Travel burden increases the risk of advanced stage at diagnosis of Breast Cancer in Kashgar, China

Authors: Shengyuan Liang, Zhiqiang Jia, Rezia Emir, Xinyue Chen, Yishuai Peng, Xuwei Tian*, Ailing Ma*

Date of Data Compilation: July 2, 2025

Contact for Corresponding Authors: Xuwei Tian (tianxw2024@163.com), Ailing Ma (kdyymal@163.com)

This file provides a minimal, anonymized dataset for the replication of the primary statistical analyses in the manuscript titled, "Travel burden increases the risk of advanced stage at diagnosis of Breast Cancer in Kashgar, China." The data were sourced from a retrospective study cohort at the Breast Cancer Center at the First People's Hospital of Kashgar (FPHK), Xinjiang, China. To protect patient confidentiality, this dataset has been fully anonymized. All direct identifiers have been removed. Each row in this dataset represents a single, anonymized patient.

Table 1 description of each variable

Variable Name	Variable Label	Codes / Value Description
tnm_stage	Detailed TNM clinical stage (8th AJCC)	0: Stage 0 (in situ); 1: Stage IA; 2: Stage IB; 3: Stage IIA; 4: Stage IIB; 5: Stage IIIA; 6: Stage IIIB; 7: Stage IIIC; 8: Stage IV
late3	Three-category outcome for stage at diagnosis	dependent variable for the multinomial logistic regression. 1: Early-stage (TNM Stage 0, I, II) 2: Advanced-stage (TNM Stage III, IV) 3: Unconfirmed stage
late2	Binary outcome for stage at diagnosis	dependent variable for the binary logistic regression. 1: Advanced-stage (TNM Stage III, IV) 0: Non-advanced stage (Early or Unconfirmed)
distance_secondary	Driving distance to nearest secondary hospital	Yes (>20 km) 0: No (≤ 20 km)
distance_ter_80	Driving distance to nearest tertiary hospital	1: Yes (>80 km) 0: No (≤ 80 km)
agegroup	Patient's age group at diagnosis	1: ≤40 years; 2: 41-50 years; 3: 51-60 years; 4: >60 years
peasant	Occupation	1: Farmer; 0: Non-farmer
married	Marital status	1: Married; 0: Other (e.g., unmarried, divorced, widowed)
year	Year of diagnosis	ranging from 2019 to 2022.