HIGH SCHOOL ACADEMIC RESEARCH COMPETITION (SARC) 2026
Research Proposal Submission

Research Topic:
Beyond the Cancer Train: A Composite Geospatial Risk Index for Multi-Contaminant Exposure in Bathinda District, Punjab: A Pilot Study with Satellite Validation

Category: STEM (Category #1)

Submitted by:
Rabaani Mehra

Email:
rabaanimehra@gmail.com

Submission Date: April 30, 2026

1. Background and Motivation

I am from Amritsar. I am not from Malwa, the southern belt of Punjab where the cancer crisis I want to study is concentrated. But every Punjabi family I know has heard about the Cancer Train, and it would be hard to grow up in this state without forming a view about it. The train is real. It is the 339-up Bathinda to Bikaner passenger, and on most mornings it leaves with around 60 cancer patients on board, headed to a free hospital in Rajasthan. Bathinda district, where the train starts, reports cancer prevalence near 1,089 per million each year, compared with a national average closer to 800 (Mittal et al., 2014). A 2024 groundwater study found that 73 percent of wells in Malwa are over the WHO uranium safety limit of 30 µg/L, with one well at 299 (Coyte et al., 2024).

Composite environmental risk indices already exist. EJScreen does it. So does India's AQI. The issue with how this approach has been used in Malwa is not the idea, it is what goes into it. Most pesticide-exposure numbers in the existing literature come from self-reported district-level consumption, and self-reporting in rural India is unreliable for obvious reasons. So the input is wrong before the model even starts. What I want to do is replace that input with a satellite-based measurement of agricultural intensity from Sentinel-2 NDVI, then combine it with groundwater contamination in a Bayesian spatial model. Even at this scope, the data cleaning is the slowest part, so I am limiting the pilot to one district, Bathinda, and writing the protocol in a way that the rest of Malwa can be added later.

2. Research Question and Hypothesis

Central question: At the block level within Bathinda district, does a composite exposure index built from satellite-validated agricultural intensity and groundwater contamination explain spatial variation in cancer incidence more accurately than any single contaminant?

Hypothesis: Joint exposure to satellite-derived agricultural intensity, groundwater uranium, arsenic, and fluoride predicts block-level cancer incidence with greater statistical fit than any one of these variables alone, after controlling for age structure, rural population share, and per-capita income.

3. Literature Review and Identified Gap

I searched PubMed, Scopus, and Google Scholar for the period 2005 to 2025, using combinations of "Malwa," "Punjab," "cancer," "pesticide," "uranium," and "groundwater." 47 papers came back as relevant. 41 of them looked at one exposure pathway only. Mittal et al. (2014) is the foundational pesticide review for the region. Thakur et al. (2008) compared a high-pesticide block with a low-pesticide block and found cancer prevalence of 125 vs. 72 per 100,000. Coyte et al. (2024) measured uranium up to 299 µg/L in wells across Malwa. Sharma et al. (2017) ran uranium-specific risk assessments for Sangrur. Each of these papers is solid. The bigger problem cuts across all of them: every quantitative pesticide-exposure number in this entire literature is based on self-reported consumption or sales records, and none has ever been independently verified. So the gap I want to close is narrower than "nobody has built a composite index." It is the absence of any Malwa cancer study where the agricultural-exposure input has been checked against an independent source before being plugged into a model.

4. Methodology

Data sources. All datasets are public and Bathinda-specific where possible. Block-level pesticide consumption from the Punjab Department of Agriculture and Farmers' Welfare. Groundwater uranium, arsenic, and fluoride concentrations from Central Ground Water Board (CGWB) reports for Bathinda blocks and from peer-reviewed sampling studies. Cancer incidence at the block level from the Punjab Cancer Atlas and the Punjab State Cancer & Drug Addiction Treatment Centre at Bathinda. Demographic and socioeconomic controls from the 2011 Census and updated projections.

Satellite-based exposure measurement (the main methodological contribution). Self-reported pesticide use is hard to trust. To get an objective number instead, I will compute Sentinel-2 NDVI time-series and cropping-intensity proxies for every block in Bathinda over the 2018 to 2024 growing seasons. More cropping cycles per year is a known proxy for higher agrochemical input. The result is a satellite-derived agricultural-intensity score that nobody in Bathinda has to fill out a form for. I will then use that score as a check on every exposure model. From my literature search, no Malwa cancer study has done this yet.

Index construction. Each contamination variable is normalized to a 0 to 1 scale and combined into a Composite Environmental Risk Index (CERI). I will build two versions: one where everything contributes equally, and one where each variable is weighted by its WHO and US EPA reference doses. Building both lets me see whether the answer depends on how I picked the weights.

Spatial-statistical analysis. Plain regression won't work. Cancer rates in neighboring blocks are not statistically independent of each other, and pretending they are would distort everything downstream. So I will run the analysis in three stages. First, global Moran's I and Local Indicators of Spatial Association (LISA) to test for spatial clustering and locate where it sits. Second, a spatial-lag regression and a Bayesian hierarchical model with spatially structured random effects, comparing each single-contaminant model against the CERI. Third, model performance through DIC and out-of-sample prediction error using leave-one-block-out cross-validation.

Limitations and the ecological fallacy. A finding at the block level cannot prove individual-level causation. This is the well-known ecological fallacy in spatial epidemiology, and the output here should be read as a prioritization tool, not as a claim that any specific contaminant caused any specific cancer. Other limitations: CGWB sampling is uneven from one block to the next; the Cancer Atlas under-reports in poorer rural areas; and there is the latency problem (cancers being diagnosed today were caused by exposures from decades ago). I will handle each through sensitivity analyses, including bootstrapping over plausible under-reporting rates.

5. Expected Contributions

There are three things this work would add. The most important one is using satellite NDVI as an independent exposure measurement, in place of the self-reported pesticide numbers that every Malwa cancer study so far has relied on. The second is the dataset itself: a Bathinda block-level table that pulls satellite, groundwater, and cancer registry data into one place, with code and a public risk map released open-license. The third is a real statistical comparison. By formally testing whether a composite multi-contaminant model beats the single-pollutant ones, this would give the existing single-cause framing of Malwa's cancer crisis its first head-to-head check. The pilot also produces a written protocol that can be reused for the remaining ten Malwa districts.

6. Feasibility and Timeline

The pilot scope is what makes this doable at high-school scale. Bathinda has only 9 development blocks, which keeps manual data harmonization possible. The work is desk-based: no fieldwork, no human subjects, no ethics-board approval. Every tool I plan to use is free and open-source. The full project should take about 9 months and end with a public release of the dataset, code, and map.

References

Coyte, R. M., Singh, A., Kumar, R., Vengosh, A., & others. (2024). Assessment and mitigation of heavy toxic elements with emphasis on uranium in the Malwa region of Punjab, India. *ACS ES&T Water, 4*(11), 4861–4872. https://doi.org/10.1021/acsestwater.4c00900

Government of India, National Centre for Disease Informatics and Research. (2014). *Report of Punjab Cancer Atlas*. National Cancer Registry Programme, Indian Council of Medical Research.

Jamwal, N. (2015, January 9). On the "Cancer Train" of India's pesticides. *Al Jazeera*. https://www.aljazeera.com/features/2015/1/9/on-the-cancer-train-of-indias-pesticides

Mittal, S., Kaur, G., & Vishwakarma, G. S. (2014). Effects of environmental pesticides on the health of rural communities in the Malwa region of Punjab, India: A review. *Human and Ecological Risk Assessment, 20*(2), 366–387. https://doi.org/10.1080/10807039.2013.788972

Sharma, D. A., Rishi, M. S., & Keesari, T. (2017). Uranium anomalies in groundwater of Sangrur district of Punjab (India) for cancer risk assessment. *Environmental Earth Sciences, 76*, 678. https://doi.org/10.1007/s12665-017-7011-0

Thakur, J. S., Rao, B. T., Rajwanshi, A., Parwana, H. K., & Kumar, R. (2008). Epidemiological study of high cancer among rural agricultural community of Punjab in northern India. *International Journal of Environmental Research and Public Health, 5*(5), 399–407. https://doi.org/10.3390/ijerph5050399

World Health Organization. (2022). *Guidelines for drinking-water quality: Fourth edition incorporating the first and second addenda*. World Health Organization. https://www.who.int/publications/i/item/9789240045064