

Original Article

## Mantoux test revisited: Variability in reading tuberculin test in pediatric population

Dimple Goel<sup>1,2,3</sup>, Mukta Mantan<sup>1</sup>, Gulshan Rai Sethi<sup>1</sup>

<sup>1</sup> Department of Pediatrics, Maulana Azad Medical College and associated Lok Nayak, Hospital, New Delhi, India

<sup>2</sup> Pediatric and Child Health Department, University of Sydney, Sydney, Australia

<sup>3</sup> Sydney Children's Hospital Network, Westmead, Sydney, Australia

### Abstract

**Introduction:** Mantoux test aids in the diagnosis of tuberculosis (TB), however its application and interpretation are dependent on multiple factors.

**Methodology:** A prospective study enrolling 400 children (aged 2-12) suspected to have tuberculosis. All participants received Mantoux test with two different strengths (1 TU and 5 TU) of Purified Protein Derivative (PPD) on different forearms. The test was read by two readers after  $48 \pm 2$  and  $72 \pm 2$  hours. Primary outcome was difference in the size of induration when read by two readers (interobserver variability). Secondary outcomes were difference in the size of induration at different intervals, with different strengths of PPD and percentage positivity of Mantoux test in TB patients.

**Results:** Statistically significant difference was seen in the size of induration when read by two different readers, with fair to moderate agreement when read at 48 and 72 hours (1 TU:  $p = 0.002$ ,  $k = 0.52$  and  $p = 0.1$ ,  $k = 0.73$  respectively, 5 TU:  $p = 0.001$ ,  $k = 0.39$  and  $p = 0.0009$ ,  $k = 0.33$  respectively). Tendency of under-reading occurred when size of induration was close to significant level (10-14 mm). Size of induration was similar when read at 48 or 72 hours (1 TU:  $p = 0.9$ , 5 TU:  $p = 1.0$ ). Mantoux positivity rate in patients with TB was more with 5 TU as compared to 1 TU (61.2% vs. 16.3%).

**Conclusions:** There is significant interobserver variability with a tendency to under-read around the cutoff point. The use of 5 TU PPD at 48 hours by a trained physician can aid in early and more reliable diagnosis of TB.

**Key words:** interobserver variability; Mantoux test; tuberculin test; tuberculosis; PPD.

*J Infect Dev Ctries* 2018; 12(8):625-630. doi:10.3855/jidc.9601

(Received 13 July 2017 – Accepted 18 April 2018)

Copyright © 2018 Goel *et al.* This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Introduction

The Mantoux test (tuberculin test) has been the traditional method for detection of infection with tubercular bacilli. However, its application is often undermined by difficulties in interpretation of results. Variation is inherent in all phases of tuberculin testing procedure and size of reaction is dependent on many factors besides actual sensitivity. According to a study, 2% error in measurement reduces the accuracy of Mantoux test by 25% and the impact exceeds 50% for 5% error [1]. The unreliability of tuberculin test readings due to variability among readers is of continued concern [1]. Other important confounding factors are the use of different methods for measuring the size of induration, time variability in reading the test and the strength of Purified Protein Derivative (PPD) used for the test [2]. The Mantoux test is one of the important supportive tests for diagnosis of tuberculosis (TB) in the pediatric population and variability in the test results may lead to a dilemma in the diagnosis of

childhood TB. The present study was planned to assess the interobserver variation in reading the Mantoux test with currently available PPD.

### Methodology

This prospective observational study was conducted between August 2009 and February 2010 in the pediatrics department of a tertiary care hospital in northern India. Patients of age-group 2-12 years, seen in either pediatric outpatient clinics or admitted in the pediatric wards, who were suspected to have TB by the treating physician and requested for a Mantoux test, were enrolled. Patients in whom either strength of the PPD [1 Tuberculin Unit (TU) or 5 TU] could not be administered correctly, i.e. given subcutaneously, or if leakage occurred at the injection site, or if the size of raised bleb was inadequate, were excluded from final analysis.

Tuberculin vials containing 1 TU and 5 TU doses per 0.1 mL were procured from Span Diagnostics (173-

B, New industrial estate, Udhna, Surat-394210, Gujarat, India) which is a commercial manufacturer of PPD in India. All vials belonged to the same lot (lot no-3409 for 1 TU and 3640 for 5 TU with expiry in January 2011) and each vial contained 5mL of the product, phenol (< 0.5%) as a preservative and Tween 80 (0.005%) as a stabilizer. All the vials were preserved in a personal refrigerator at 2-8°C. The Mantoux test was administered by standard technique as recommended by the WHO<sup>3</sup>. The researcher was trained for three months, carrying out over 200 Mantoux tests under expert supervision at a validated training centre (NDTB centre), till she was certified to perform the Mantoux test accurately. This included appropriate administration of PPD as recommended by the WHO [3] and reading the Mantoux test by the ballpoint pen technique [4] for measuring the maximum transverse margin of the induration (using a transparent flexible ruler).

The test was performed by injecting 0.1 mL of tuberculin intradermally into the middle one-third of the volar surface of right and left forearm with 1 TU and 5 TU, the arms receiving 1 or 5 TU of PPD being alternated between patients. Airtight plastic disposable tuberculin syringes of 1 mL fitted with a 26-gauge needle were used to make the injection, producing a distinct pale elevation on the skin (wheal) of 6-10 mm in diameter. It was ensured that the site selected was free from any scar or vein and was at a significant distance from the intravenous line.

Subsequently, the size of induration was read by two readers after 48 ± 2 hours and 72 ± 2 hours on both forearms. Two observers in the study were, research reader (X) who was trained earlier and the other reader (Y) was the pediatric resident on duty in the ward from where cases were enrolled. The size of induration was first measured by reader Y on both forearms and the readings were noted on a paper by reader Y. Subsequently, the pen marks were removed (if any) with a spirit swab and the size of induration was re-measured by reader X within two hours. Only after this, reader Y was requested to give the readings to reader X. It was ensured that the same resident (Y) read the Mantoux test at 48 hours and 72 hours. In this way, both readers X and Y were blinded to each other’s reading at the time of performing the measurement. The readings measured by reader X was taken as standard. The results were unblinded for the purpose of case management after all the readings were recorded. The presence of unpleasant skin reactions such as tenderness, patient discomfort, site infection etc. were also recorded. The clinical details of the patient were

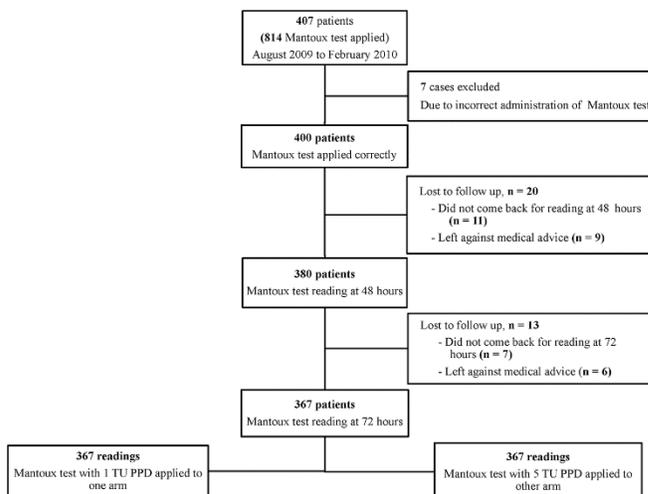
noted at the time of reading in a predesigned proforma. The final diagnosis such as TB or other was made by the treating clinician based on clinical evaluation, circumstantial evidence (history of contact, immunization status and malnutrition), radiological suggestion of TB and/or Mantoux test positivity (with 5 TU at 48 hours). Attempts were made to follow up all patients until a final diagnosis was made.

The primary outcome measured was variability in reading the size of induration of the Mantoux test among the two readers (interobserver variability). Secondary outcomes were the difference in the size of induration at 48 hours and 72 hours, the difference in size of induration after administration of two different strengths of PPD (1 TU PPD and 5 TU PPD) and the percentage positivity of Mantoux test in cases diagnosed as TB.

*Statistical Analysis*

To detect a difference of 2.5 mm with SD of 4 mm (based on existing literature) [5] at α error of 0.05 and a power of 90%, we required 300 paired observations (reading made by two different readers) with any given PPD administration. Wilcoxon Signed Rank Test was performed using SPSS Software package (version 19.0) to compare the frequency distribution of tuberculin reaction sizes. For interobserver comparison of tuberculin reaction size, the Chi-square test was applied. The Kruskal-Wallis test was applied for comparing the mean values. For all these tests, a p value of < 0.05 was considered significant and p value < 0.001 was considered highly significant. To study the level of agreement between two variables kappa statistic was applied.

**Figure 1.** The study design.



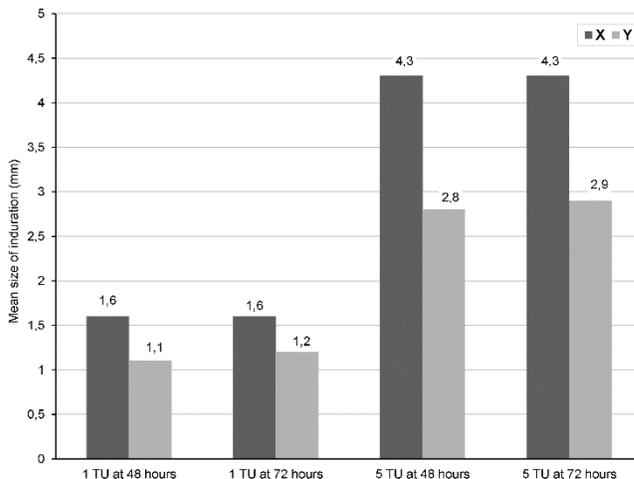
**Results**

We enrolled 407 patients, out of which 7 were excluded and 33 (8.25%) were lost to follow up (Figure 1). Baseline characteristics of the study population (n = 367) showed that 47.4% of patients were less than 5 years and 52.6% were 5-12 years of age. There were 198 males and 169 females (male to female ratio = 1.2:1). History of contact with sputum-positive cases was present in 10.3% of patients. History of BCG immunization was recorded in 185 patients (50.4%) with BCG scar found in 61.6% of immunized patients, 41.6% of cases were suffering from grade I-III and 13.3% of cases were suffering from grade IV malnutrition.

Out of 367 patients, 98 were finally diagnosed as TB. The majority of cases (33.7%) were tubercular meningitis and the rest included, 25% disseminated TB and miliary TB, 19.8% pulmonary TB (primary complex, progressive primary, pleural effusion), 9.2% abdominal TB, 8.2% tuberculoma and 4.1% lymph node TB. Among these only 6 (6.1%) were acid-fast bacilli (AFB) positive, of which 3 were diagnosed by sputum, 1 by lymph node biopsy, 2 by bronchoalveolar lavage. None of the gastric aspirates showed AFB. Out of 92 AFB negative patients, 60 (65.2%) were diagnosed on basis of clinical evaluation, circumstantial evidence (history of contact, immunization status, and malnutrition), radiology suggestive of TB and Mantoux test positivity (with 5 TU at 48 hours). In the remaining 32 patients (34.7%) despite the Mantoux test being negative, diagnosis of TB was made on basis of clinical evaluation, circumstantial evidence, and suggestive radiology.

There was fair to moderate agreement between two readers when read at 48 and 72 hours respectively (p =

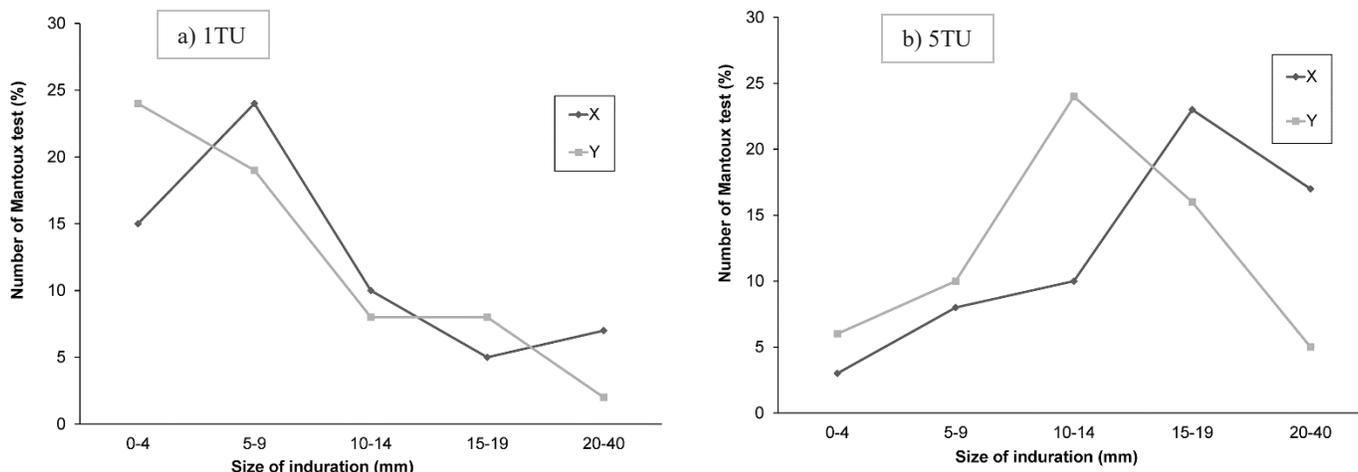
**Figure 2.** Variability in mean size of induration as read by X (research reader) and Y (other reader), at 48 and 72 hours with 1 TU and 5 TU



0.002, k = 0.52 and p = 0.1, k = 0.73 respectively) with 1 TU of PPD. With 5 TU of PPD there was fair agreement between two readers when read at 48 and 72 hours respectively (p = 0.001, k = 0.39 and p = 0.0009, k = 0.33 respectively) (Figure 2). On subgroup analysis of interobserver variability by applying Kappa to measurable readings of the Mantoux test (size >0 mm), a decrease in agreement between all four paired-readings groups were noted (5 TU at 48 hours, k = 0.39 and 72 hours, k = 0.33, 1 TU at 48 hours, k = 0.33 and 72 hours, k = 0.57).

The difference of > 5 mm between the two readers was present more frequently with 5 TU (Figure 3). At 48 hours with 1 TU PPD, 5 out of 22 Mantoux-positive cases (22.7%) were wrongly assigned as negative by reader Y. Additionally, with 5 TU PPD 17 out of 77 Mantoux-positive patients (22.1%) were wrongly assigned as Mantoux-negative by reader Y.

**Figure 3.** Variation in size of induration as read by X (research reader) and Y (other reader), at 48 hours with a) 1 TU and b) 5 TU



**Table 1.** Correlation of Mantoux test positivity with final diagnosis.

| Final diagnosis | Size of induration with 1 TU PPD |                | Size of induration with 5 TU PPD |                   |
|-----------------|----------------------------------|----------------|----------------------------------|-------------------|
|                 | <10 mm                           | ≥ 10 mm        | < 10 mm                          | ≥ 10 mm           |
| No TB           | 263 (97.8%)                      | 6 (2.2%)       | 258 (95.9%)                      | 11 (4.1%)         |
| TB              | 82 (83.7%)                       | 16 (16.3%)     | 38 (38.8%)                       | 60 (61.2%)        |
| <b>Total</b>    | <b>345 (94%)</b>                 | <b>22 (6%)</b> | <b>296 (80.7%)</b>               | <b>71 (19.3%)</b> |

TB-Tuberculosis; PPD- purified protein derivative (tuberculin); TU-Tuberculin Unit.

To study variation in the size of induration at 48 and 72 hours, readings of reader X were analyzed. Comparison of size of induration at 48 and 72 hours showed similar mean values in the 1 TU group (1.6 vs 1.6 mm) as well as in the 5 TU group (4.3 vs 4.3 mm),  $p = 0.9$  and  $p = 1.0$  respectively (Figure 2). In only 2 out of 367 cases, the readings which were negative (< 10 mm) at 48 hours became positive (≥ 10 mm) at 72 hours. However, no cases which were initially positive at 48 hours became negative at 72 hours.

Comparison of the Mantoux test with two strengths of PPD (1 TU and 5 TU) showed statistically significant difference in mean values (1.6 vs 4.3 mm,  $p = 0.0001$ ). In 53% of the cases the size of induration was the same with both 1 and 5 TU, in 11% of cases the difference was < 2 mm and in 16% cases, the difference was 3-5 mm. In a significant number of patients (16%) the difference was > 5 mm. Using a cut-off of ≥ 10 mm, 16.3% of TB cases were classified positive using 1 TU, while with 5 TU this number increased to 61.2% (Table 1 and 2). It was also seen that Mantoux reading of 0 mm was more frequent with 1 TU (72.4%) as compared to 5 TU (54.5%). False positive rate was similar with 1 TU and 5 TU (2.2% vs 4.1%,  $p = 0.23$ ). In total, 11 patients had a false positive test, out of which in 6 it was falsely positive with both 5 TU and 1 TU.

There was no significant impact of BCG vaccination on the size of induration. Median values in tubercular cases with BCG-vaccinated patients were 11 mm and with non-vaccinated were 13 mm, while in non-tubercular groups, medians were 0 mm in both groups. All patients with false positive results were BCG vaccinated. None of the patients had nontuberculous mycobacteria (NTM) infection. Severe malnutrition was as frequent in the non-tubercular group (13%) as in patients diagnosed to have TB (14%).

No significant adverse effects were seen in any of the patients. However, it was observed that the reaction size of >20 mm was usually associated with mild tenderness at the site of administration.

**Discussion**

We demonstrated significant interobserver variability in the reading of Mantoux test which impacted on the final result of the test. There was no significant difference in the size of induration when reading at either 48 or 72 hours. Additionally, Mantoux testing with 5 TU caused more reactivity when compared to 1 TU.

The tuberculin test has been the traditional method for detection of infection with tubercular bacilli. This test aids in the diagnosis of TB among children and to guide clinicians regarding the administration of chemoprophylaxis, which requires further confirmation with a more specific test such as Interferon-γ release assays (IGRAs) [6]. The diagnostic test for TB is the isolation of AFB from body fluids. Although reported AFB positivity is high in research studies [7-8], in routine practice the yield rate of isolation of AFB from body fluids in children is very poor. Therefore, the Mantoux test is crucial as a supportive test for diagnosis of TB in children. Variation is inherent in all phases of tuberculin testing procedure such as improper test administration, different readers, and strengths of PPD [5]. Additionally, BCG vaccination and NTM infection can make the interpretation of the test difficult due to false positive results [6]. This has been further confounded by non-availability of standardized PPD. We used two strengths of commercially available PPD (1 and 5 TU) and readings were done at both 48 and 72 hours, as an early result of Mantoux testing can help in early diagnosis. We found a statistically significant

**Table 2.** Variation in Mantoux test positivity with 1 TU and 5 TU in patients with tuberculosis.

| Size of induration with 1 TU | Size of induration with 5 TU |                    |
|------------------------------|------------------------------|--------------------|
|                              | Negative (< 10 mm)           | Positive (≥ 10 mm) |
| Negative (< 10 mm)           | 38 (38.8%)                   | 44 (73.3%)         |
| Positive (≥ 10 mm)           | 0 (0%)                       | 16 (26.7%)         |
| <b>Total</b>                 | <b>38 (38.8%)</b>            | <b>60 (61.2%)</b>  |

difference in the readings of two readers, both with 1 TU and 5 TU. However, the difference in readings was more statistically significant with 5 TU as compared with 1 TU. It was due to higher frequency of induration size of 0 mm with 1 TU (72.4%) as compared to with 5 TU (54.5%). Loudon *et al.* also suggested that in a population of a large number of strong reaction, the quantitative disagreement increases and found classification disagreements between pairs of readers in 9% of readings [8]. Similarly, Fine and colleagues found classification disagreements in 12% of cases [9]. In contrast to our study, the sample size in both of these studies was too small to draw a conclusion.

It was also observed that variation in reading was less among reaction sizes of < 5 mm and  $\geq$  15 mm. There is a tendency of under-reading when the size of induration is between 5-14 mm, which is around the standard cutoff point ( $\geq$  10 mm). Edwin L *et al.* also reported a similar tendency to under-read size of induration among health professionals [10]. In their study, 33% of positive cases were wrongly assigned as Mantoux-negative which is slightly higher than that observed in our study (22.1%).

We also noted that mean readings at 48 and 72 hours with both 1 TU and 5 TU were similar. However, two patients later diagnosed with TB were negative (< 10 mm) at 48 hours and became positive ( $\geq$  10 mm) at 72 hours. None of the cases which were initially positive at 48 hours became negative at 72 hours. Similar were the findings of Gopi *et al.*, who reported an insignificant difference between the reaction sizes at 48 and 72 hours [11].

In terms of difference with different strengths of PPD, we found that reactions with 1 TU PPD ( $1.6 \pm 4.2$  mm) were lower than with 5 TU ( $4.3 \pm 6.5$  mm) with increased true positive rate with 5 TU. The overall Mantoux positivity rate with 5 TU was 19.3% and with 1 TU it was 6%. This raises concerns whether tuberculin PPD RT-23 is losing its potency due to non-standardization of PPD, leading to low reactivity. A similar drop in potency has been reported in Korea [12]. Other reasons could be, that our study was a hospital-based study and the majority of diagnosed patients had a severe form of TB (tubercular meningitis, miliary TB or disseminated TB) and many had associated grade IV malnutrition. All these are known factors to cause false negative Mantoux test [13].

There exists variable practice across the world regarding the use of different strengths of PPD (1, 2, 5, 10 TU) and recommendations for different diagnostic cutoffs [14-17]. Chada *et al.* studied the Mantoux test with two strengths of PPD (1 and 2 TU) [18]. Their

results did not support the use of 2 TU (BCG lab Guindy), as many children with small reaction to 1 TU showed a moderate reaction to 2 TU and presumably got merged with the group of the truly infected children. They used standardized PPD as available at the time (year 2000), which might be different to currently available preparations or it is possible that the baseline population characteristics have changed over the years with differences in immune response. In our study, Mantoux testing with 5 TU led to higher true positive reactivity when compared to 1 TU but to recommend the most useful strength of PPD, further larger population-based studies are needed.

Our study is a large study in the pediatric population reflecting the current practice. It was conducted in a tertiary centre with a researcher adequately trained in Mantoux testing. It was a well-designed study with additional evaluation of the impact of Mantoux test variabilities on the final diagnosis of TB. The pragmatic approach of the study also made the result generalizable and useful for the clinicians in the given setting. The limitation of the study was inherent hospital bias with presumably more unwell TB patients.

## Conclusions

There is a general decline in the incidence of tuberculosis in developed countries. However, it continues to be a serious public health issue in developing countries. It is often challenging to have an early diagnosis in the pediatric population and the Mantoux test is an important diagnostic tool for the clinicians. We found significant interobserver variability, with a tendency to under-read the test around the cutoff point, which can be of clinical importance. In our population, we conclude that the use of 5 TU PPD for the Mantoux test and reading by a trained physician at 48 hours, can aid in early and more reliable diagnostic clue which can thereby facilitate early targeted treatment.

## Acknowledgements

We thank the National Delhi Tuberculosis Center (NDTB), Lok Nayak Hospital, New Delhi, India for the training and guidance provided to conduct the research.

We thank Dr Pankaj Gupta for his valuable expert advice and support throughout the process of research.

## Ethics

The study was approved by an institutional review board (expedited approval on May 2009). Additionally, the research study was discussed with both parents and child and

an informed written consent was taken from the parent or the guardian of the subject prior to administration of Mantoux test.

### Author's contribution

Dr Dimple Goel was the primary researcher who conducted the research, acquired data, analyzed and interpreted the data, prepared the first draft of the manuscript and reviewed the literature. Dr Gulshan Rai Sethi and Dr Mukta Manthan supervised the study design, data interpretation, and revised the manuscript.

### References

1. Toman K (1981) Sensitivity, specificity and predictive value of diagnostic tests. *Bull Internat Union against Tuberc* 56: 18.
2. Rhoades ER, Bryant RE (1980) The influence of local factors on the reaction to tuberculin. *Chest* 77: 190-193.
3. American Thoracic Society/Centers for Disease Control and Prevention (1981) The tuberculin skin test. *American Rev of Respir Dis* 124: 356-363.
4. Sokal JE (1975) Measurement of delayed skin test responses. *N Engl J Med* 293: 501-502.
5. Loudon RG, Lawson JR, Brown J (1963) Variation in tuberculin test reading. *Am Rev Respir Dis* 87: 852-861.
6. Starke JR, Committee on Infectious Diseases (2014) Interferon- $\gamma$  release assays for diagnosis of tuberculosis infection and disease in children. *Pediatrics* 134: 1763-1773.
7. American Thoracic Society (2000) Diagnostic standards and classification of tuberculosis in adults and children. *Am J Respir Crit Care Med* 161: 1376-1395.
8. Singh M, Moosa NV, Kumar L, Sharma M (2000) Role of gastric lavage and bronchoalveolar lavage in the bacteriological diagnosis of childhood pulmonary TB. *Indian Pediatrics* 37: 947-951.
9. Kendig EL, Kirkpatrick BV, Carter WH, Hill FA, Caldwell K, Entwistle M (1998) Under-reading of the tuberculin skin test reaction. *Chest* 113: 1175-1177.
10. Fine MH, Furcolow ML, Chick EW, Bauman DS, Arik M (1972) Tuberculin skin test reactions. *Am Rev Respir Dis* 106: 752-758.
11. Gopi PG, Vasantha M, Kolappan C, Narayanan PR (2007) Comparison of tuberculin reaction sizes at 48 and 72 hours among children in Tiruvallur district, South India. *Indian J Tuberc* 54: 152-156.
12. Kim SJ, Hong YP, Bai GH, Lee EK, Lew WJ (1998) Tuberculin PPD RT23: has it lost some of its potency? *IJTL D* 2: 857-860.
13. Sharma U, Saxena S (1982) Evaluation of BCG test in malnourished. *Indian J Pediatr* 49: 123-127
14. Ballal SG (1995) International controversies in interpreting the Mantoux test with special reference to Saudi Arabia. *J Fam Community Med* 2:11-8.
15. Larsson LO, Bentzon MW, Lind A, Magnusson M, Sandegard G, Skoogh BE, Boëthius G (1993) Sensitivity to sensitins and tuberculin in Swedish children. Part 5: a study of school children in an inland rural area. *Tubercle and Lung Dis* 74: 371-376
16. Morse DL, Hansen RE, Swalbach WG, Redmond SR, Grabau JC (1982) High rate of tuberculin conversion in Indochinese refugees. *JAMA* 248: 2983 - 2986.
17. Rowlands DF (1984) Tuberculin sensitivity in a Saudi military school population. *Saudi Medical Journal* 5: 183 - 189.
18. Chadha VK, Jagannatha PS, Nagaraj AV, Narayana Prasad D, Anantha N (2000) A comparative study of tuberculin reactions to 1 TU and 2 TU of PPD - RT23. *Indian J Tuberc* 47: 15-20.

### Corresponding author

Dr Dimple Goel  
Department of Respiratory and Sleep Medicine  
The Children's Hospital at Westmead  
Level 3, Corner of Hawkesbury Road and Hainsworth Street  
Westmead, New South Wales, Australia 2145  
Phone: 0061-469938682  
Fax: 02-88907490  
Email: dr\_dimplegoel@yahoo.co.in

**Conflict of interests:** No conflict of interests is declared.