

分类号: R
学号: 20232114052

密级: 公开
单位代码: 10759

石河子大学

硕士学位论文



新疆生产建设兵团 2008-2024 年肺结核流行特 征分析及发病趋势预测

学位申请人	同晓蕾
指导教师	王奎 副教授
申请学位类别	专业硕士
专业名称	公共卫生
研究领域	流行病学与卫生统计学
所在学院	公共卫生学院

中国·新疆·石河子

2026 年 5 月

分类号: R
学号: 20232114052

密级: 公开
单位代码: 10759

石河子大学

硕士学位论文



新疆生产建设兵团 2008-2024 年肺结核流行特征分析及发病趋势预测

学位申请人	同晓蕾
指导教师	王奎 副教授
申请学位类别	专业硕士
专业名称	公共卫生
研究领域	流行病学与卫生统计学
所在学院	公共卫生学院

中国·新疆·石河子

2026 年 5 月

**Analysis of the Epidemiological Characteristics of Tuberculosis and
Forecast of Incidence Trends in the Xinjiang Production and
Construction Corps, 2008–2024**

A Dissertation Submitted to

Shihezi University

In Partial Fulfillment of the Requirements

for the Degree of

Master of Public Health

By

Tong Xiao-lei

(Epidemiology and Health Statistics)

Dissertation Supervisor: A/Prof. Wang Kui

May, 2026

石河子大学学位论文独创性声明及使用授权声明

学位论文独创性声明

本人所提交的学位论文是在我导师的指导下进行的研究工作及取得的研究成果。据我所知，除文中已经注明引用的内容外，本论文不包含其他个人已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中作了明确的说明并表示谢意。

研究生签名：同晓雷

时间：2026年5月17日

使用授权声明

本人完全了解石河子大学有关保留、使用学位论文的规定，学校有权保留学位论文并向国家主管部门或指定机构送交论文的电子版和纸质版。有权将学位论文在学校图书馆保存并允许被查阅。有权自行或许可他人将学位论文编入有关数据库提供检索服务。有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

研究生签名：同晓雷

时间：2026年5月17日

导师签名：同晓雷

时间：2026年5月17日

摘要

目的：本研究基于新疆生产建设兵团 2008-2024 年肺结核流行现状，描述其时间、人群及空间分布特征，构建发病预测模型预测未来发病趋势，为兵团及同类边疆管理单元制定肺结核防控策略提供数据支持。

方法：收集 2008 年 1 月 1 日至 2024 年 12 月 31 日新疆生产建设兵团肺结核法定报告病例数据及同期人口学资料，采用描述性流行病学方法描述疾病流行特征。运用 Joinpoint 回归模型识别发病率变化拐点及变化阶段。构建 ARIMA、SARIMA、Holt-Winters 等传统时间序列模型，以及随机森林、XGBoost 等机器学习模型，并构建 SARIMA-RF-XGBoost-GBDT 优化加权组合模型。按照 7:3 比例划分训练集与测试集，通过 RMSE、MAE、 R^2 、MAPE 等指标评价模型拟合精度与泛化能力，结合敏感性分析验证模型稳健性。最后利用最优模型预测 2025 年兵团各师肺结核发病情况。

结果：

1. 2008-2024 年兵团累计报告肺结核病例 32120 例，年均发病率 63.22/10 万，整体呈波动下降趋势（AAPC=-32.71%），2022-2023 年出现回升（APC=97.49%），2024 年降至 26.86/10 万，疾病呈现冬春高、夏秋低的季节特征。

2. 人群分布方面， ≥ 60 岁老年组累计发病占比 42.36%，70~79 岁年龄组发病达峰值；男性发病率（81.50/10 万）高于女性（51.23/10 万）；职业分布以离退休人员（31.15%）、农民（24.57%）为主。空间分布上第八师（93.46/10 万）、第九师（91.60/10 万）、第三师（89.94/10 万）为高发师市。

3. 模型性能对比显示，机器学习模型整体优于传统时间序列模型，XGBoost 为性能最优的单一模型，测试集 $R^2=0.93$ 、MAPE=10.80%；SARIMA-RF-XGBoost-GBDT 组合模型综合性能最优，测试集 $R^2=0.93$ 、MAPE=10.79%，与 XGBoost 精度接近，但对数据扰动及异常值敏感性更低，模型总体稳健性良好。

4. 最优模型预测显示，2025 年兵团肺结核月发病率处于 1.63-2.61/10 万的低水平区间，季节波动幅度减小；兵团各师可划分为高、中、低三个发病风险区，高风险区包含七师等 4 个师，其发病高峰呈现冬春高发的差异化模式。

结论：

1. 新疆生产建设兵团肺结核发病率整体呈波动下降趋势，防控成效显著，但存在阶段性回升，疾病流行呈现明确的时间、人群及空间分布规律，老年男性、离退休人员、农民及高发师市人群为重点防控对象。

2. SARIMA-RF-XGBoost-GBDT 优化加权组合模型在肺结核发病预测中具备良好的精度与稳健性；基于模型预测结果可为不同地区差异化防控与资源调配提供相应的数据支撑。

关键词：新疆生产建设兵团；肺结核；流行特征；预测模型；机器学习

Abstract

Objective:This study is based on the current epidemiological situation of tuberculosis in the Xinjiang Production and Construction Corps from 2008 to 2024, describing its temporal, demographic, and spatial distribution characteristics. It constructs a disease prediction model to forecast future trends and provides data support for formulating tuberculosis prevention and control strategies in the Corps and similar border management units.

Methods:Data on legally reported tuberculosis cases in the Xinjiang Production and Construction Corps from January 1, 2008, to December 31,2024, along with corresponding demographic data for the same period, were collected. Descriptive epidemiological methods were used to characterize the epidemiological features of the disease. The Joinpoint regression model was employed to identify inflection points and phases of change in incidence rates. Traditional time series models, including ARIMA, SARIMA, and Holt-Winters, as well as machine learning models such as Random Forest and XGBoost, were constructed. An optimized weighted combination model comprising SARIMA-RF-XGBoost-GBDT was developed. The dataset was split into training and testing sets in a 7:3 ratio. Model fit and generalization ability were evaluated using metrics such as RMSE,MAE, R^2 , and MAPE, and model robustness was verified through sensitivity analysis. Finally, the optimal model was used to forecast tuberculosis incidence across all divisions of the Corps in 2025.

Results:

1. Between 2008 and 2024, the Corps reported 32,120 pulmonary tuberculosis cases, with an average annual incidence of 63.22/100,000. The overall incidence showed a fluctuating downward trend (AAPC=-32.71%), a significant recovery was seen in 2022-2023 (APC=97.49%),falling to 26.86/100,000 in 2024. A distinct seasonal pattern was observed, with peaks in winter and spring and troughs in summer and autumn.
2. Regarding population distribution, the cumulative incidence in the elderly group aged 60 years and over accounted for 42.36%, with the incidence peaking in the 70-79 age group; the incidence rate among men (81.50/100,000) was higher than that among women (51.23/100,000); In terms of occupational distribution, the majority were retired personnel (31.15%) and farmers (24.57%). In terms of geographical distribution, the 8th Division (93.46/100,000), the 9th Division (91.60/100,000) and the 3rd Division (89.94/100,000) are the divisions with the highest incidence rates.
3. Machine learning models outperformed traditional time series models, with XGBoost being the optimal single model (test $R^2=0.93$, MAPE=10.80%). The SARIMA-RF-XGBoost-GBDT ensemble model achieved comparable accuracy (test $R^2=0.93$, MAPE=10.79%) while exhibiting lower sensitivity to data perturbation and outliers, indicating good robustness.

4. The optimal model predicts that the monthly incidence of pulmonary tuberculosis in the Corps will stay within a low range of 1.63-2.61/100,000 in 2025, with diminished seasonal variation. Divisions can be classified into high-, medium-, and low-risk zones. The high-risk zone, comprising four divisions including the 7th Division, shows a distinct pattern of winter – spring peaks.

Conclusions:

1. The incidence of tuberculosis in the Xinjiang Production and Construction Corps shows an overall fluctuating downward trend, with significant achievements in prevention and control; however, there have been periodic resurgences. The epidemiology of the disease exhibits clear temporal, population and spatial distribution patterns, with elderly males, retired personnel, farmers and populations in high-incidence divisions and cities identified as key targets for prevention and control.

2. The SARIMA-RF-XGBoost-GBDT optimized weighted ensemble model demonstrates good accuracy and robustness in predicting pulmonary tuberculosis incidence. Based on the model's prediction results, it can provide corresponding data support for differentiated prevention and control strategies and resource allocation across different regions.

Key words: Xinjiang Production and Construction Corps; tuberculosis; epidemiological characteristics; prediction model; machine learning

目录

摘要.....	I
Abstract.....	II
目录.....	IV
缩略词表.....	VI
第 1 章 前言.....	1
1.1 研究背景.....	1
1.2 肺结核感染的流行分布现状.....	2
1.3 肺结核发病趋势预测研究现状.....	3
1.4 研究目的与意义.....	4
第 2 章 资料与方法.....	5
2.1 研究区域资料.....	5
2.2 研究对象与数据来源.....	6
2.3 纳排标准.....	6
2.4 肺结核诊断标准.....	6
2.5 统计分析方法.....	6
2.5.1 发病率.....	6
2.5.2 统计分析.....	7
2.5.3 流行病学特征分析.....	7
2.5.4 预测模型的构建与评价.....	7
2.5.5 预测模型介绍及建立.....	8
2.5.6 预测模型评价指标.....	13
2.5.7 质量控制.....	13
2.5.8 技术路线图.....	15
第 3 章 结果.....	16
3.1 新疆生产建设兵团 2008-2024 年肺结核流行特征分析.....	16
3.1.1 肺结核发病总体情况.....	16
3.1.2 时间分布.....	17
3.1.3 人群分布.....	21
3.1.4 地区分布.....	25

3.2 肺结核预测模型构建与结果	28
3.2.1 ARIMA 模型	28
3.2.2 SARIMA 模型	31
3.2.3 霍尔特温特 (Holt-Winters) 季节模型	34
3.2.4 随机森林模型	37
3.2.5 XGBoost 模型	41
3.2.6 SARIMA-RF-XGBoost-GBDT 组合模型	44
3.3 模型评价与对比	48
3.3.1 单一模型性能评价	48
3.3.2 组合模型与单一模型对比	49
3.3.3 模型敏感性分析	50
3.3.4 最优模型筛选及发病趋势预测	51
3.4 最优模型应用	51
第 4 章 讨论	54
4.1 新疆生产建设兵团肺结核流行现状	54
4.2 预测模型应用与评价	55
4.3 创新性、局限性与研究展望	57
4.3.1 研究创新性与局限性	57
4.3.2 研究启示与未来展望	57
第 5 章 结论	59
文献综述	60

缩略词表

缩略词	英文全称	中文全称
AAPC	Average Annual Percentage Change	平均年度百分比变化
ACF	Autocorrelation Function	自相关函数
ADF	Augmented Dickey-Fuller	单位根检验
AIC	Akaike Information Criterion	赤池信息准则
APC	Annual Percentage Change	年度变化百分比
ARIMA	Autoregressive Integrated Moving Average Model	自回归差分移动平均模型
BIC	Bayesian Information Criterion	贝叶斯信息准则
CI	Confidence Interval	置信区间
CNN	Convolutional Neural Network	卷积神经网络
CV	Cross Validation	交叉验证
GBDT	Gradient Boosting Decision Tree	梯度提升决策树
GRU	Gated Recurrent Unit	门控循环单元
LSTM	Long Short-Term Memory	长短期记忆网络
MAE	Mean Absolute Error	平均绝对误差
MAPE	Mean Absolute Percentage Error	平均绝对百分比误差
MASE	Mean Absolute Scaled Error	平均绝对标准化误差
MSPE	Mean Squared Percentage Error	平均平方百分比误差
PACF	Partial Autocorrelation Function	偏自相关函数
RF	Random Forest	随机森林
RMSE	Root Mean Squared Error	均方根误差
RR	Relative Risk	相对危险度
R ²	Coefficient of Determination	决定系数
SARIMA	Seasonal Autoregressive Integrated Moving Average	季节性自回归差分移动平均模型
TB	Tuberculosis	结核病
XGBoost	Xtreme Gradient Boosting	极端梯度提升

第1章 前言

1.1 研究背景

结核病（Tuberculosis, TB）是由结核分枝杆菌感染引起的慢性传染性疾病，可波及多种脏器，主要为肺结核感染^[1]。结核分枝杆菌主要通过呼吸道传染，含有结核分枝杆菌的飞沫悬浮于空气中，通过空气播散感染新的宿主^[2]。肺结核是严重危害人民健康的重大公共卫生问题^[3]。在世界范围内，结核分枝杆菌是最具威胁性的细菌感染病原体之一。尽管通常仅在患病、老年或免疫功能低下者中发展为临床疾病，但其易于传播的特性给医学界带来了严峻挑战。

根据世界卫生组织（WHO）2020年的数据显示，结核病是全球范围内致死率居高不下的传染性疾病，尤其在发展中国家，其发病率与死亡率长期处于较高水平^[4]。2025年的最新报告进一步指出，结核病仍是全球十大死因之一，也是单一传染源导致死亡的首位传染病。2024年全球新发结核病患者达1070万例，死亡病例超123万人，我国估算有69.6万例新发结核病患者，在30个结核病高负担国家中，中国估算结核病发病例数从第3位下降至第4位。按照全球疾病流行分类，中国首次进入中低流行国家行列。尽管全球结核病发病率在2021-2023年连续增长后首次出现回落，但距离《终止结核病战略》提出的全球终结结核病流行目标仍面临严峻挑战^[5,6]。

国内的肺结核流行病学调查数据显示，我国肺结核患病率较高，且分布广泛。肺结核发病情况在不同省、市、自治区之间分布差异明显。过去30年，我国在全国范围内开展肺结核结核病控制措施，实施效果显著，痰涂片阳性结核病的患病情况有明显下降^[7-9]。但是目前肺结核仍然是我国大多数地区的主要传染病，利福平耐药结核病也成为新挑战^[10]。随着今年交通便利，肺结核患者迁移以及跨区域疫情传播显著增加^[11]。因此怎样及时抑制肺结核疫情传播，如何降低肺结核疾病负担成为我国需要解决的关键问题。

根据中国疾病预防控制中心发布的2023年1-12月的全国法定传染病疫情概况显示：2023年肺结核病的报告发病613091例，报告死亡2167人；报告发病率为43.489/10万，报告死亡率为0.1537/10万^[12]。肺结核的疾病数据仍远高于WHO提出的“终止结核病战略”目标（到2035年发病率降至100/10万以下）。我国肺结核疫情呈现“三高一低”的特点，即高感染率、高患病率、高耐药率以及低诊断率，尤其是在经济欠发达地区和医疗资源薄弱地区，疫情控制难度更大。在全国乙类传染病中，肺结核的发病数和死亡数均位居第二，其中发病率仅次于病毒性肝炎，死亡率仅次于艾滋病。相关数据显示，其发病报告数总体呈逐年下降趋势。过去20年来，中国为减轻肺结核的负担做了大量的工作。卫生行政部门先后下发了《全国结核病防治规划2011-2015年》^[13]，《“十

三五”全国结核病防治规划（2016-2020年）》^[14]和《遏制结核病行动计划（2019-2022年）》^[15]等政策文件，防治工作取得显著进展。

在我国结核病发病呈明显区域差异，新疆位于我国的西北边陲，肺结核发病率一直居全国前列，约为全国平均水平的2.4倍，疫情形势依然严峻^[16]。新疆生产建设兵团（以下简称“兵团”）作为我国具有特殊管理体制的组织，其独特的属性决定了肺结核防控工作的特殊性与复杂性。兵团地理跨度广阔，覆盖新疆多个区域，以农垦产业为核心产业，人口构成具有鲜明的职业特征与地域特征，职工群众多从事农业、畜牧业等户外集体生产作业，生产生活的聚集性较强，这一特点在一定程度上增加了肺结核的传播风险。同时，兵团各师分布在不同区域，受自然环境、经济发展水平、医疗资源配置等因素影响，各师之间发展差异明显，防控能力也存在不均衡现象，给全域性防控工作带来了挑战。相关学者^[17-19]研究显示，新疆和青海地区的肺结核发病在近年来仍呈现出逐年增长的趋势。我国传染病网络直报系统显示，2004-2019年新疆累计肺结核病例668753例，年平均发病率为189.21/10万，位列全国首位^[20,21]。目前，新疆生产建设兵团的结核病长期流行趋势及未来发展尚缺乏系统性的预测研究。虽然近年该地区结核病疫情得到一定控制，但仍处于全国较高水平。随着耐药菌株的扩散及艾滋病交叉感染等因素的影响，结核病发病存在上升风险，仍需引起重视。因此，针对兵团的特殊属性，本文主要阐述了肺结核的流行特征、构建科学的发病预测模型、总结各类预测模型的应用效果，进而为兵团制定个性化、精准化的防控策略提供相应的数据及理论支持。

1.2 肺结核感染的流行分布现状

肺结核是全球性严峻的公共卫生问题，其疫情在全球及我国均存在显著区域差异。WHO《2025年全球结核病报告》显示，2024年全球新发结核病患者1070万例，发病率131/10万，死亡超123万人，发病数在连续三年（2021-2023年）增长后首次回落。其中30个高负担国家占全球发病总数的87%，8个国家贡献67%新发病例，中国占比6.5%。目前已有62个国家发病率低于10/10万，提前实现WHO2035年目标，主要集中在美洲区和欧洲区；而30个高负担国家多数发病率为150-400/10万，部分国家超500/10万。值得注意的是，印度尼西亚被重新划归西太平洋区域后，该区域患者占比升至全球27%，成为第二大高负担区域^[22]。

我国结核病防控成效显著，2024年估算新发69.6万例，较2023年下降5.8%，发病率降至49/10万，较2020年累计下降15.5%，在高负担国家中排位从第3位降至第4位，首次进入中低流行国家行列。此外，我国估算结核病死亡2.5万例，死亡率1.9/10万，新发耐多药/利福平耐药病例2.8万例，占全球7.1%^[22]。我国先后开展五次全国结核病流行病学调查，为防控工作提供重要数据支撑^[23]，前三次明确了当时疫情真实情况，

后两次显示 2010 年较 2000 年涂阳肺结核患病率下降 45.9%^[24]，截至 2019 年全国发病率降至 55.55/10 万，提前实现 2020 年目标^[25]。

2018 年提出的全球 2035 年终止结核病目标为：较 2015 年发病率下降 90%、死亡率下降 95%，无家庭因结核病面临灾难性支出^[26]。我国肺结核在乙类传染病中发病数始终居第二位，总体呈下降趋势，但东中西部差异显著，西部地区活动性肺结核患病率 695/10 万，远高于中东部^[24]。我国肺结核存在明显空间聚集区，主要分布在新疆、西藏、青海等中西部地区^[27,28]，2018 年高发省份为新疆、西藏、青海和贵州。相关研究显示，新疆南北部发病率差异明显^[21]，武汉远城区、贵州望谟县周边、广州西部和南部外来人口聚集区为高发区域^[29-31]，提示经济欠发达地区高发、发达地区受流动病例影响大，我国实现 2035 年目标仍具挑战，需加强西部地区防控。

1.3 肺结核发病趋势预测研究现状

肺结核发病趋势预测是结核病精准防控的核心，其研究兼具理论创新与公共卫生实践价值。疾病预测模型整合统计学、机器学习、人工智能等技术，通过分析历史疾病数据与风险因素识别流行趋势，为公共卫生部门提供早期预警支持。

传统预测模型以逻辑回归模型、Cox 比例风险模型、时间序列模型为代表。逻辑回归模型将线性回归结果转化为概率值，假设因变量服从二项分布，构建自变量与因变量的逻辑关系以描述事件发生概率^[32,33]，优势在于结构简单、易于解释且可处理多类别问题，但对数据分布有特定要求，复杂模型易过拟合。

随着技术发展，新型模型逐步应用于疾病预测领域。深度学习模型（如 CNN、RNN、LSTM 等）可高效处理医学图像、电子病历等大规模数据，提升预测准确性与效率^[34,35]；机器学习模型中，随机森林作为多决策树集成方法^[36]，具备高维数据处理能力与抗过拟合性，却存在计算成本高、解释需专业知识的问题；决策树通过直观规则实现分类预测^[37]，但易受噪声影响过拟合；支持向量机通过最优超平面处理高维数据^[38]，泛化能力强但对参数敏感、计算复杂；朴素贝叶斯模型基于特征独立假设快速分类，但假设可能与实际数据不符^[39]；多层感知机通过神经网络学习复杂模式^[40]，表达能力突出但小样本下易过拟合、训练耗时等。

国内外研究呈现多样化特征。国内方面，康颖等人、申采玉等人运用机器学习模型挖掘数据规律^[41,42]，刘文博等人采用时间序列模型捕捉时间演变趋势^[43]，中浦洁等人借助随机森林提升预测稳定性^[44]。作为结核病高负担国家，我国当前多采用 ARIMA 模型^[45,46]，但单一模型存在计算不充分、考虑因素片面等问题，导致预测精准度不足，因此开发高准确性、高稳定性的预测系统成为研究重点，模型性能提升、算法架构优化及结果可解释性增强是核心挑战。国外研究中，Wang F、Norrman A 等人^[47,48]的机器学习模

型、Mirahmadizadeh A 等人^[49]的时间序列模型、Yu X Y 等人^[50]的 Cox 回归模型应用广泛，Maki S 等人^[51]的多模态数据融合模型通过整合多类型数据挖掘潜在信息，显著提升了预测的全面性与准确性，为国外相关研究提供了重要支撑。

1.4 研究目的与意义

本研究聚焦新疆生产建设兵团肺结核防控需求，以 2008-2024 年肺结核报告发病数据为核心，旨在分析该区域肺结核的时间、人群及空间分布特征、聚集性与变化趋势。构建多类预测模型并进行对比，筛选出最优模型。结果可为兵团肺结核防控策略制定、医疗资源配置及预警体系建立提供数据参考。

本研究聚焦新疆生产建设兵团这一特殊管理地区，基于 2008-2024 年肺结核发病数据，分析其时间、空间及人群流行特征，可为我国边疆地区的结核病流行病学研究补充基础数据。针对兵团数据的复杂性，研究整合了时间序列、机器学习和组合模型，对比了不同模型在该区域的拟合精度、稳定性与适用条件，明确了各类模型的优劣及优化方向，为同类边疆区域的传染病预测提供一定方法学参考。研究结果进一步明确了各师团的高发区域、高发时段及高危人群，卫生健康部门可据此制定差异化防控策略、开展针对性筛查；发病趋势预测结果可用于指导防控资源向高发区域和时段倾斜，有助于缓解边疆医疗资源分布不均的矛盾。筛选出的最优模型可作为贴合兵团实际的预警工具，用于提前预判聚集性疫情风险，为及时采取防控措施提供数据支持。综上，本研究可为兵团制定结核病防治目标、细化防控方案提供参考，也可为我国边疆同类区域的肺结核防控工作提供借鉴。

第 2 章 资料与方法

2.1 研究区域资料

新疆生产建设兵团下辖 14 个师，空间分布于新疆天山南北的 14 个地州市，总面积约 7 万余平方公里，总人口约 270 万。其中，第一师阿拉尔市、第二师铁门关市、第三师图木舒克市、第十四师昆玉市分布于南疆塔里木盆地边缘及和田地区绿洲带；第六师五家渠市、第七师胡杨河市、第八师石河子市、第十一师建工师、第十二师布局于北疆准噶尔盆地南缘及乌鲁木齐周边；第四师可克达拉市、第五师双河市、第九师白杨市、第十师北屯市坐落于伊犁河谷、博尔塔拉及中哈、中蒙边境区域；第十三师新星市则位于东疆哈密盆地^[52]，见图 2-1。



图 2-1 新疆维吾尔自治区地图示意图

注：该图基于新疆维吾尔自治区标准地图服务网站下载的审图号为 GS（2024）0568 号的标准地图制作，底图边界无修改。

2.2 研究对象与数据来源

本研究对象为2008年1月1日至2024年12月31日期间，新疆生产建设兵团疾病预防控制中心通过传染病疫情信息网络直报系统报告的全部肺结核患者病例。肺结核发病数据来源于该直报系统，患者信息包括性别、年龄、职业、出生日期、工作单位、现住详细地址、发病日期等；报告发病数以病例的发病日期所属年度为准进行统计。同期人口学数据来自新疆生产建设兵团统计年鉴^[53]。

2.3 纳排标准

肺结核病例按传染病报告卡发病日期统计，纳入2008年1月1日-2024年12月31日肺结核报告数据中，病例分类为临床诊断病例、确诊病例，且报告现住址为兵团各地市的对象；排出：报告信息错误、查无此人、重复报告或存在逻辑错误的病例；以及被排除陈旧性结核、经检测排除肺结核患者的病例。

2.4 肺结核诊断标准

根据中华人民共和国国家卫生和计划生育委员会发布的《肺结核诊断》（WS288-2017，2018年5月1日实施，替WS288-2008）标准^[54]，肺结核诊断以病原学（细菌学、分子生物学）检查和病理学结果为确诊核心依据，结合流行病学史（如肺结核患者密切接触史、免疫功能低下等高危因素）、临床表现（咳嗽咳痰 ≥ 2 周、痰中带血或咯血等可疑症状及相应体征）、胸部影像学检查（不同类型肺结核的特征性病灶表现，如原发性肺结核的肺内原发病灶与胸内淋巴结肿大、继发性肺结核的上叶尖后段/下叶背段浸润影或空洞等）、辅助检查（结核菌素皮肤试验、 γ -干扰素释放试验、胸腔积液腺苷脱氨酶升高等）进行综合分析，同时需与肺炎、肺癌、支气管扩张等肺部疾病相鉴别；儿童肺结核诊断除痰液病原学检查外，还需重视胃液病原学检查，确诊需严格遵循疑似病例、临床诊断病例、确诊病例的分类判定标准。

2.5 统计分析方法

2.5.1 发病率

在兵团总体发病率和各师发病率的计算过程中，使用年末人口数做分母来估计月发病率和年发病率。