment in India. The proportion is even almost 10% points lower in the rural India and likewise for the poorest wealth quintiles. Hence, health-seeking behaviour is majorly contingent upon the socioeconomic and regional gradients[12].

The utilisation of primary care, speciality care, and hospital services by Cancer-affected households can significantly differ from households having members afflicted with other leading diseases. The former may require frequent and intensive treatments, surgeries, hospitalisations, and specialised care, often over an extended

[1]Institute for Human Development, Delhi, NCR, India. [2]Centre for the Study of Regional Development, Jawaharlal Nehru University, New Delhi, India. ✉email: indranircjnu@gmail.com

| Cancer incidence (per 100,000) | Total | Male | Female |
|---|---|---|---|
| NCRP (2022)* | 100.4 | 95.6 | 105.4 |
| NSS (2017-2018)** | 86.5 | 102.05 | 69.96 |
| NFHS (2019-2021)*** | 145.47 | 281.8 | 124.5 |

**Table 1**. **Incidence of cancer in India.** *Crude rate estimated from NCRP 2022 data. **Authors' calculation from National Sample Survey 75th Round (2017–2018) (using appropriate sampling weights). ***Authors' calculation from the National Family Health Survey (NFHS 2019–2021) (using appropriate sampling weights). Incidence for the age group 15–49 years only. #The numbers present an overview of the cancer incidence from different data sources. These are however not comparable due to the nature of the data. NCRP uses reported cancer cases as per Population-Based Cancer Registries (PBCRs) and Hospital-Based Cancer Registries (HBCRs). NSS and NFHS on the other hand are sample surveys based on specific sampling designs. Moreover, NFHS provides data for the age group of 15–49 years only.

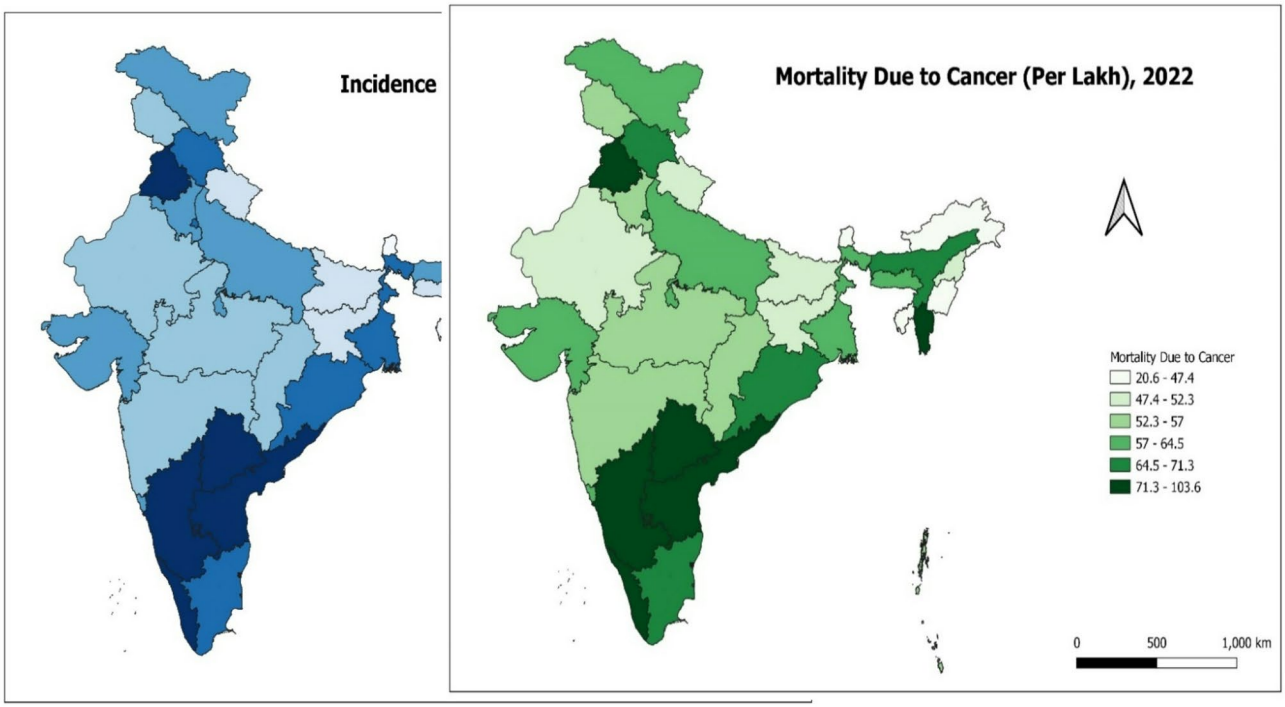


**Fig. 1**. **Estimated Incidence and Mortality of Cancer per Lakh Population in India, 2022.** *Source* Open Government Data Platform, 2022, India

period. Table 2 highlights the utilization of healthcare by individuals (in inpatient and outpatient facilities) with categorised ailments (including Cancer). We can see that the intensity of healthcare utilized (in terms of length of hospital stay, medicine, and other diagnostic tests) in Cancer outweighs other major diseases. This pattern may have serious implications in terms of financial stress to the cancer-affected households as compared to the households with reported non-cancer diseases. Moreover, the economic burden is supposed to be significantly high on households, from the vulnerable socioeconomic background.

Cancer-detection therapy generally involves recurrent and exorbitantly costly healthcare interventions. Along with the direct and indirect costs of treatment, it also entails substantial opportunity costs in terms of loss of employment, income, and consumption rationing for the *Cancer-affected (CA) households*[13–16]. The hospitalisation expenditure for Cancer is 2.5 times more than the average hospitalisation expenditure in India[17]. According to India's NSS 75th Round (2017-18) database, more than 75% of *CA households* incur catastrophic expenditures, and close to 44% meet their medical expense primarily through borrowing and selling assets (see Table 3). Moreover, Cancer strongly dominates any other major diseases in terms of the intensity of healthcare utilization (like inpatient visits, outpatient visits, length of hospital stay, duration of treatment, screening diagnostics etc.) and in terms of overwhelmingly high economic burden (direct inpatient expenditure, catastrophic health expenditure, financial distress etc.) (see Tables 2 and 3). In the face of such catastrophic effects, social security support (health insurance) is also very limited and inadequate. In India, the budgetary allocation of public-health expenditure is slightly above 1% of GDP, and only 14% of Indians are covered by any health insurance scheme (NSS Report 2017–18[18]). Although there is an explicit provision for Cancer

| Diseases | Observation | Inpatient visit (%) | Length of stay in hospital | Outpatient visit (%) | Surgeries (%) | Medicine (%) | Screening (%) | Diagnostic (%) |
|---|---|---|---|---|---|---|---|---|
| Infection | 30,402 | 26 | 1.32 | 79 | 5 | 96 | 23 | 41 |
| Cancer | 1726 | 81 | 10.95 | 47 | 39 | 100 | 81 | 83 |
| Blood-diseases | 1460 | 48 | 2.56 | 64 | 6 | 94 | 43 | 70 |
| Endocrine-metabolic-nutritional | 6494 | 7 | 0.48 | 100 | 5 | 100 | 10 | 35 |
| Psychiatric-neurological | 5195 | 34 | 3.01 | 84 | 8 | 96 | 37 | 40 |
| Eye | 2391 | 56 | 1.70 | 67 | 49 | 101 | 49 | 59 |
| Ear | 483 | 30 | 1.93 | 100 | 21 | 100 | 32 | 33 |
| Cardio vascular | 11,021 | 18 | 1.23 | 100 | 7 | 100 | 22 | 31 |
| Respiratory | 6230 | 13 | 0.80 | 97 | 6 | 97 | 15 | 20 |
| Gastro-intestinal | 8173 | 47 | 3.04 | 68 | 23 | 99 | 49 | 53 |
| Skin | 1249 | 33 | 4.29 | 40 | 23 | 43 | 32 | 40 |
| Musculoskeletal | 5319 | 37 | 4.69 | 59 | 32 | 55 | 43 | 44 |
| Genito-urinary | 4335 | 66 | 4.61 | 54 | 37 | 97 | 69 | 74 |
| Injuries | 10,214 | 69 | 5.54 | 47 | 37 | 100 | 69 | 69 |

**Table 2**. **Treatment sought by the individuals with reported ailments in leading diseases.** Authors' calculation; Source: Social Consumption: Health Survey- NSS 75th Round (2017–2018).

| Diseases | Observation | ME (INR) | TE (INR) | Total IE (INR) | CHE (%) | Finan. dist. (%) |
|---|---|---|---|---|---|---|
| Infection | 20,651 | 9085.8 | 1356.2 | 10,442 | 20.15 | 20.42 |
| Cancers | 1531 | 62525.7 | 5818.0 | 68343.7 | 75.31 | 44.42 |
| Blood diseases | 1190 | 13561.36 | 1809.4 | 15370.76 | 38.15 | 25.46 |
| Endocrine metabolic nutritional | 1827 | 17359.77 | 2273.9 | 19633.69 | 41.54 | 23.37 |
| Psychiatric neurological | 3793 | 26644.06 | 2992.5 | 29636.57 | 52.54 | 31.24 |
| Eye | 2083 | 10897.74 | 1289.5 | 12187.19 | 40.37 | 19.73 |
| Ear | 336 | 14428.69 | 1849.7 | 16278.38 | 41.07 | 23.81 |
| Cardio-vascular | 6081 | 36311.29 | 2914.7 | 39225.95 | 49.79 | 26.54 |
| Respiratory | 2655 | 14099.84 | 1894.8 | 15994.66 | 30.09 | 23.43 |
| Gastro-intestinal | 6899 | 19680.42 | 2472.5 | 22152.9 | 44.05 | 26.84 |
| Skin | 630 | 20577.83 | 2354.5 | 22932.29 | 34.29 | 22.86 |
| Musculoskeletal | 2886 | 32476.37 | 3494.5 | 35970.88 | 47.61 | 27.55 |
| Genito-urinary | 3907 | 24634.31 | 2471.6 | 27105.95 | 57.00 | 32.02 |
| Injuries | 9353 | 26825.83 | 2858.4 | 29684.24 | 50.06 | 30.89 |

**Table 3**. **Economic burden of leading diseases in India (2017-2018)** Authors' calculation. Source: Social Consumption: Health Survey- NSS 75th Round (2017-18). *INR* India Rupee (Rs.), ME, TE, and Total IE are medical expenditure, transport and other miscellaneous expenditure, and total inpatient health expenditure over the period of 365 days, *CHE* Catastrophic health expenditure—The percentage of household incurring catastrophic expenditure at 10% expenditure threshold level, *Finan. Dist* Financial distress in the form of debt and selling of assets.

treatment in health benefit packages (HBPs) under social health insurance coverage, it is confined to cover inpatient expenditure alone, leaving the affected households exposed to high expenditure spread in outpatient care, diagnostics and follow-up stages[19].

## Literature review

The economic burden of cancer is multifaceted, encompassing direct medical costs, productivity losses, and intangible costs that vary significantly across regions and populations. In the international context, the financial burden of cancer for the EU was €126 billion in 2009, with healthcare accounting for 40% of the expenses, varying from €16 to €184 per person across the member states[20]. The excess economic burden on U.S. cancer survivors, with those recently diagnosed facing higher annual costs, is primarily driven by medical expenses[21]. A steady rise in Canadian cancer care costs on account of hospital-based care was observed from $2.9 billion in 2005 to $7.5 billion in 2012[22]. Studies highlight the methodological challenges in assessing comprehensive economic impact including direct and indirect costs of global burden of cancer, where indirect costs dominate the economic burden of most cancers, except prostate and breast cancer[13,23]. Chronic conditions among cancer survivors compound their financial hardship, with increased medical costs, lost productivity and poor quality of life[21,24]. The financial strain extends to caregivers as well, stemming from employment loss and care complexity[25].

Studies demonstrated high out-of-pocket expenses, especially in low- and middle-income countries, where patients and caregivers spend significant portions of their income on cancer care[26].

The existing literature on the burden of Cancer in India is documented in the context of epidemiology, demographic, and catastrophic expenditure[16,17,19,27–30] however its impact has not been comprehensively addressed at the household level. These studies mostly have not undertaken any causal analysis and examined the economic burden of disease in exclusion and hence potentially overestimated the impact of cancer in the presence of multiple morbidities of other members in the household. Moreover, these studies overlooked the indirect impact of Cancer beyond the healthcare utilisation and corresponding catastrophic healthcare expenditure. Thus, the coping behavior of the CA households with financial distress (in terms of indebtedness, borrowing, selling of assets, compromised consumption etc.) and a potential adverse spill-over effect on the other members of the households (in terms of compromised healthcare utilisation, consumption, withdrawal from the labour market for caregiving, etc.) are ignored. To the best of our knowledge, there are only a few empirical studies which attempted to investigate the causal impact of Cancer on CA households in India[16,31,32]. However, these studies have methodological limitations. They may have overestimated the impact of cancer due to the inclusion of households without any episode of ailment as a benchmark category. Further, these studies are limited in capturing the possible heterogeneous effect across socioeconomic groups with poor health insurance coverage. Besides, they fail to check the sensitivity of their results due to the possible influence of unobservable confounding effects on the treatment group. None of these studies reflects on the institutional aspects (public or private healthcare) of healthcare utilization and the implications on the OOP expenditure.

Therefore, the paper contributes to the narrow literature on the causal impact on economic burden of households due to Cancer health shocks. By addressing all methodological limitations in the existing literature, we investigate the healthcare utilization, healthcare expenditure, distress financing, and other potential adverse spill-over effects on the affected households. Hence, we seek to address the following research questions:

1. Do the health care utilizations of the CA households differ significantly compared to NCA households?
2. Do the CA households bear excess economic burden as compared to NCA households?
3. Given the inadequacy of health security coverage, whether there is heterogeneous effect of Cancer on households?

We addressed the first two questions using a quasi-experimental technique, i.e., the *Propensity Score Matching (PSM)* method and the last question through *Regression Adjusted Matching*.

The paper is organised as follows: Sect. 2 presents a brief literature review, and Sect. 3 discusses the dataset used for the study and the estimation technique. Section 4 presents the empirical findings, followed by a discussion of the results. Section 5 concludes the study and states the policy implications.

## Data and research methodology

### Data

The study utilises unit-level large-sample data extracted from the National Sample Survey (NSS) (Social Consumption: Health Survey-NSS 75th Round (2017-18)). It is a multi-stage stratified survey based on a meticulous survey design that collects data on Indian households' and individuals' socioeconomic and demographic background, morbidity, deaths, healthcare utilization, health expenditure etc. A total of 113,823 households are surveyed in this dataset. Since this study tries to investigate the impact of the economic burden of Cancer on households, the treatment group is CA households, i.e., the households that reported cancer by any member who were seeking treatment and the control group is non-cancer affected (NCA) households, i.e., the households that reported any chronic disease other than cancer by any member who were seeking treatment for the same. By restricting the sample to households with reported chronic ailments, we maintain the integrity of the comparative analysis, ensuring that the economic burden reflects real treatment-seeking behaviour and associated costs.

The detailed classification of outcome and control variables is provided in Appendix: Tables A1 and A2. In total, our sample size is 69,329 (treatment and control), of which 1,217 are CA households (treatment) and 68,112 are NCA households (control).

### Estimation technique

We follow a framework proposed by Anderson's Health Behavioural model[33,34], presented in Fig. 2, which offers a robust conceptual foundation for understanding healthcare utilization by categorizing determinants into *predisposing*, *enabling*, *and need factors*. In the context of cancer, these determinants shape household's decision-making process regarding whether and how to seek treatment optimally. Based on this framework, we propose a simple theoretical model (in a static framework) (see Supplementary Appendix A1) for optimal healthcare utilization in the presence (or absence) of healthcare insurance and other provisions of informal insurance (like social networks for easing healthcare accessibility, financial assistance from relatives, friends and neighbours). The static model formalizes a utility-maximizing framework, thereby offering a microeconomic insight into cancer healthcare utilization decisions under resource constraints. The out-of-pocket expenditure is the source of financing the cancer treatment in the absence of any insurance. In such a situation, the household faces severe economic hardship and may face catastrophic health expenditure and financial distress due to depletion of savings or indebtedness.

The empirical exercise presented in the following section is conceived from this theoretical framework.

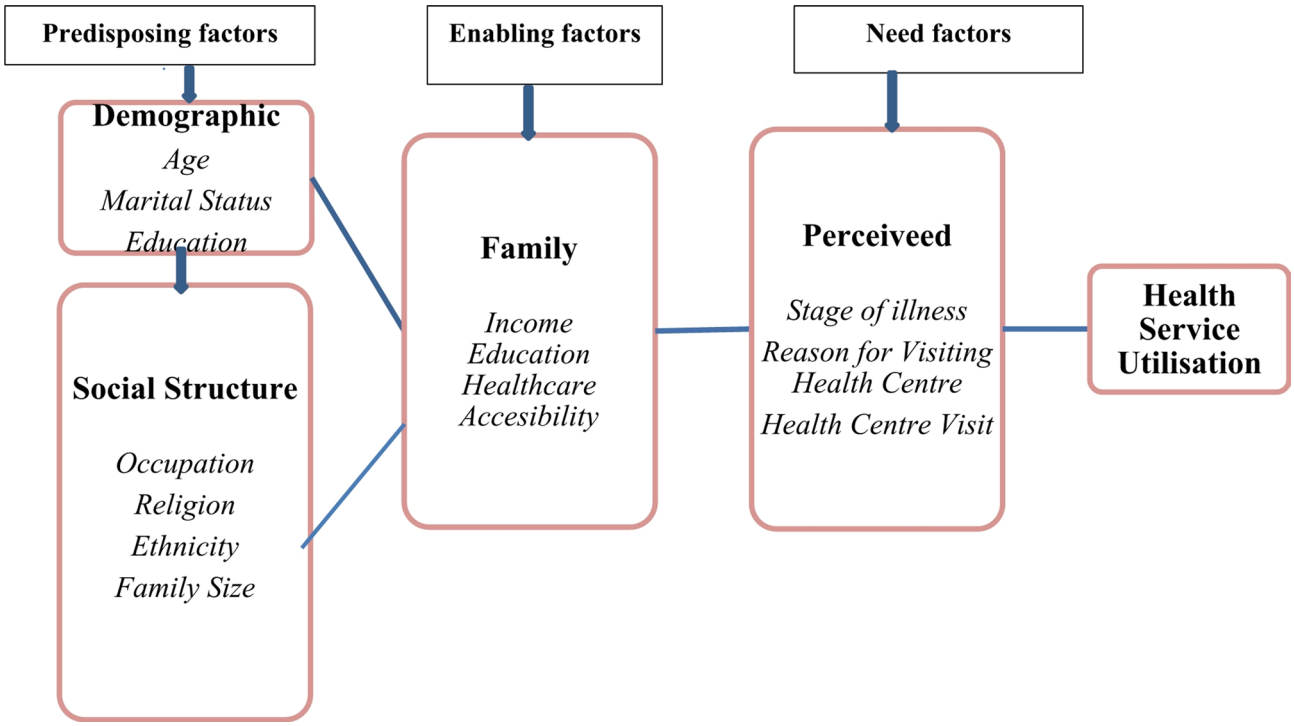


**Fig. 2. Andersen's behavioural model of healthcare services.**

*Propensity score matching*

The economic burden of Cancer on households can be attributed in terms of *healthcare utilisation and healthcare expenditure.* The outcome variables for capturing the impact of healthcare utilisation include inpatient visits, length of stay in the hospital, number of surgeries, medicine uptake, diagnostic tests and outpatient visits. The outcome variables for economic burden are characterised under healthcare expenditure (includes medical expenditure, transportation expenditure and total expenditure in respective inpatient and outpatient cares), and financial hardship as well as coping strategies (include borrowing and selling of assets, and effect on income, work, and non-medical consumption) (see Table A1 and A2 in the Appendix). It is also likely that the CA households may experience induced adverse spill-over effects on the rest of the family members due to financial constraints (owing to the exorbitant OOP healthcare expenditure of the Cancer-affected member). Therefore, we explore the effect by analysing the rest of the household members' healthcare utilisation (highlighted as *non-cancer diseases* (all major diseases except cancer) and *non-major diseases* (all diseases with the exception of cancer, diabetes, stroke, and heart disease)) and healthcare expenditure.

Since there can be multiple ailing members with multiple healthcare utilisation in a single household, we have standardised the outcome with respect to household size to make outcomes comparable between CA and NCA. We have also analysed the institutional differences in healthcare delivery between public and private healthcare. The treatment effect (of Cancer) is thus interpreted in terms of *per household member (p.m).* For example, *'Inpatient visit p.m.'* implies, average inpatient visits of household members when affected by ailments.

The effect of Cancer on households in terms of healthcare utilisation, healthcare expenditure, and financial hardships (including coping strategy) is explained by the following equation:

$$Y_i = f\left(X_i, D_i\right) \tag{1}$$

Where $Y$ denotes the outcome variables (various indicators of healthcare utilisation or economic burden) for household *i*, *D* is a dummy which is equal to 1 if any member of a household is affected by Cancer and 0 otherwise, and X is a vector of covariates that might influence the outcome. The assignment of treatment and control is not random as the CA and NCA households may differ in terms of demographic, socioeconomic, and regional characteristics and in the presence of unobservable heterogeneous Cancer risk factors. Given that the treatment (Cancer) is not likely to be random ensuring internal validation is challenging. Furthermore, health-seeking behaviour in the absence of social health security becomes self-selective in nature and is subject to the socioeconomic gradient. This may lead to potential selection bias due to the violation of OLS assumptions and may yield inconsistent estimates. In other words, the covariance of D and the error term would be non-zero ($Cov\left(D,\ \epsilon\right) \neq\ 0$) if the unobservable factors influence the self-reporting of Cancer. Therefore, we propose to adopt the *propensity score matching* (PSM) method to estimate Eq. (1). The quasi-experimental matching technique attempts to "balance" the distribution of covariates in the treated and the control group to mimic the randomisation framework (see the Supplementary Appendix A2) for the methodological details of PSM). The

advantage of PSM compared to other parametric methods is that it does not rely on any functional assumption between the treatment and outcome[35], and is also robust to selection bias[36].

The primary objective is to estimate the *Average Treatment Effect on the Treated (*ATT) on health outcomes between CA and NCA households by using a non-experimental secondary dataset. Thus,

$$ATT_i = E\left(Y_{1i}|D_i = 1\right) - E\left(Y_{0i}|D_i = 1\right) \tag{2}$$

Where $Y_{1i}$ is a health outcome of a household *i* when any member has Cancer and $Y_{0i}$ is what the health outcome of a CA household *i* would have experienced when the member is afflicted with non-cancer diseases. The parameter $E\left(Y_{1i}|D_i = 1\right)$ is an average health outcome when a household is afflicted with cancer, while $E\left(Y_{0i}|D_i = 1\right)$ is a counterfactual, i.e., what would have been an outcome of a CA household in the absence of Cancer (but with other ailments). However, $E\left(Y_{0i}|D_i = 1\right)$ is unobserved, so cannot be measured. The estimator $E\left(Y_{0i}|D_i = 0\right)$ can be a measurable proxy for the counterfactual, but it would lead to a biased estimates because of the inherent pre-treatment difference between CA and NCA households. In order to estimate the a*verage treatment effect of the treated (ATT)*, the matching requires the unobserved $E\left(Y_{0i}|D = 1\right)$ to be replaced with observed household, say *t*, from the control group, such that it is matched to the observation *i* on a vector of pre-treatment X (i.e., $X_i \approx X_t$). The unmatched observations are pruned from the data set before further causal analysis. The implicit assumption is that the reporting for Cancer is conditioned on observed covariates. The PSM identifies the matched counterfactual by summarising the effect of vector X into a scalar quantity represented by conditional probability called propensity scores. It is specified as;

$$p\left(X_i\right) = Pr\left(D_i = 1|X_i\right) = E\left[D_i|\, X_i\right] \tag{3}$$

We predict the probability (propensity score) of a household with at least one member reporting for Cancer (CA) conditioned on X using the logit model,

$$p\left(X_{ij}\right) = \frac{e^{\beta\, X_i}}{1 + e^{\beta\, X_i}} \tag{4}$$

The households in the treatment and control groups with similar observable characteristics in the pre-treatment are supposed to have similar propensity scores. Thus, it ensures a situation similar to a quasi-randomisation where $\beta$ is a vector of parameters to be estimated for the treatment effect.

PSM requires that the mean of the respective covariates used for matching *should not be statistically different* for the control and the treated groups. The balancing property and area of common support assumption are checked before estimating ATT. We estimate the causal effect of cancer on households compared to other diseases using four popular propensity score matching techniques (*nearest-neighborhood*, *radius*, *kernel*, *and stratified matching*), given that each algorithm has its own advantage[37]. The ATT measures the average treatment effect on the treated, i.e., the effect of cancer on CA households.

$$ATT_i = E\left[E\left\{P_{1i}|D_{ij} = 1, p\left(X_i\right)\right\} - E\left\{P_{0i}|D_i = 0, p\left(X_i\right)\right\}|D_i = 1\right] \tag{5}$$

*Regression adjusted matching*

Given the negligible social health insurance coverage in India, the impact of cancer might have socio-economically heterogeneous effect on households. To check such heterogeneous effect of Cancer on households with different socioeconomic groups (based on education and social groups), we conduct a separate analysis by running a *two-stage regression-adjusted matching method*. To understand whether there is a significant difference in healthcare utilisation among the CA and NCA households based on the heterogeneous effect of socioeconomic status: (1) measured by educational attainment (Low Socioeconomic Status (LSES) for education below primary and (2) social groups (disadvantaged social categories in India -Scheduled Castes (SCs) and Scheduled Tribes (STs). These social categories are the historically discriminated and marginalized sections and have been therefore declared reserved categories for affirmative actions after India's independence.

We first identify the matched counterfactuals and drop the remaining unmatched samples before applying parametric regression models. The detailed estimation process is provided in the Appendix (see section A3).

The following regression-adjusted matching equation:

$$Y = \alpha + \beta\, CA + \gamma\, LSES\ (or\ ST/SC) + \delta\, CA * LSES\ (or\ CA * ST/SC) + \epsilon \tag{6}$$

Here, Y is an outcome variable, and CA is a binary indicator variable with 1 as cancer-affected household and 0 otherwise; (LSES)/(ST/SC) is a binary indicator variable with 1 being household is LSES, 0 otherwise (with 1 means household being ST/SC, 0 otherwise); and CA*LSES is an interaction term; $\epsilon$ is an error term; $\alpha, \beta, \gamma\ and\ \delta$ are the parameter of our interest. The coefficient $\beta$ is the difference in outcome for CA and matched NCA households belonging to the non-LSES (non-ST/SC) households. The sum of coefficient $\beta + \delta$ can be interpreted as the difference in outcome for CA households and matched NCA households for LSES (ST/SC). Thus, the test of equivalence or no difference in healthcare outcome under the Null hypothesis—$H_0$: $\delta = 0$, which assume that there no statistical '*difference*' in outcomes for CA households in LSES (ST/SC) as compared to Non LSES (ST/SC) households (i.e., equality in healthcare utilisation, healthcare expenditure, and financial hardship across groups, compared to the reference group) (see Appendix: Section A3 for methodological detail).

*Robustness and sensitivity analysis*

The given data reports that majority of the Cancer cases are hospitalisation cases. Moreover, NSS database does not provide the cause of death in a household. If we ignore these facts, it may overestimate the effect due to possible differences in healthcare utilization and expenditure due to death during the survey round and also the risks of retaining outliers in healthcare visit and health expenditure due to end stage treatment (during the last 365 day recall period). So, we have adopted different robustness strategies to avoid such pitfall. We conducted more heterogeneity analysis through several sub-sample estimation of ATTs by (1) including hospitalisation, (2) excluding hospitalisation, (3) dropping the households reporting for the death of any member; and (4) dropping the households in top 1% of the total expenditure incurred due to Cancer (to avoid potential upward bias in the impact of Cancer on households), and then estimating ATT using the scores respectively. We conducted couple of subgroups analysis separately for possible heterogeneous effect of Cancer on the disadvantaged households by taking education (household head with primary or below primary education level.

In a quasi-experimental framework, the assumption of *conditional mean independence* would be violated if there are unobservable factors affecting the assignment of treatment and control simultaneously, thereby leading to biased estimations. Therefore, it is important to check the sensitivity of the estimated result with respect to deviation from the identified assumption. We conduct a sensitivity analysis on PSM-based estimates to determine the potential bias due to the presence of unobservable. In Rosenbaum's sensitivity framework, the parameter $(\Gamma)$ is introduced to quantify the strength of potential unobserved confounders, representing the extent to which the study's conclusions might be influenced by factors not accounted for in the observed data[38].

## Empirical analysis

### Selection of matching variable

We identify the counterfactuals/comparison group (i.e., NCA households) that closely resembles the treatment group (CA households) in terms of observed characteristics (or confounding covariates). We broadly classify the observables confounding covariates in terms of demographic, socioeconomic, and regional profiles (see Supplementary Appendix Table A3). Table 4 captures the balancing statistics in terms of the PSM scores of the CA, matched NCA and unmatched CA households, percentage of the bias reduction through the matching and the corresponding t-value of the mean differences between the CA and matched NCA households across the selected covariates on which PSM is estimated. The Nearest-Neighborhood Matching Method is used to identify the matched NCA households. Since medical expenditure is also included in monthly household expenditure, we deliberately did not match households on consumption expenditure. This is because we also want to capture whether the CA households are rationing their non-medical consumption significantly (to meet the exorbitant OOP medical expenditure on account of Cancer treatment) as compared to NCA households. So, we use proxy variables like cooking and sanitation facilities (piped water, septic tank, garbage disposal) to determine socioeconomic status. The standard mean difference and variance ratio of covariates approached (0, 1) after matching (see Supplementary Appendix: Fig. A2), indicating a reduction in bias. The insignificant mean differences between CA and matched NCA households on covariates thus validate the required assumption of conditional mean independence. Sample means of covariates corresponding to matched NCA households are much closer to CA households than the unmatched households, indicating the overlap of common support. Thus, it reflects that the propensity scores estimated using the covariates have a similar distribution in treated (CA households) and comparison groups (matched NCA households) (see Supplementary Appendix Figure A3).

### Result and discussion

*The impact of cancer on households' healthcare utilisation*

We estimate the impact of cancer on healthcare utilisation and the corresponding economic burden (see Tables 5 and 6). We find that a CA household significantly has more per-member private and public inpatient healthcare visits compared to matched NCA households during the last 365-day recall period. Likewise, the intensity of healthcare utilization of the CA households, in terms of per-member length of hospital stay, surgeries, medicine, screening and other diagnostic tests, are significantly higher compared to per-member matched NCA households. The significant spill-over effect in terms of lower inpatient and outpatient visits for non-cancer and non-major diseases among other members of CA households is observed compared to NCA households. There might be two possible scenarios: either average CA households have a lower probability of disease incidence or the average CA households suffer *severe financial constraints and staggering opportunity costs that restrict the other members from seeking treatment*. However, further investigation is needed as it is difficult to decipher the relative intensity with the available data. The critical level of $\Gamma$ is highlighted in the last column to tell us the level at which estimates of the treatment effect are questionable. The higher value of $\Gamma$ indicates that treatment effects are robust and insensitive to hidden bias. For example, the critical levels of $\Gamma$ for outcome variables like length of hospital stay, inpatient visits for a non-cancer ailment, and OPD visits are greater than 2, implying that unobserved factors would have to increase the odds ratio of Cancer incidence by more than 100% percent before the estimates become questionable and dilute the causal effect under PSM. Hence, based on the critical values of $\Gamma$, we can claim that the matching is robust to unobservable selection except for inpatient visits for non-major ailments.

Table 5 shows the economic burden of Cancer in terms of healthcare expenditure and the corresponding financial hardships CA households bear compared to matched NCA households. The per-member medical expenditure for inpatient and outpatient care is significantly greater for CA households (Rs 27,080 and Rs 889.96) compared to matched NCA households (Rs 6275 and Rs 303). The per-member OPD expenditure by CA households is significantly higher, thus indicating substantial uninsured OOP costs borne by CA households in the course of treatment. Although the per-member healthcare expenditure in public inpatients care is low

| Variable | CA scores | Matched NCA scores | Unmatched NCA scores | % bias (matched NCA) | %reduction \|bias\| | t-value for CA and matched NCA |
|---|---|---|---|---|---|---|
| Demographics | | | | | | |
| Hospitalised | 0.7904 | 0.8060 | 0.8683 | -4.2 | 79.9 | − 0.96 |
| Prop_Age15_39 | 0.3832 | 0.3793 | 0.4138 | 1.8 | 87.3 | 0.45 |
| Prop_Age40_59 | 0.2792 | 0.2775 | 0.2496 | 0.7 | 94.3 | 0.17 |
| Prop_Age60_ | 0.1575 | 0.1523 | 0.1114 | 2.3 | 88.9 | 0.52 |
| Prop. female | 0.4903 | 0.4954 | 0.4905 | − 2.9 | − 1773.1 | − 0.73 |
| Household size | 4.84 | 4.82 | 4.91 | 0.7 | 79.3 | 0.17 |
| Socioeconomic | | | | | | |
| Primary education | 0.3873 | 0.4013 | 0.3827 | − 2.9 | − 176.6 | − 0.71 |
| Secondary education | 0.2327 | 0.2245 | 0.2386 | 1.9 | − 35.2 | 0.48 |
| Higher education | 0.1637 | 0.1505 | 0.1360 | 3.7 | 52.1 | 0.89 |
| Cooking gas/electric | 0.6406 | 0.6373 | 0.6370 | 0.7 | 14.9 | 0.17 |
| Piped water | 0.3841 | 0.3906 | 0.3946 | − 1.3 | 39.7 | − 0.33 |
| Septic tank | 0.5000 | 0.5263 | 0.4986 | − 5.3 | − 1371.9 | − 1.3 |
| Garbage | 0.3849 | 0.3799 | 0.3730 | 1 | 57.1 | 0.25 |
| OBC | 0.3660 | 0.3668 | 0.3981 | − 0.2 | 97.5 | − 0.04 |
| Unreserved | 0.3956 | 0.3799 | 0.3204 | 3.3 | 79.3 | 0.79 |
| Hindu | 0.7673 | 0.7788 | 0.7544 | − 2.7 | 11.2 | − 0.68 |
| Self-employed | 0.3742 | 0.3775 | 0.4287 | − 0.7 | 94 | − 0.17 |
| Salaried | 0.1538 | 0.1521 | 0.1699 | 0.4 | 89.9 | 0.11 |
| HE support | 0.2591 | 0.2558 | 0.2355 | 0.8 | 86.4 | 0.19 |
| Region | | | | | | |
| Urban | 0.4605 | 0.4597 | 0.4412 | 0.2 | 95.7 | 0.04 |
| Central-region | 0.0452 | 0.0461 | 0.0683 | − 0.4 | 96.5 | − 0.1 |
| West-region | 0.1743 | 0.0461 | 0.1692 | − 3.5 | − 166.6 | − 0.84 |
| East-region | 0.2245 | 0.2237 | 0.1764 | 0.2 | 98.3 | 0.05 |
| South-region | 0.1941 | 0.1801 | 0.2336 | 3.4 | 64.8 | 0.88 |
| North east-region | 0.0806 | 0.0798 | 0.1127 | 0.3 | 97.4 | 0.07 |
| Observations | 1217 | 1220 | 66,892 | | | |

**Table 4**. **List of variables used for balancing the matched NCA households**. Authors' own calculation. Source: Social Consumption: Health Survey-NSS 75th Round (2017-18); Using the same covariates, the propensity score is estimated with the inclusion of hospitalisation, without hospitalisation, dropping reported death of any household members, and dropping top 1% expenditure on healthcare, respectively. Further, for spatial heterogeneity, we have controlled for 88 NSS sub-regions instead of the state. The SE is calculated using the bootstrapping technique. The t-value provides a statistical test of the mean significance of variables between CA and matched NCA households. *CA* cancer-affected household, *NCA* non-cancer-affected households.

compared to private, CA households (Rs 27,276 and Rs 13,575) still have to bear significantly higher medical expenditures compared to NCA households (Rs 8528 and Rs 2086). Overall, the total inpatient expenditure per member for CA households (Rs 25,627) is significantly higher than the matched per-member NCA households (Rs 6113). The critical value of $\Gamma$ indicates that these outcomes are robust and insensitive to hidden bias. However, the exact spill-over effect can't be captured due to the limitations of the database. The per-member CA households bear relatively higher healthcare expenditure for both inpatient and outpatient visits for the treatment of non-major diseases compared to NCA households, but this is insignificant. The corresponding $\Gamma$ is found to be sensitive to unobservable selection when odds increase beyond 30%.

The healthcare utilization can leave CA households reeling in chronic debt if they are not covered with formal or informal insurance. We therefore examine the magnitude of distress an average CA household undergoes compared to matched NCA households. We observed that 62% of the CA households borrow or sell assets while only 28% of the matched NCA households bear the debt burden while seeking inpatient treatment. There is not much difference in terms of debt financing of the CA (60% in public and 55% in private) and matched NCA (34% in public and 21% in private) households while visiting public or private healthcare respectively, nevertheless, it is much higher compared to matched NCA households (~ 30%). Likewise, loss of income for CA households is significantly higher while seeking treatment in inpatient care. The impact is also reflected in the workforce participation rate of the CA households. The workforce participation rate of non-cancer members of CA households is lesser as compared to matched NCA households. Only 27% of non-cancer members of CA households are involved in the workforce, while 33% of matched NCA households are in the workforce. The non-medical consumption expenditure during the 15-day recall period is also found to be significantly less for CA households (Rs 1565) as compared to matched NCA households (Rs 2263), indicating possible redistribution

| Outcome variables | CA | Matched NCA | Unmatched NCA | t-value for CA and matched NCA | Critical level of (Γ) |
|---|---|---|---|---|---|
| Inpatient visit p.m. | 0.2168 | 0.2291 | 0.2318 | − **2.44** | **1.5–1.6** |
| Inpatient visit p.m. (public) | 0.3368 | 0.2889 | 0.2830 | **5.08** | **1.3–1.4** |
| Inpatient visit p.m. (private) | 0.3366 | 0.2885 | 0.2844 | **5.46** | **1.3–1.4** |
| Length of inpatient stay p.m. | 4.9150 | 1.4894 | 1.4631 | **16.41** | **2>** |
| Surgeries p.m. | 0.1688 | 0.0710 | 0.0682 | **15.98** | **2>** |
| Medicine uptake p.m. | 0.5028 | 0.3870 | 0.3706 | **10.38** | **1.4–1.5** |
| Screening tests p.m. | 0.3573 | 0.2195 | 0.2165 | **17.88** | **1.6–1.7** |
| Other diagnostic tests p.m. | 0.3764 | 0.2516 | 0.2475 | **14.90** | **1.6–1.7** |
| Non-cancer Inpatient visit p.m. | 0.0343 | 0.2384 | 0.2376 | − **63.41** | **2>** |
| Non-major inpatient visit p.m. | 0.0206 | 0.1664 | 0.1685 | − **57.97** | **1.1–1.2** |
| Outpatient visit p.m. | 0.1708 | 0.1654 | 0.1480 | 0.69 | 2> |
| Outpatient visit (public) p.m. | 0.3202 | 0.3377 | 0.3273 | − 1.15 | 2> |
| Outpatient visit (private) p.m. | 0.3407 | 0.3519 | 0.3431 | − 0.74 | 2> |
| Non-cancer outpatient visits p.m. | 0.0955 | 0.1654 | 0.1480 | − **11.19** | **2>** |
| Non-major outpatient visits p.m. | 0.0463 | 0.0940 | 0.0865 | − **12.39** | **2>** |

**Table 5**. **The impact of cancer on households' healthcare utilisation.** Authors' calculation. Source: Social Consumption: Health Survey-NSS 75th Round (2017-2018). Treatment Observations = 1217 (varies depending on the healthcare utilisation); The ATT is based on propensity scores derived without hospitalisation; We have used the nearest neighbourhood matching method to estimate the propensity score; *CA* Cancer-affected household; *NCA* Non-cancer-affected household; The SE is calculated using a bootstrapping technique. The t-value provides a statistical test of the mean significance of health outcome variables between CA and matched NCA households. .

| Outcome variables | CA | Matched NCA | Unmatched NCA | t-value for CA and matched NCA | Critical level of gamma (Γ) |
|---|---|---|---|---|---|
| Inpatient medical exp. p.m. | 27080.44 | 6275.43 | 5782.13 | **10.95** | **2>** |
| Inpatient medical exp. (private) p.m. | 27276.16 | 8528.96 | 8109.18 | **7.58** | **2>** |
| Inpatient medical Exp. (public) p.m. | 13575.06 | 2086.03 | 1939.31 | **8.47** | **2>** |
| Inpatient transport exp. p.m. | 853.70 | 178.06 | 138.71 | **9.34** | **2>** |
| Total inpatient exp. p.m. | 25627.59 | 6113.01 | 5706.28 | **11.31** | **2>** |
| Non-cancer inpatient exp. p.m. | 9767.31 | 6170.66 | 5706.28 | 1.68 | 1.2–1.3 |
| Non-major disease inpatient exp. p.m. | 8847.83 | 4673.58 | 4281.88 | 1.40 | 1.2–1.3 |
| Outpatient medical exp. p.m. | 808.06 | 270.37 | 257.49 | **4.68** | **2>** |
| Outpatient transport exp. p.m. | 62.81 | 22.86 | 22.14 | **4.44** | **1.5–1.6** |
| Total outpatient exp. p.m. | 889.96 | 302.98 | 290.30 | **4.88** | **2>** |
| Non-cancer outpatient exp. p.m. | 387.65 | 304.79 | 290.30 | 1.01 | 1.1–1.2 |
| Non-major disease outpatient exp. p.m. | 394.47 | 237.17 | 232.98 | 1.21 | 1.1–1.2 |
| Coping and financial hardship | | | | | |
| Borrow/sale of assets (inpatient) | 0.6227 | 0.2898 | 0.2924 | **13.75** | **2>** |
| Borrow/sale of assets (inpatient public) | 0.6032 | 0.3465 | 0.3551 | **8.86** | **2>** |
| Borrow/sale of assets (inpatient private) | 0.5553 | 0.2140 | 0.2120 | **10.30** | **2>** |
| Borrow/sale of assets (outpatient) | 0.0544 | 0.0243 | 0.0239 | **2.97** | **2>** |
| Income loss (inpatient) p.m. | 1554.53 | 531.00 | 525.73 | **8.37** | **2>** |
| Income loss (outpatient) p.m. | 104.01 | 46.15 | 46.89 | **2.43** | **1.2–1.3** |
| Members currently working (%) | 0.2957 | 0.3037 | 0.3050 | − 1.47 | 2> |
| Non-cancer members currently working (%) p.m. | 0.2733 | 0.3309 | 0.3280 | − **10.65** | **2>** |
| Non-medical consumption expenditure p.m. | 1565.67 | 2263.29 | 2217.63 | − **2.89** | **2>** |

**Table 6**. **The impact of cancer on households' medical expenditure and coping strategy.** Authors' calculation. Source: Social Consumption: Health Survey- NSS 75th Round (2017-18); The expenditure figures (Col 1, 2 & 3) are presented in Indian Rupee (INR (Rs.)). The ATT is based on propensity scores derived without hospitalisation; We have used the nearest neighbourhood matching method to estimate the propensity score; CA: Cancer-affected household; NCA: Non-cancer-affected household; The SE is calculated using a bootstrapping technique. The t-value provides a statistical test of the mean significance of health outcome variables between CA and matched NCA households.

of subsistence resources towards cancer healthcare. The critical level of $\Gamma$ suggests that most of the estimates of financial hardship are robust to unobservable selections except for income loss due to OPD visits.

*Robustness check*

Tables A5 and A6 (in the Appendix) show the impact of Cancer on healthcare utilisation, healthcare expenditure, and coping strategy of CA households compared to matched NCA households using four different matching methods (i.e., Nearest neighbourhood matching;, Radius matching, Kernel matching, Stratified matching) and also after adjusting the propensity score (with inpatient hospitalisation, without hospitalisation, excluding reported death, and excluding outlier health expenditure (truncated) see details in Sect. 3.2.2). The estimated results (see Table A5 and A6) broadly reflect a similar pattern of the baseline results (Tables 5 and 6). Here, we see that per-member CA households have comparatively more inpatient visits (by 0.033) (See Table A5 column 6; excluding death). The intensity of medical intervention in terms of per-member length of hospital stay (between 2.7 and 3.7 days more), surgeries (between 0.09 and 0.10), medicine uptake between 0.08 and 0.14 p.m.), screening (between 0.11 and 0.14 p.m.) and other diagnostic test (between 0.1–0.13p.m.) are significantly higher for CA households compared to matched NCA households. The aggressive healthcare utilisation by the CA households indicates the severity of the disease and the resources it demands. We observed consistently fewer inpatient and outpatient visits by members having non-cancer or non-major diseases from CA households compared to matched NCA households. It is likely because of the severe resource constraint non-cancer members in a CA households are forced to postpone their treatment-seeking till it reaches a desperation.

The per-member total inpatient expenditure incurred by CA households is between Rs. 16,000 and Rs. 19,000 on average in excess compared to matched NCA households. The affected households have to incur more inpatient expenditure in private hospitals than the public, but in both setups CA households have to spend significantly more than the matched NCA households. Similarly, CA households spend between Rs. 540 and Rs. 600 per member in outpatient visits more than NCA households. Although per-member CA households spend more on non-cancer or non-major diseases while visiting both inpatient and outpatient care, it is statistically insignificant. As a result of the excessive medical expenditure in inpatient care, (between 24% and 33%) more CA households fall for distress financing compared to matched NCA households. The severity of distress financing is more observable when they seek treatment to private facilities (more than 33%). Moreover, we also find that while visiting inpatient care, CA households incur a loss of income of more than Rs 1000 per member compared to NCA households. On the employment front, the per-member CA households experience a (5–8) percent less workforce participation rate among non-cancer members as compared to the matched NCA households, however, results are not statistically significant. The per-member non-medical consumption expenditure on average is significantly less by Rs 757 – Rs 989 compared to matched NCA households, thus indicating potential subsistence resource redistribution towards healthcare.

*Healthcare utilisation based on socioeconomic status of CA households*

Table 7 shows the result of the subgroup analysis separately for the possible heterogeneous effect of Cancer on disadvantaged households with low education and backward social categories. Columns 5 and 9 show the results from the interaction of the Cancer dummy with education (Household head with primary or below primary education level takes the value 1 and 0 otherwise) and caste (Household in reserved social group (SC/ST) takes the value 1 and 0 otherwise) dummy, respectively. We did not find any statistically significant difference in the utilisation of health care services between the two education groups (except for private outpatient visits, which is borderline significant). However, there is a significant difference between the SC/ST and non-SC/ST households, the former having lower utilisation of some of the healthcare services as compared to the latter (e.g., overall inpatient visit (− 0.04*)) and inpatient visits in public hospital (− 0.05*). The SC/ST groups also have a lesser number of medicine uptake and other diagnostic tests, indicating financial constraints experienced by these disadvantaged groups even on vital healthcare services. Apart from poor affordability, lower uptake of healthcare of the households with lower education and under-privileged social categories also reflects the absence of social security, higher opportunity costs of treatment-seeking and poor understanding of diseases as compared to more educated and economically well-off households from upper social category[39,40]. The financial constraints imposed on cancer-affected households (belonging to disadvantaged categories) due to the high treatment cost of cancer, often leads to financial toxicity and impoverishment, which may restrict their healthcare-seeking behaviour for NCA members[41,42].

Table 8 shows the corresponding difference in healthcare expenditure and financial hardship for the two groups. The results show that the per-member inpatient medical, transport and total expenditure are significantly lower than NCA households with education below primary or belonging to ST/SCs. The higher out-of-pocket expenditure (OOPE) by the advantaged socio-economic groups is due to their location in a higher socio-economic gradient, enabling them to have higher education, higher income, higher social security, and higher awareness[43]. Moreover, we also find that both the disadvantaged groups have to bear significantly more distress financing in terms of borrowing and selling assets for inpatient care, more so for CA households with below primary education. The cumulative levels of deprivation and discrimination of the socially backward categories manifest into low uptake of human capital (like health and education) culminating into a vicious circle of high disease burden, low years of education, low labour market outcome and impoverishment). However, we acknowledge the results are grossly indicative and need in-depth focus for a better understanding of the problem, which is beyond the scope of the paper.

## Conclusion

We have attempted to comprehensively study the impact of Cancer on households' healthcare utilisation and economic burden (in terms of healthcare expenditure, financial hardships, distress financing and coping

| Outcome variables | Based on the level of education | | | | Based on social group | | | |
|---|---|---|---|---|---|---|---|---|
| | Const. | CA | Education ≤ primary | CA*education ≤ primary | Const. | CA | ST/SC | CA*ST/SC |
| Inpatient visit p.m. | 0.3011*** (0.0090) | 0.0619*** (0.0145) | − 0.0020 (0.0123) | − 0.0065 (0.0191) | 0.3007*** (0.0074) | 0.0727*** (0.0113) | 0.0011 (0.0151) | **− 0.0444* (0.0219)** |
| Inpatient visit p.m. (public) | 0.2930*** (0.0154) | 0.0257 (0.0211) | 0.0050 (0.0184) | 0.0198 (0.0264) | 0.2940*** (0.0109) | 0.0575*** (0.0159) | 0.0137 (0.0202) | **− 0.0546* (0.0291)** |
| Inpatient visit p.m. (private) | 0.2925*** (0.0102) | 0.0678*** (0.0177) | − 0.0219 (0.0237) | 0.2925 (0.0102) | 0.2878*** (0.0093) | 0.0587*** (0.0139) | − 0.0169 (0.0188) | − 0.0271 (0.0270) |
| Length of inpatient stay p.m. | 1.67*** (0.1087) | 3.12*** (0.3600) | 0.4730 (0.3592) | − 0.0091 (0.5708) | 1.89*** (0.1192) | 2.88*** (0.2554) | − 0.0195 (0.1777) | 0.7998 (0.5755) |
| Surgeries p.m. | 0.0865*** (0.0088) | 0.0923*** (0.0133) | − 0.0174 (0.0107) | − 0.0063 (0.0163) | 0.0830*** (0.0056) | 0.0900*** (0.0090) | − 0.0280*** (0.0101) | − 0.0081 (0.01620) |
| Medicine uptake p.m. | 0.2989*** (0.0091) | 0.0601*** (0.0146) | − 0.0009 (0.0123) | − 0.0058 (0.0191) | 0.2996*** (0.0074) | 0.0713*** (0.0113) | 0.0012 (0.0152) | **− 0.0457** (0.0219)** |
| Screening tests p.m. | 0.2794*** (0.0097) | 0.0721*** (0.0150) | − 0.0180 (0.0129) | 0.0060 (0.0196) | 0.2622*** (0.0074) | 0.0958*** (0.0114) | − 0.0165 (0.0159) | − 0.0224 (0.0227) |
| Other diagnostic tests p.m. | 0.2848*** (0.0095) | 0.0666*** (0.0150) | − 0.0049 (0.0128) | − 0.0081 (0.0196) | 0.2714*** (0.0075) | 0.0856*** (0.0114) | 0.0025 (0.0159) | **− 0.0409* (0.0227)** |
| Non-cancer Inpatient visit p.m. | 0.2706*** (0.0074) | − 0.0326** (0.0156) | − 0.0204** (0.0096) | − 0.0113 (0.0197) | 0.2683*** (0.0065) | − 0.0507*** (0.0116) | 0.0072 (0.0132) | − 0.0149 (0.0232) |
| Disease inpatient visit p.m. | 0.2843*** (0.0072) | 0.0203 (0.0296) | 0.0075 (0.0096) | − 0.0186 (0.0335) | 0.2811*** (0.0042) | 0.0204 (0.0175) | − 0.0054 (0.0089) | − 0.0183 (0.0244) |
| Outpatient visit p.m. | 0.4215*** (0.0247) | − 0.0478 (0.0312) | − 0.0732** (0.0290) | 0.0345 (0.0372) | 0.4116*** (0.0214) | − 0.0473* (0.0251) | − 0.0651** (0.0313) | 0.0006 (0.0388) |
| Outpatient visit (public) p.m. | 0.3485*** (0.0301) | − 0.0121 (0.0406) | − 0.0069 (0.0374) | − 0.0257 (0.0490) | 0.3898*** (0.320) | − 0.0648* (0.0362) | − 0.0571 (0.0422) | 0.011 (0.0525) |
| Outpatient visit (private) p.m. | 0.4333*** (0.0301) | − 0.0722* (0.0378) | − 0.1312*** (0.0334) | **0.0752* (0.0436)** | 0.3868*** (0.0218) | − 0.0427 (0.0270) | − 0.0584 (0.0393) | − 0.0210 (0.0475) |
| Non-cancer outpatient visits p.m. | 0.4215*** (0.0247) | − 0.0504 (0.0344) | − 0.0732** (0.0290) | 0.0275 (0.0420) | 0.4116*** (0.0214) | − 0.00565** (0.0275) | − 0.0651** (0.0313) | 0.0019 (0.0455) |
| Non-major outpatient visits p.m. | 0.3433*** (0.0188) | − 0.0451 (0.0290) | − 0.0456** (0.0231) | 0.0094 (0.0361) | 0.3086*** (0.0161) | − 0.0264 (0.0161) | 0.0161 (0.0318) | − 0.0353 (0.0453) |

**Table 7**. **Difference in healthcare utilisation between CA and matched NCA households by socioeconomic status**. Authors' own calculation. Source: Social Consumption: Health Survey- NSS 75th Round (2017–2018). ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; Treated Observations = 1,217 (varies depending on the healthcare utilisation); *CA* Cancer affected household; Robust standard error is in parenthesis; Education ≤ Primary: A household head with below primary and primary level of education; *ST/SC* A household head under Schedule Tribe and Schedule Caste social category; We have control spatial heterogeneity for 88 NSS sub-region instead of states. We have used nearest-neighbourhood matching to estimate the propensity score and identify the matched NCA household. The methodology to obtain these estimates is discussed in the Appendix; Robust standard error is in parenthesis.

strategies) relative to NCA households. The CA households have significantly higher institutional inpatient and outpatient visits, reported longer hospital stays, more surgery, medicine uptake, and diagnostic tests as compared to NCA households. The spill over effect is reflected in significantly lower inpatient and outpatient visits for non-cancer and non-major ailments among the CA households. The corresponding healthcare expenditure is also relatively high for CA households in both inpatient and outpatient visits. Whether it is private or public, CA households bear a significantly higher treatment costs on average. As a result, the CA households experience significantly higher incidence of borrowing and selling of assets, greater loss of income while visiting inpatient care, lower workforce participation for non-cancer members, and lower non-medical consumption expenditure compared to NCA households. Given the low social security coverage, we also find heterogeneous effect of cancer across vulnerable socioeconomic groups. The CA households from ST/SC backgrounds are found to have significantly (weak) fewer inpatient visits, medicine uptake, and diagnostic tests, and spend significantly less for inpatients compared to non-ST/SC households. They are more likely to end up with borrowing and distressed selling of assets. Hence, the study suggests that interventions should be tailored according to disease hazards that address their access to timely and quality care. The coverage over inpatient care without addressing the excess health expenditure incurred in OPD (at diagnosis and follow-up) would dilute the policy effort.

A key strength of our study is the use of a nationally representative household survey, which includes information on inpatient and outpatient healthcare visits. The estimation of the average treatment effects on the treated (ATT) through matching CA with NCA households. addresses the confounding effect arising due to the non-random assignment of the treatment (here, Cancer). Further, we have done heterogeneity analysis to estimate a series of sub-sample effects for the robustness check of our results. Moreover, we have used four different matching methods to support out baseline results.

Given the limited health security coverage, the pronounced heterogeneity in the economic burden of cancer across vulnerable socioeconomic groups, well targeted disease-specific interventions, are critical particularly for diseases (like cancer), which involve huge economic burden. We emphasize the policy frameworks that

| Outcome variables | Based on the level of education | | | | Based on social group | | | |
|---|---|---|---|---|---|---|---|---|
| | Const. | CA | Education ≤ primary | CA*Education ≤ primary | Const. | CA | ST/SC | CA*ST/SC |
| Inpatient medical exp. p.m. | 8348.25*** (1027.39) | 29653.12*** (3648.08) | − 3454.56*** (1131.30) | **− 14663.22*** (4314.98)** | 6839.12*** (591.15) | 22574.84*** (2469.33) | − 3557.351*** (751.59) | **− 5883.88*** (3909.12)** |
| Inpatient medical exp. (private) p.m. | 8889.55*** (933.40) | 23682.92*** (3473.53) | − 1997.39* (1160.45) | − 6576.02 (5170.55) | 8968.576*** (791.87) | 21032.21*** (3285.13) | − 1918.78 (1358.16) | **− 10893.58*** (791.87)** |
| Inpatient medical exp. (public) p.m. | 1923.06*** (341.88) | 18397.75*** (3487.91) | 259.45 (415.49) | **− 10394.38*** (3648.76)** | 2130.86*** (241.31) | 12855.48*** (1912.01) | − 540.23 (392.78) | − 3726.01 (2269.37) |
| Inpatient transport exp. p.m. | 226.34*** (21.79) | 816.25*** (139.90) | 11.47 (28.21) | **− 300.06* (161.36)** | 255.23*** (21.09) | 657.21*** (93.47) | − 54.84* (30.41) | − 47.38 (141.17) |
| Total inpatient exp. p.m. | 7386.22*** (739.75) | 24815.15*** (2885.63) | − 2166.53** (890.36) | **− 8486.02** (3729.33)** | 6734.66*** (553.65) | 21788.18*** (2357.91) | − 3043.39*** (734.67) | **− 7953.871*** (2906.64)** |
| Non-cancer inpatient exp. p.m. | 7386.22*** (740.41) | 4308.39 (5787.06) | − 2166.53** (891.16) | − 866.90 (6011.45) | 6734.66*** (554.14) | 3734.95 (2832.90) | − 3043.39*** (2832.90) | 3734.952 (2832.90) |
| Non-major disease inpatient exp. p.m. | 5855.61*** (657.45) | 8015.13 (8688.92) | − 2129.35*** (743.42) | − 5848.40 (8793.78) | 5010.41*** (501.68) | 5440.96 (4283.57) | − 2639.97*** (571.11) | − 4215.53 (4371.57) |
| Outpatient medical exp. p.m. | 313.69*** (44.63) | 623.81** (199.27) | − 93.18 (48.20) | − 228.93 (221.26) | 301.38*** (42.27) | 522.61*** (129.10) | − 127.73** (49.81) | − 231.81 (154.82) |
| Outpatient transport exp. p.m. | 25.83 (7.00) | 53.18 (21.39) | − 3.43 (7.66) | − 28.72 (22.34) | 24.75*** (3.10) | 41.45*** (11.90) | − 7.32 (4.03) | − 19.05 (15.16) |
| Total outpatient exp. p.m. | 353.27*** (49.06) | 673.16*** (209.71) | − 98.87* (53.21) | − 239.87 (233.36) | 302.43*** (26.36) | 603.75*** (131.29) | − 96.99** (39.20) | **− 275.87* (160.57)** |
| Non-cancer outpatient exp. p.m. | 353.27*** (49.089) | 46.29 (91.87) | − 98.87* (53.24) | − 46.63 (106.56) | 302.43*** (26.38) | 22.23 (50.89) | − 96.987** (39.23) | 47.49 (143.41) |
| Non-major disease outpatient exp. p.m. | 342.14*** (73.90) | 14.87 (141.94) | − 117.41 (77.22) | − 3.87 (161.42) | 213.82*** (17.75) | 64.798 (68.54) | − 25.18 (38.82) | 53.26 (184.35) |
| Coping and financial hardship | | | | | | | | |
| Borrow/sale of assets (inpatient) | 0.1884*** (0.0247) | 0.2513*** (0.0415) | 0.1846*** (0.0367) | **0.1255** (0.0598)** | 0.2839*** (0.0207) | 0.3171*** (0.0353) | 0.0272 (0.0413) | **0.0893*** (0.0705)** |
| Borrow/sale of assets (inpatient public) | 0.1721*** (0.0361) | 0.2717*** (0.0597) | 0.1181** (0.0478) | 0.0582 (0.0804) | 0.2012*** (0.0241) | 0.3381*** (0.0472) | 0.0828 (0.0470) | − 0.0068 (0.0877) |
| Borrow/sale of assets (inpatient private) | 0.1888*** (0.0290) | 0.2068*** (0.0474) | 0.2475*** (0.0479) | 0.1303 (0.0739) | 0.3358*** (0.0293) | 0.2424*** (0.0447) | 0.0047 (0.0692) | 0.1403 (0.0978) |
| Borrow/sale of assets (outpatient) | 0.0234 (0.0155) | 0.0116 (0.0224) | − 0.0092 (0.0172) | 0.0427 (0.0272) | 0.0219** (0.0087) | 0.0186 (0.0139) | − 0.0135 (0.0121) | **0.079** (0.0314)** |
| Income loss (inpatient) p.m. | 529.99*** (74.83) | 1324.32*** (291.46) | 96.45 (111.76) | **− 584.65* (318.30)** | 435.99*** (42.83) | 1198.32*** (166.60) | 54.60 (83.68) | − 296.44 (231.25) |
| Income loss (outpatient) p.m. | 33.799*** (9.36) | 25.77 (17.26) | 36.26* (19.11) | 42.25 (46.52) | 35.69*** (9.53) | 50.55*** (16.61) | 53.88 (33.89) | 41.18 (105.15) |
| Members currently working (%) | 0.2839*** (0.0091) | 0.0018 (0.0160) | − 0.0014 (0.0124) | 0.0209 (0.0116) | 0.2764*** (0.0066) | 0.0079 (0.0091) | 0.0546*** (0.0141) | − 0.0013 (0.0192) |
| Non-cancer members currently working (%) p.m. | 0.3293*** (0.0085) | − 0.0597*** (0.0124) | 0.0015 (0.0110) | 0.0072 (0.0158) | 0.3102*** (0.0064) | − 0.0379*** (0.0090) | 0.0503*** (0.0137) | **− 0.0366* (0.0189)** |
| Non-medical consumption expenditure p.m. | 3207.94*** (246.95) | − 583.75 (515.72) | − 1447.71*** (258.86) | − 149.24 (561.47) | 2281.19*** (93.20) | − 574.79* (295.02) | − 521.69** (178.09) | 219.04 (393.163) |

**Table 8**. **Difference in out-of-pocket healthcare expenditure between CA and matched NCA households by socioeconomic status.** Authors' own calculation. Source: Social Consumption: Health Survey- NSS 75th Round (2017-2018); The estimates of expenditures are expressed in Indian Ruppee (INR, Rs.) ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; Treated Observations = 1217 (varies depending on the healthcare utilisation); Robust standard error is in parenthesis.

differentiate between healthcare utilization patterns driven by the nature of diseases. Thus policies should consider the unique cost dynamics of cancer, which involve both high hospitalization costs and long-term treatment, compared to the more short-term frequent yet lower-cost OPD-centric nature of non-cancer diseases.

The current emphasis of health insurance policies on inpatient care neglects significant out-of-pocket expenditures associated with outpatient services, including diagnostics and follow-up care, thereby undermining the effectiveness of these policy efforts. A more comprehensive framework, as shown by state-mandated insurance coverage for mammograms and the removal of cost-sharing for preventive services. It highlights the potential benefits of incorporating preventive measures into insurance plans[44], Such an approach could alleviate the economic burden by promoting early detection at OPD, reducing disease incidence, and minimizing subsequent hospitalizations and associated high costs.

There are some limitations to our study. The CA households are identified based on their self-reported Cancer morbidity. Therefore, due to the respondents' lack of knowledge about earlier symptoms and delay in treatment-seeking due to financial constraints, the disease might have remained undiagnosed for some respondents during the survey period. So, estimates are likely to be downward biased. Moreover, we acknowledge the possibility of latent chronic diseases among the members in those households not reporting any ailment; however, we

expect this limitation applies similarly to cancer as well. Like many chronic diseases cancer often manifests its symptoms after a prolonged gestation period. Thus, the exclusion criteria are consistent with the study's objective of accurately assessing the differential economic burden of cancer relative to other ailments. NSS does not record the cause of death of any member in the household. Therefore, the potential undercounting of CA households could cause a downward bias in the estimation of economic burden as there may be a drastic increase in healthcare utilisation and out-of-pocket expenditure incurred immediately before death. Such potential CA households in our analysis either fall into the control group or are dropped from the study. However, we attempt to address these potential biases by analysing the economic burden of cancer by dropping households reporting for death and including hospitalisation as an additional indicator in the estimation of propensity score matching. The survey does not provide information on Cancer risk factors, and therefore, the matching method does not account for and control for such unobservable factors. Additionally, we do not have data on the occupational history of the household members, which can reveal the intensity of carcinogens the households are exposed to.

The analysis presented in the current paper highlights the institutional differences in healthcare delivery and their implications for health expenditure. Specifically, our findings show that cancer-affected (CA) households incur significantly higher medical expenditures in both public and private inpatient care compared to non-cancer-affected (NCA) households, with private care imposing a substantially greater financial burden. Additionally, the severity of distress financing remains high for CA households across both institutional types, further underscoring the economic strain associated with catastrophic illnesses like cancer. While these findings underscore the critical role of institutional variables, analysing the economic burden of healthcare across institutional differences in a spatial context (rural versus urban) would require a distinct research framework. Such an analysis involves additional dimensions, including the spatial variation in healthcare infrastructure, the accessibility of specialized services, and socioeconomic disparities, which are beyond the scope of the current paper.

## Data availability

Analysis is solely based on the secondary database on India. These data sets are available in the public domain of the Government of India, Ministry of Statistics and Program Implementation (MOSPI).: Unit Level Data & Report on NSS 75th Round for Schedule- 25.0, July 2017 -June 2018, (Social Consumption: Health) and can be accessed freely through registration. The link is as follows : Link: https://microdata.gov.in/nada43/index.php/catalog/152; https://microdata.gov.in/nada43/index.php/auth/login/?destination=catalog/152/get_microdata.

Received: 11 September 2024; Accepted: 5 May 2025
Published online: 14 May 2025

## References

1. Tan, B. H., Birdsell, L. A., Martin, L., Baracos, V. E. & Fearon, K. C. Sarcopenia in an overweight or obese patient is an adverse prognostic factor in pancreatic cancer. *Clin. Cancer Res.* **15** (22), 6973–6979. https://doi.org/10.1158/1078-0432.CCR-09-1525 (2009).
2. Rahib, L. et al. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the united States. *Cancer Res.* **74** (11), 2913–2921. https://doi.org/10.1158/0008-5472.CAN-14-0155 (2014).
3. World Health Organization. *Non-Communicable Diseases Country Profiles* (2018).
4. Vineis, P. & Wild, C. P. Global cancer patterns: causes and prevention. *Lancet* **383** (9916), 549–557. https://doi.org/10.1016/S0140-6736(13)62224-2 (2014).
5. Dhillon, P. K. et al. The burden of cancers and their variations across the States of India: the global burden of disease study 1990–2016. *Lancet Oncol.* **19** (10), 1289–1306 (2018).
6. Watts, N. et al. The 2018 report of the lancet countdown on health and climate change: shaping the health of nations for centuries to come. *Lancet* **392** (10163), 2479–2514 (2018).
7. Chandramohan, K. & Thomas, B. Cancer trends and burden in India. *Lancet Oncol.* **19** (12), e663 (2018).
8. Sathishkumar, K., Chaturvedi, M., Das, P., Stephen, S. & Mathur, P. Cancer incidence estimates for 2022 & projection for 2025: result from National Cancer registry programme, India. *Indian J. Med. Res.* **156**, 598–607 (2022).
9. WHO. *International Agency for Research on Cancer, Cancer Today, India*.
10. Indian Council for Medical Research. *National Centre for Disease Informatics and Research (ICMR-NCDIR), World Cancer Day 2024, Close the Care Gap—Addressing Cancer Care in India, 04th February*. https://ncdirindia.org/display/wcd.aspx (2024).
11. Verma, M., Sheoran, P. & Chaudhury, A. Application of nanotechnology for cancer treatment. In *Advances in Animal Biotechnology and its Applications* 161–178 (Springer, 2018).
12. Ward, E. et al. Association of insurance with cancer care utilisation and outcomes. *CA Cancer J. Clin.* **58** (1), 9–31 (2008).
13. Brown, M. L., Lipscomb, J. & Snyder, C. The burden of illness of cancer: economic cost and quality of life. *Annu. Rev. Public Health*. **22** (1), 91–113 (2001).
14. Sikora, K. & James, N. Top-up payments in cancer care. *Clin. Oncol.* **21** (1), 1–5 (2009).
15. Mehnert, A., de Boer, A. & Feuerstein, M. Employment challenges for cancer survivors. *Cancer* **119**, 2151–2159 (2013).
16. Goyanka, R., Yadav, J. & Sharma, P. Financial burden and coping strategies for cancer care in India. *Clin. Epidemiol. Global Health*. **20**, 101259 (2023).
17. Rajpal, S., Kumar, A. & Joe, W. Economic burden of cancer in India: evidence from cross-sectional nationally representative household survey, 2014. *PLoS ONE* **13** (2), e0193320 (2018).
18. Kaur, B. Disasters and exemplified vulnerabilities in a cramped public health infrastructure in India. *Int. J. Disaster Risk Manage.* **2** (1), 15–22 (2020).
19. Prinja, S. et al. Financial toxicity of cancer treatment in India: towards closing the cancer care gap. *Front. Public. Health*. **11**, 1065737 (2023).
20. Luengo-Fernandez, R., Leal, J., Gray, A. & Sullivan, R. Economic burden of cancer across the European union: a population-based cost analysis. *Lancet Oncol.* **14** (12), 1165–1174 (2013).
21. Guy, G. P. Jr et al. Economic burden of chronic conditions among survivors of cancer in the united States. *J. Clin. Oncol.* **35** (18), 2053–2061 (2017).
22. De Oliveira, C. et al. The economic burden of cancer care in Canada: a population-based cost study. *Can. Med. Association Open. Access. J.* **6** (1), E1–E10 (2018).

23. Huang, S. Y., Chen, H. M., Liao, K. H., Ko, B. S. & Hsiao, F. Y. Economic burden of cancers in Taiwan: a direct and indirect cost estimate for 2007–2017. *BMJ Open.* **10** (10), e036341 (2020).
24. Fenn, K. M. et al. Impact of financial burden of cancer on survivors' quality of life. *J. Oncol. Pract.* **10** (5), 332–338 (2014).
25. Bradley, C. J. Economic burden associated with cancer caregiving. *Semin. Oncol. Nurs.* **35**, 333–336 (2019).
26. Iragorri, N., de Oliveira, C., Fitzgerald, N. & Essue, B. The out-of-pocket cost burden of cancer care—a systematic literature review. *Curr. Oncol.* **28** (2), 1216–1248 (2021).
27. Mallath, M. K. et al. The growing burden of cancer in India: epidemiology and social context. *Lancet Oncol.* **15** (6), e205–e212 (2014).
28. Chakrabarty, J., Pai, M. S., Ranjith, V. K. & Fernandes, D. Economic burden of cancer in India. *Indian J. Public. Health Res. Dev.* **8**, 137–141 (2017).
29. Tamizhazhagan, V. et al. Social and economic burden of cancer on 2020-mini-review. *J. Biol. Med. Sci.* **1** (103), 2 (2017).
30. Maurya, P. K., Murali, S., Jayaseelan, V., Thulasingam, M. & Pandjatcharam, J. Economic burden of cancer treatment in a region in South India: a cross-sectional analytical study. *Asian Pac. J. Cancer Prev.* **22** (12), 3755 (2021).
31. Mahal, A., Karan, A., Fan, V. Y. & Engelgau, M. The economic burden of cancers on Indian households. *PLoS ONE.* **8** (8), e71853. https://doi.org/10.1371/journal.pone.0071853 (2013).
32. Karan, A., Engelgau, M. & Mahal, A. The household-level economic burden of heart disease in India. *Tropical Med. Int. Health*. **19** (5), 581–591 (2014).
33. Andersen, R. Revisiting the behavioral model and access to medical care: does it matter? *J. Health Soc. Behav.* **36** (1), 1–10. https://doi.org/10.2307/2137284 (1995).
34. Andersen, R. & Newman, J. F. Societal and individual determinants of medical care utilization in the united States. *Milbank Meml. Fund Q. Health Soc.* **1**, 95–124 (1973).
35. Becerril, J. & Abdulai, A. The impact of improved Maise varieties on poverty in Mexico: A propensity score-matching approach. *World Dev.* **38** (7), 1024–1035 (2010).
36. Dehejia, R. H. & Wahba, S. Propensity score-matching methods for non-experimental causal studies. *Rev. Econ. Stat.* **84** (1), 151–161 (2002).
37. Baser, O. Too much ado about propensity score models? Comparing methods of propensity score matching. *Value Health.* **9** (6), 377–385. https://doi.org/10.1111/j.1524-4733.2006.00130.x (2006).
38. DiPrete, T. A. & Gangl, M. Assessing bias in the Estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables Estimation with imperfect instruments. *Sociol. Methodol.* **34** (1), 271–310 (2004).
39. Fu, W. et al. Effects of cancer treatment on household impoverishment: a multicentre cross-sectional study in China. *BMJ Open.*, **11** (6), e044322 (2021).
40. Rijal, A., Adhikari, T. B., Khan, J. A. & Berg-Beckhoff, G. The economic impact of non-communicable diseases among households in South Asia and their coping strategy: A systematic review. *PLoS ONE* **13**(11), e0205745 (2018).
41. Kastor, A. & Mohanty, S. K. Disease-specific out-of-pocket and catastrophic health expenditure on hospitalization in India: do Indian households face distress health financing? *PLoS ONE* **13** (5), e0196106 (2018).
42. Mohanty, S. K., Ladusingh, L., Kastor, A., Chauhan, R. K. & Bloom, D. E. Pattern, growth and determinant of household health spending in India, 1993–2012. *J. Public Health*. **24**, 215–229 (2016).
43. Baird, K. E. The financial burden of out-of-pocket expenses in the united States and Canada: how different is the united States? *SAGE Open. Med.* **4**, 2050312115623792. https://doi.org/10.1177/2050312115623792 (2016).
44. Bitler, M. P. & Carpenter, C. S. Health insurance mandates, mammography, and breast cancer diagnoses. *Am. Economic J. Economic Policy*. **8** (3), 39–68 (2016).


## Acknowledgements

We are thankful to the editors and the two annonymous referees for their insughtful comments. The revision in the light of the comments has improved the draft substantially. However, error if there is any is exclusively our own.

## Author contributions

The first author is primarily responsible for conceptualization, data curation, formalization of methodology, preliminary analysis, drafting of the manuscript, statistical analysis, interpretation of data. The second author was involved in, write up design, formalization of methodology, interpretation of data, drafting of the manuscript, critical revision, validation, and supervision. The third author was involved in data curation, statistical analysis, interpretation of data, and critical revision. All authors reviewed the final manuscript before submission.

## Declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-01279-6.

**Correspondence** and requests for materials should be addressed to I.R.C.