

Original Article

Precision of tuberculosis diagnosis codes in the Central Denmark Region

Victor Næstholt Dahl¹, Pernille Grand Moestrup², Anders Koch^{3, 4}, Dorte Bek Folkvardsen⁵, Frauke Rudolf¹, Tina Nørregaard Gissel⁶ & Andreas Fløe²

1) Department of Infectious Diseases, Aarhus University Hospital, 2) Department of Respiratory Diseases and Allergy, Aarhus University Hospital, 3) Department of Infectious Disease Epidemiology and Prevention, Statens Serum Institut, 4) Department of Infectious Diseases, Copenhagen University Hospital – Rigshospitalet, 5) International Reference Laboratory of Mycobacteriology, Statens Serum Institut, 6) Department of Respiratory Diseases, Viborg Regional Hospital, Denmark

Dan Med J 2025;72(7):A12240847. doi: 10.61409/A12240847

ABSTRACT

INTRODUCTION. Correct use of tuberculosis (TB) diagnosis codes is essential for patient care, surveillance and resource allocation. We aimed to assess the positive predictive value (PPV) of TB diagnosis codes.

METHODS. In this retrospective cohort study, we identified patients with International Classification of Diseases, tenth version (ICD-10) TB diagnosis codes from 1 July 2020 to 30 June 2023, at two TB centres in the Central Denmark Region. Confirmed TB was defined as microbiologically confirmed TB, prescription of ≥ 3 first- or second-line TB drugs, or TB notification. All patients who did not meet these criteria and those who received fewer than three TB drugs or lacked TB notification underwent manual hospital record review to verify or exclude the TB diagnosis. PPVs were calculated as the proportion of confirmed TB diagnoses among all patients with a TB diagnosis code.

RESULTS. In total, 185/230 patients were confirmed to have TB, yielding a PPV of 80% (95% CI: 75; 85). The PPVs for TB microbiology, TB prescriptions and TB notification exceeded 95% individually. Excluding TB lupus codes increased the PPV to 89% (95% CI: 84; 93). Patients with more than one different type of TB diagnosis code had a PPV of 100% (95% CI: 93; 100). Additionally, PPVs were high when TB diagnosis codes appeared on multiple occasions, increasing with the number of occurrences (≥ 2 : 85%, ≥ 3 : 89%, ≥ 4 : 93%).

CONCLUSION. TB ICD-10 diagnosis codes demonstrate a moderately high PPV in Denmark, particularly when excluding TB lupus codes, highlighting the importance and complexities of diagnostic coding.

FUNDING. None.

TRIAL REGISTRATION. Not relevant.

Diagnosis codes in health registers are used mainly for healthcare administration but may be valuable tools for tuberculosis (TB) disease surveillance and research [1-3]. Precise coding is essential for correct patient care administration, effective public health monitoring and efficient resource allocation. However, previous studies have highlighted problems with the use of TB diagnosis [4, 5]. Over- or underreporting of TB diagnoses results in unreliable public health data, undermining disease surveillance and research efforts. Various factors may contribute to coding errors [6].

Denmark has a strong tradition of population-based research using diagnosis codes to identify patients with various diseases, such as mycobacterial disease [7, 8], and to evaluate comorbidity burdens [9, 10], among others. Nevertheless, TB diagnosis codes have not yet been validated in our setting. Therefore, in this study, we assessed

the positive predictive value (PPV) of TB diagnosis codes registered over three years at two TB centres covering the Central Denmark Region.

Methods

In this retrospective cohort study, we identified all patients with at least one TB International Classification of Diseases, 10th revision (ICD-10) diagnosis code in the hospital record system at Aarhus University Hospital, Aarhus, and Viborg Regional Hospital, Viborg, Denmark, between 1 July 2020 and 30 June 2023. Using data from a local data hub, the Business Intelligence Data Warehouse, we evaluated patients with the following ICD-10 diagnosis codes [11]: A15-A19.9 (Tuberculosis), B20.0A (HIV disease resulting in tuberculosis), J65 (Pneumoconiosis associated with tuberculosis), K67.3 (Tuberculous peritonitis), M49 (Tuberculosis of the spine), M90 (Tuberculosis of bone), and N74 (Tuberculous infection of cervix uteri and female tuberculous pelvic inflammatory disease).

For each patient with a TB ICD-10 diagnosis code, we counted the number of TB diagnosis codes and recorded the number of types of codes. For all patients, we collected microbiological and notification information from the centralised International Reference Laboratory of Mycobacteriology and the nationwide TB notification register at Statens Serum Institut, Copenhagen, respectively, to verify the diagnosis. This data collection was expanded by administrative hospital data on TB treatment, including both first-line (ethambutol, isoniazid, pyrazinamide and rifamycins (rifabutin, rifampicin, and rifapentine)) and second-line drugs (amikacin, bedaquiline, capreomycin, cycloserine, delamanid, ethionamide, imipenem/cilastatin, levofloxacin, linezolid, meropenem, moxifloxacin, para-aminosalicylic acid (PAS), pretomanid, protionamide), including combination drugs. To include all relevant diagnostic instances, all data points were gathered within one year before and after the study period. Data were merged using a unique identifier for each patient [12].

Patients were defined as having confirmed TB if they had 1) at least one polymerase chain reaction (PCR) or culture positive for *Mycobacterium tuberculosis*, 2) were prescribed three or more first- or second-line TB drugs or 3) were notified with TB. All patients who did not meet these criteria and those who received fewer than three TB drugs or lacked TB notification underwent manual hospital record review to verify or exclude the TB diagnosis.

Descriptive statistics were presented using numbers and proportions for categorical variables and medians and IQRs for continuous variables. The Wilcoxon rank sum test, Pearson's Chi-squared test and Fisher's exact test were used to compare patients with confirmed TB and those without TB, as appropriate. A PPV, defined as the proportion of true positives out of the sum of true and false positives, was calculated by determining the proportion of patients with confirmed TB among all with a TB diagnosis code. PPVs were also calculated separately for all those with TB microbiology, TB prescriptions and TB notification suggestive of TB to evaluate the predictive precision of each individually. In addition, PPVs were calculated for patients where the TB diagnosis code appeared on multiple occasions and for patients with more than one type of TB diagnosis code. Wilson's method for binomial proportions was used to compute 95% confidence intervals (CIs).

Trial registration: not relevant.

Results

During the study period, 230 patients were identified with an ICD-10 TB diagnosis code. The median patient age was 45 years (IQR: 30-59, range: 0-95 years), with the majority being males (56%, $n = 129$). The median number of TB diagnosis codes per unique patient was nine (IQR: 5-13) (from one year before to one year after the study

period). A minority had two (20%, n = 46) or three (1.7%, n = 4) different types of TB diagnosis codes.

In total, 185 patients had confirmed TB. Among these, 172 patients (93%) satisfied at least one of the following conditions: TB was microbiologically confirmed by PCR or culture in 75% (n = 139), 90% (n = 167) received three or more first- or second-line TB drugs and 86% (n = 159) were notified with TB, while a few (7%, n = 13) did not meet any of these conditions (**Figure 1**). Thirteen additional patients had confirmed TB after reviewing hospital records (**Table 1**). Many of the inappropriately coded patients had cutaneous or systemic lupus erythematosus (56%, n = 22/39). Among patients who were not notified as having TB but had microbiologically confirmed TB or received three or more first- or second-line TB drugs (n = 19), three had infections with non-tuberculous mycobacteria (NTM), and three had Bacillus Calmette-Guérin infection following intravesical immunotherapy, while the rest had TB (n = 13). Three of these were notified with TB more than a year before the study period. Significant differences were observed between patients with confirmed TB and those without TB concerning sex, department, type and number of TB codes, and sampling for mycobacteriological examination (**Table 2**). All patients who were prescribed fewer than three first- or second-line TB drugs (n = 5) were confirmed to have TB but were in the continuation phase of treatment.

FIGURE 1 Venn diagram of patients with confirmed tuberculosis (n = 172/185; 13 were in neither of the groups), stratified by microbiological confirmation of tuberculosis (TB), first- or second-line tuberculosis treatment and notification.

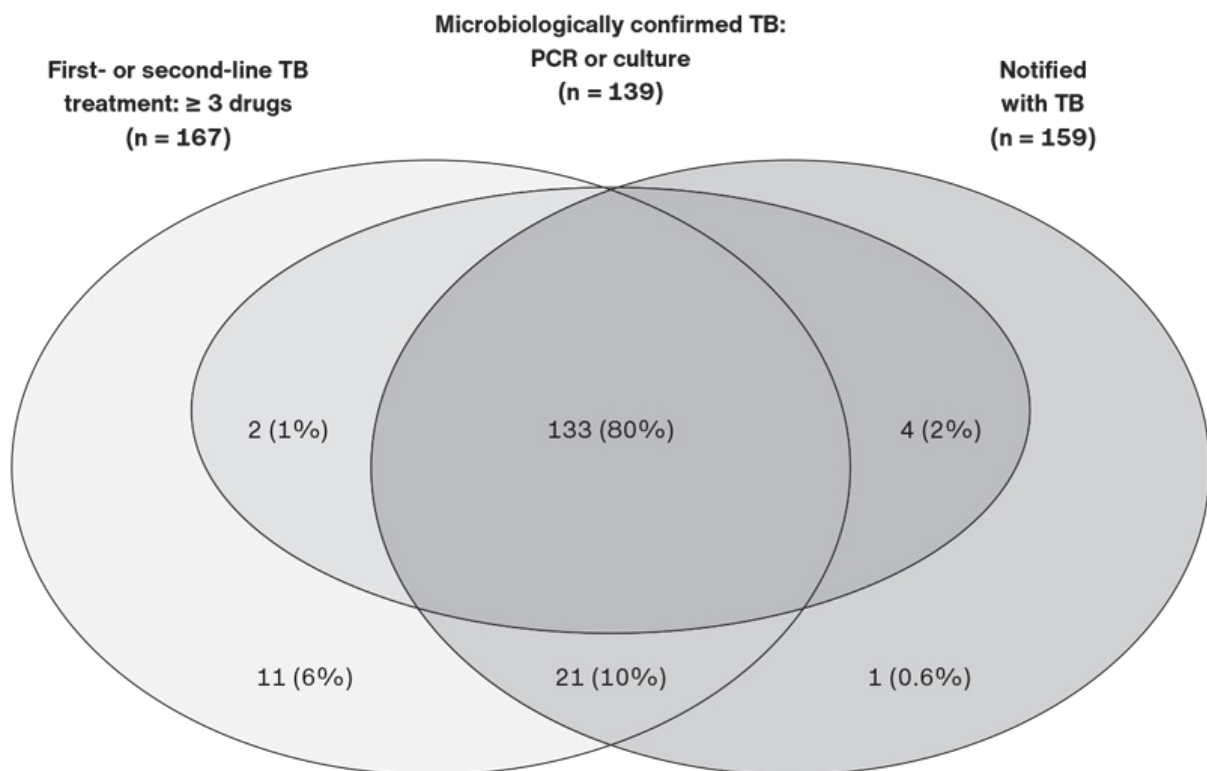


TABLE 1 Final diagnosis after manual hospital record review for 52 patients^a.

Final diagnosis	n (%)
Error code without relevance	3 (5.8)
Cutaneous or systemic lupus erythematosus	21 (40)
Cutaneous or systemic lupus erythematosus + TB infection ^b	1 (1.9)
NTM disease	1 (1.9)
Provisional TB diagnosis	2 (5.0)
TB disease ^c	13 (25)
TB infection ^b	11 (21)
Total	52 (100)

NTM = non-tuberculous mycobacteria; TB = tuberculosis.

a) Patients were defined as having confirmed TB if they had: 1) ≥ 1 PCR or culture positive for *Mycobacterium tuberculosis*, 2) were prescribed ≥ 3 1st- or 2nd-line TB drugs or 3) were notified with TB; all patients who did not meet these criteria and those who received < 3 TB drugs or lacked TB notification underwent manual hospital record review (n = 52) to verify or exclude the TB diagnosis.

b) Previously termed latent TB.

c) Previously termed active TB.

TABLE 2 Characteristics of patients with confirmed tuberculosis and without tuberculosis^a.

	Confirmed TB (n = 185)	Without TB (n = 45)	p value ^b
Age, median (IQR), yrs	44 (30, 56)	50 (35, 70)	0.4
Males, n (%)	108 (58)	21 (47)	0.023
Hospital, n (%)			0.3
Hospital 1	145 (78)	31 (69)	
Hospital 2	40 (22)	14 (31)	
Department, n (%)			< 0.001
Respiratory medicine	85 (46)	3 (6.7)	
Internal medicine: Hospital 2	39 (21)	11 (24)	
Infectious diseases	41 (22)	5 (11)	
Paediatrics	16 (8.6)	4 (8.9)	
Dermatology	0	13 (29)	
Other ^c	4 (2.2)	9 (20)	
TB diagnosis code, n (%)			< 0.001
Lung	41 (22)	1 (2.2)	
Lung, confirmed by culture only	38 (21)	3 (6.7)	
Lung, confirmed by unspecified means	22 (12)	5 (11)	
Lupus	0	22 (49)	
Peripheral lymphadenopathy	14 (7.6)	0	
Lung, bacteriologically and histologically negative	10 (5.4)	1 (2.2)	
Other ^d	60 (32)	13 (29)	
TB diagnosis codes, median (IQR), n	10 (7-14)	2 (1-4)	< 0.001
≥ 1 type of TB diagnosis codes, n (%)	50 (27)	0	< 0.001
Sample sent for mycobacteriological examination, n (%)	175 (95)	11 (24)	< 0.001

TB = tuberculosis.

a) Patients were defined as having confirmed TB if they had: 1) ≥ 1 PCR or culture positive for *Mycobacterium tuberculosis*, 2) were prescribed ≥ 3 1st- or 2nd-line TB drugs or 3) were notified with TB; all patients who did not meet these criteria and those who received < 3 TB drugs or lacked TB notification underwent manual hospital record review (n = 52) for verification or de-verification of the TB diagnosis.

b) Wilcoxon rank sum test, Pearson's χ^2 test and Fisher's exact test, as appropriate.

c) Departments with < 10 observations; the departments included a Centre of Diagnostic Investigation, Department of Cardiology, Department of Rheumatology, Department of Surgery, Department of Nephrology, and Department of Ear-, Nose and Throat Surgery.

d) TB diagnosis codes with < 10 observations. Please see supplementary for details.

The overall PPV of the TB diagnosis codes was 80% (95% CI: 75; 85). The PPVs for TB microbiology, TB prescriptions and TB notification exceeded 95% individually (Table 3). As many patients had a TB lupus diagnosis code (A18.4A), we recalculated the PPV as a post hoc sensitivity analysis after excluding these patients. When the TB lupus diagnosis code was excluded, the PPV increased to 89% (95% CI: 84; 93). Patients with more than one type of TB diagnosis code had a PPV of 100% (95% CI: 93; 100). Additionally, PPVs were high when TB diagnosis codes appeared on multiple occasions, increasing with the number of occurrences (≥ 2: 85%, ≥ 3: 89%, ≥ 4: 93%).

TABLE 3 Positive predictive value of tuberculosis diagnosis codes for identifying confirmed tuberculosis, overall and stratified by microbiology, treatment, notification and diagnosis code frequency.

Category	PPV (95% CI), % ^a	Confirmed/total, n/N ^b
TB diagnosis code overall	80.4 (74.8; 85.0)	185/230
TB diagnosis code excluding lupus	88.9 (84.0; 92.5)	185/208
<i>Microbiology</i>		
With	98.6 (95.0; 99.6)	139/141
Without	51.7 (41.5; 61.8)	46/89
<i>Treatment</i>		
With	96.5 (92.6; 98.4)	167/173
Without	31.6 (21.0; 44.5)	18/57
<i>Notification</i>		
With	100.0 (97.6; 100.0)	159/159
Without	36.6 (26.4; 48.2)	26/71
<i>TB diagnosis codes</i>		
≥ 2	85.0 (79.6; 89.2)	181/213
≥ 3	89.4 (84.4; 93.0)	178/199
≥ 4	92.6 (88.0; 95.5)	175/189
> 1 type	100.0 (92.9; 100.0)	50/50

CI = confidence interval; PPV = positive predictive value; TB = tuberculosis.

a) PPV = (confirmed cases/total cases) × 100, where the numerator represents the number of confirmed TB cases in each category, and the denominator includes the total number of cases within each category, e.g., those with and without microbiology suggestive of TB, those with and without notification of TB, those with and without treatment and those with and without > 3 diagnosis codes, etc. 95% CI calculations were performed using Wilson's method for binomial proportions with R package binom (version 1.1-1.1).

b) Patients were defined as having confirmed TB if they had: 1) ≥ 1 PCR or culture positive for *Mycobacterium tuberculosis*, 2) were prescribed ≥ 3 1st- or 2nd-line TB drugs or 3) were notified with TB; all patients who did not meet these criteria and those who received < 3 TB drugs or lacked TB notification underwent manual hospital record review (n = 52) to verify or exclude the TB diagnosis.

Discussion

In this study, 75% of patients with a TB diagnosis code had a confirmed diagnosis by PCR or culture, TB treatment or TB notification. When combining this information with findings from manual review of hospital records, the PPV of the TB diagnosis codes was to 80%. Most false positives (56%) were attributed to patients with cutaneous or systemic lupus erythematosus, who were mistakenly coded as having lupus vulgaris (a form of cutaneous TB). When TB lupus diagnosis codes were excluded, the PPV increased to 89%, highlighting the overall usefulness of TB diagnosis codes - except for TB lupus-related codes - for disease surveillance and research.

Administrative health data are increasingly used for epidemiological research and provide a low-cost approach to obtaining large amounts of real-life data, especially for rare diseases such as TB, which remains uncommon in high-income countries. However, as data are not obtained for research purposes, they may be incomplete or

imprecise [5]. A smaller study from Northern California and Portland, United States, used the US Centres for Disease Control and Prevention criteria for reference to confirm TB and found notably lower PPVs of 54% and 9%, respectively, for patients having at least one TB diagnosis code [13]. The PPVs increased to 87% and 46%, respectively, when including medication dispensing data, and to 71% and 21% when including only patients with at least two TB codes [13]. Patients with TB infection (previously termed latent TB) or patients whose TB diagnosis was later disconfirmed were discussed as the main reasons for incorrectly coded diagnoses. We also observed that those with a TB diagnosis code appearing on multiple occasions had higher PPVs, increasing with the number of occurrences (≥ 2 : 85%, ≥ 3 : 89%, ≥ 4 : 93%). Moreover, we found that all patients with more than one type of TB diagnosis code had confirmed TB.

A review from 2017 summarised the diagnostic accuracy of TB codes and found that PPVs ranged widely, 1.3-100%, and sensitivities fell in the 20-100% range, while specificities and negative predictive values were unavailable [5]. The study concluded that combining diagnosis codes and pharmacological data augmented the PPV considerably. In our study, we also included information on TB drug prescription to validate the diagnosis and found a high PPV of TB treatment. A small group of patients ($n = 5$) received less than three TB drugs. However, through the review of hospital records, we confirmed that these patients had TB disease (previously termed active TB) and not TB infection, as they were in the continuation treatment phase (with two drugs) or were clinically monitored after the end of TB treatment.

In many EU countries, including Denmark, laboratories are, in addition to the diagnosing clinician, required to notify confirmed cases of *M. tuberculosis* [14]. Moreover, TB incidence rates are low, and diagnostics and treatment are centralised. Most TB patients are presumably in contact with the hospital at some point, leading to an almost complete overlap between the notification system, administrative hospital register and mycobacteriological database (if confirmed), which all have nationwide coverage [14]. A capture-recapture study estimated the completeness of TB notifications in Denmark at 98.4% [14]. In that study, however, treatment data were not used. Since the TB notification system is already complete, relying on diagnosis codes may not add much value and could introduce false positives. Still, treatment data could improve case ascertainment and help validate or complement TB notifications. A study from the Southern Region of Denmark spanning five years characterised patients with a TB diagnosis code not found in the TB notification register [2]. It found that 28.9% ($n = 30$) of clinically diagnosed, culture-negative TB cases were not notified, corresponding to an underreporting rate of 7.5% [2]. Additionally, the study observed that 71.1% ($n = 96/135$) of all cases with a TB diagnosis code who were not notified of TB were misclassified. These included 20% misdiagnoses, 18.5% NTM infections, 18.5% withdrawn suspicions of TB and 14.1% TB infection. We found that 5% were not notified among those with confirmed TB (including notifications before the study period). Therefore, relying solely on the TB notifications register potentially entails a small risk of excluding patients with a clinical diagnosis of TB – although presumably few patients – while relying on diagnosis codes may lead to overestimations. Detailed clinical information, such as treatment data, is not available in the notification register. Additionally, study permissions and data access may be more readily and quickly obtainable at the local level than nationwide.

The main limitation of the study was that we did not include patients without TB diagnosis codes, which prevented us from estimating sensitivity, specificity and negative predictive values. Future studies should include such patients, alongside microbiology, treatment and notification information to enable a more comprehensive evaluation of diagnostic accuracy and precision. Nonetheless, our findings remain valuable for those using TB diagnosis codes in disease surveillance and research, and we believe that our results are generalisable to the rest of the country and comparable settings.

Conclusion

TB ICD-10 diagnosis codes in Denmark have a moderately high PPV, increasing considerably when TB lupus diagnosis codes are excluded. While TB diagnosis codes provide a reliable tool for disease surveillance and research through administrative health data, certain codes introduce potential imprecision. This highlights both the strengths and complexities of diagnostic coding.

Correspondence Victor Næstholt Dahl. E-mail: victor.dahl@rm.dk

Accepted 15 April 2025

Published 12 June 2025

Conflicts of interest none. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. These are available together with the article at ugeskriftet.dk/dmj

References can be found with the article at ugeskriftet.dk/dmj

Cite this as Dan Med J 2025;72(7):A12240847

doi 10.61409/A12240847

Open Access under Creative Commons License [CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Supplementary material: <https://content.ugeskriftet.dk/sites/default/files/2025-04/a12240847-supplementary.pdf>

REFERENCES

1. Andersen RMØ, Bjørn-Præst SO, Gradel KO, Nielsen C, Nielsen HI. Epidemiology, diagnostic delay and outcome of tuberculosis in North Jutland, Denmark. *Dan Med Bull.* 2011;58(3):A4256
2. Thrane FD, Andersen PH, Johansen IS, Holden IK. Underreporting of patients diagnosed with tuberculosis in the Region of Southern Denmark. *Scand J Public Health.* 2020;48(8):870-876. <https://doi.org/10.1177/1403494819884433>
3. Thulstrup AM, Mølle I, Svendsen N, Sørensen HT. Incidence and prognosis of tuberculosis in patients with cirrhosis of the liver: a Danish nationwide population-based study. *Epidemiol Infect.* 2000;124(2):221-225. <https://doi.org/10.1017/S0950268899003593>
4. Iqbal SA, Isenhour CJ, Mazurek G, Truman BI. Diagnostic code agreement for electronic health records and claims data for tuberculosis. *Int J Tuberc Lung Dis.* 2020;24(7):706-711. <https://doi.org/10.5588/ijtld.19.0792>
5. Ronald LA, Ling DI, FitzGerald JM, et al. Validated methods for identifying tuberculosis patients in health administrative databases: systematic review. *Int J Tuberc Lung Dis.* 2017;21(5):517-522. <https://doi.org/10.5588/ijtld.16.0588>
6. O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM. Measuring diagnoses: ICD code accuracy. *Health Serv Res.* 2005;40(5 Pt 2):1620-1639. <https://doi.org/10.1111/j.1475-6773.2005.00444.x>
7. Fløe A, Hilberg O, Wejse C, Ibsen R, Løkke A. Comorbidities, mortality and causes of death among patients with tuberculosis in Denmark 1998-2010: a nationwide, register-based case-control study. *Thorax.* 2018;73(1):70-77. <https://doi.org/10.1136/thoraxjnl-2016-209240>
8. Pedersen AA, Løkke A, Fløe A, Ibsen R, Johansen IS, Hilberg O. Nationwide increasing incidence of nontuberculous mycobacterial diseases among adults in Denmark: eighteen years of follow-up. *Chest.* 2024;166(2):271-280. <https://doi.org/10.1016/j.chest.2024.03.023>
9. Nordholm AC, Andersen AB, Wejse C, et al. Mortality, risk factors, and causes of death among people with tuberculosis in Denmark, 1990-2018. *Int J Infect Dis.* 2023;130:76-82. <https://doi.org/10.1016/j.ijid.2023.02.024>
10. Mathiasen VD, Eiset AH, Andersen PH, Wejse C, Lillebaek T. Epidemiology of tuberculous lymphadenitis in Denmark: a nationwide register-based study. *PLoS One.* 2019;14(8):e0221232. <https://doi.org/10.1371/journal.pone.0221232>
11. International Statistical Classification of Diseases and Related Health Problems 10th Revision, version 2019. World Health Organization, 2019
12. Schmidt M, Pedersen L, Sørensen HT. The Danish Civil Registration System as a tool in epidemiology. *Eur J Epidemiol.*

- 2014;29(8):541-549. <https://doi.org/10.1007/s10654-014-9930-3>
13. Winthrop KL, Baxter R, Liu L, et al. The reliability of diagnostic coding and laboratory data to identify tuberculosis and nontuberculous mycobacterial disease among rheumatoid arthritis patients using anti-tumor necrosis factor therapy. *Pharmacoepidemiol Drug Saf.* 2011;20(3):229-235. <https://doi.org/10.1002/pds.2049>
 14. Straetemans M, Bakker MI, Alba S, et al. Completeness of tuberculosis (TB) notification: inventory studies and capture-recapture analyses, six European Union countries, 2014 to 2016. *Euro Surveill.* 2020;25(12):1900568. <https://doi.org/10.2807/1560-7917.ES.2020.25.12.1900568>