Randomised controlled trial of the effect of ventilation tubes (grommets) on quality of life at age 1–2 years

M M Rovers, P F M Krabbe, H Straatman, K Ingels, G-J van der Wilt, G A Zielhuis

Abstract

Aims—To study the effect of treatment with ventilation tubes on quality of life in children aged 1–2 years with persistent otitis media with effusion (OME), as compared to watchful waiting.

Methods—Multicentre randomised controlled trial (n = 187) with two treatment arms: ventilation tubes and watchful waiting. Children were detected by auditory screening at the age of 9–12 months, and were subsequently diagnosed as having persistent (4–6 months) bilateral OME. Quality of life (TAIQOL and Erickson scales) was measured at 0, 6, and 12 months follow up.

Results—There was improvement in quality of life, but the ventilation tube group did not improve significantly more than the watchful waiting group. Although an attempt has been made to identify possible subgroups that benefit more, we were not able to find such subgroups, which might be a result of lack of power in this study.

Conclusion—Ventilation tubes do not have a substantial incremental effect on the quality of life of infants aged 1–2 years with uncomplicated persistent bilateral OME.

(Keywords: otitis media with effusion; ventilation tubes; randomised controlled trial; quality of life)

Otitis media with effusion (OME), common in childhood, is usually associated with mild to moderate hearing loss. Over the past three decades, much research has been done into the effects of OME on children’s development. Several studies showed impaired developmental outcomes, but others failed to show such associations. This may be a result of limitations in design, methodology, or selected study populations. Only a few adequately designed prospective studies have been published on the effect of OME on development.

Treatment with ventilation tubes (grommets) lessens hearing impairment in children with OME and might therefore prevent developmental problems. Most studies on the effect of ventilation tubes addressed speech development, language development, cognitive development, or behaviour, which were measured with standardised tests. In contrast, quality of life (QoL) focused on parents’ subjective perception of their children’s health status, was mostly ignored, while on theoretical grounds it might be assumed that an impaired development will be related to the general health status. QoL broadly includes aspects of physical, psychological, and social functioning. The lack of a clear definition is reflected in the many instruments that have been proposed to measure QoL, each of them based on a different quality of life concept. QoL instruments either focus on general health measurement or on disease specific wellbeing. Specific wellbeing may be more sensitive to the effect of disease, while generic tests are more suitable for making comparisons with other diseases.

Clinicians who deal with OME frequently encounter spontaneous comments from parents after their child has had ventilation tubes inserted. They record events such as improved sleep patterns, improved awareness, and a less irritable and much happier child. In a study by Rosenfeld and Bhaya, parents reported fewer physical symptoms and less emotional distress in their children. A limitation of these studies, however, is that they lacked a control group for comparison. Another problem is that only children of 3 years or older were included. Nowadays, OME is detected at an earlier age and ventilation tubes are being inserted at a younger age. Moreover, the symptoms and rates of development are different for infants and preschool children; this might also influence the quality of life in different ways.

We therefore investigated the effect of treatment with ventilation tubes on the quality of life of infants (1–2 years of age) with persistent bilateral OME. We performed a multicentre randomised controlled trial in which we compared treatment with ventilation tubes to a period of watchful waiting.

Patients and methods

Patients

A cohort of 30 099 children born in the eastern part of the Netherlands between 1 January 1996 and 1 April 1997 were invited for routine hearing screening at the age of 9 months. For the purpose of the trial, those who failed three successive tests were referred to one of the 13 participating ear, nose and throat (ENT) outpatient clinics for diagnosis and follow up (n = 1081). The parents of infants diagnosed with persistent (four to six months) bilateral OME (confirmed by tympanometry and otoscopy) by the ENT surgeon during subsequent observations (n = 386) were invited to enter their child into our randomised controlled trial. The children for whom informed consent was obtained were randomly allocated to one of two
groups: treatment with ventilation tubes (VT group; n = 93), or a period of watchful waiting (WW group; n = 94). QoL was measured at baseline and at six and 12 months follow up by means of questionnaires and scoring of standard video recordings.

We obtained approval from all ethical committees of the 13 participating hospitals.

METHODS
A balanced allocation programme was employed to increase comparability at baseline,\textsuperscript{12} with 187 children balanced on five prognostic factors: sex, age, season at randomisation, educational level of the mother, and hospital.

To measure quality of life, we used the TAIQOL (TNO-AZL Infant Quality of Life) questionnaire.\textsuperscript{13} As parent–child interaction can be seen as a domain specific part of quality of life, we also measured this interaction by means of the Erickson scales.\textsuperscript{14} The TAIQOL questionnaire was developed for children of 1–4 years and is filled in by one parent. The questionnaire contains 46 items, covering 13 domains (lungs, stomach, skin, sleeping, appetite, eating problems, aggressive behaviour, positive emotions, emotions of panic, vitality, social behaviour, motoric problems, and communication). We did not expect three of the 13 domains (lungs, stomach, skin) to be related to OME, so excluded them from the analysis after the results within these domains did not change over time. In addition, we excluded the domain on positive emotions as this scale did not show any change. In 19 of the 46 items, we assessed only the occurrence of specific complaints, while the other 27 items form a combination of the occurrence of specific complaints and the emotional response of the parents to the complaint. In the analysis, we separated the two sets of items; the emotional response was only studied when a difference in occurrence was found. The occurrence was for almost all domains rated on a 12 point scale (the higher the score the more complaints). Two domains (motoric problems and aggressive behaviour) were not rated on a 12 point scale, but were adjusted to the other scales.

Parent–child interaction was measured with the Erickson scales. Each parent–child pair was videotaped in a structured play setting for 13 minutes in a room at the hospital. Three toys were placed in the room, representing intellectual, physical, and social areas of development. At 0, 6, and 12 months follow up, different toys were used: set A consisted of a jigsaw puzzle with five pieces, a Bo-peep farm set, and three different books; set B consisted of a jigsaw puzzle with 14 pieces, a feed the animal game, and three different books; set C consisted of a jigsaw puzzle with 14 pieces, a surprise key car game, and three different books. The parents were told that the investigator was interested in how the parent and child played together. The parents were asked to start playing with the jigsaw puzzle (for three minutes), followed by the game (for five minutes), and the books (for five minutes). The investigator then left the room and knocked on the door after three, five, and five minutes to let the parents know when to change toys.

From the tapes five child scales (negativism, compliance, affection for the mother, avoidance of the mother, and reliance on the mother) and five parent scales (supportive presence, respects autonomy, structure and limits, quality of instruction, and hostility) were scored on a seven point scale (1 = low interaction to 7 = high interaction).\textsuperscript{14} The videotapes were scored by two psychological test assistants. The interobserver reliability of the scoring was high ($r > 0.8$).

OME was defined according to the MOMES protocol,\textsuperscript{15} which is primarily based on tympanometry. If tympanometry could not be performed, or if the tympanograms were doubtful (type C2), otoscopy was used to make the diagnosis.

Hearing was assessed with a portable visual reinforcement audiometry set. It appeared to be a good alternative for standard visual reinforcement audiometry in multicentre trials. A more extensive description is given elsewhere.\textsuperscript{16}

The calculated power for this trial was primarily based on the outcomes on language development (which are described elsewhere).\textsuperscript{17} However, recalculating the power, in this trial, with 187 randomised children, we would have been able to distinguish a difference of 0.5 (on a 12 point scale) in improvement on the TAIQOL between the two groups.

STATISTICAL ANALYSIS
All analyses were performed according to the intention to treat principle. Differences in domains at 0, 6, and 12 months follow up between the groups were tested with the Wilcoxon test as the scores were not normally distributed. To adjust for potential baseline differences we also studied differences in scores between six months follow up and baseline, and as well as differences between 12 months follow up and baseline.

To adjust for potential confounders and to study possible effect modifiers, we constructed logistic models, with the dichotomised scores based on improvement (children who improved versus the others) at six months and at 12 months follow up as dependent variable. Potential confounders were: educational level of the mother (high/middle/low), older siblings (yes/no), attending day care (yes/no), sex (female/male), and adenoidectomy before randomisation (yes/no). These variables were also included as potential effect modifiers. In addition hearing level at randomisation (dB) was tested for interaction with treatment. By studying these possible interactions we tried to identify subgroups of children that may experience greater benefit from treatment with ventilation tubes.

A multivariate analysis of variance (MANOVA) was performed to detect a treatment effect for all domains combined; for this analysis we modelled the improvement scores at six and 12 months follow up.
Results

A total of 187 children were randomised: 93 to the VT group and 94 to the WW group. Nineteen children dropped out immediately after randomisation (15 from the WW group and four from the VT group). These children were excluded from the trial and so have not been included in the 187 randomised children. Another 11 children dropped out during the remainder of the follow up (eight children from the WW group and three from the VT group). In addition, 10 children from the WW group received treatment with ventilation tubes during the trial period.

The mean age of the children at randomisation was 19.5 months in the VT group and 19.4 months in the WW group. Mean hearing loss (measured over 500, 1000, 2000, and 4000 Hz) in the best ear at randomisation was 46.4 dB (95% confidence interval (CI) 44.2 to 48.6) in the VT group and 43.4 dB (95% CI 41.0 to 45.8) in the WW group. At six months follow up, the improvement in hearing level in the VT group was 10.2 dB versus 4.6 dB in the WW group; at 12 months follow up, these values were 13.1 dB and 8.5 dB, respectively.

At 3, 6, 9, and 12 months follow up, 14.6%, 29.3%, 26.9%, and 26.6% of the children in the VT group were diagnosed as having bilateral OME, respectively. In the WW group, these percentages were 77.2%, 65.9%, 57.3%, and 53.2%, respectively.

The mean scores for the nine subdomains of the TAIQOL at baseline, six, and 12 months follow up are presented in Table 1.

### Table 1: Mean scores for the nine subdomains of the TAIQOL at baseline, six, and 12 months follow up

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Vitality</th>
<th>Appetite</th>
<th>Communication</th>
<th>Motoric</th>
<th>Social</th>
<th>Anxiety</th>
<th>Aggression</th>
<th>Eating</th>
<th>Sleeping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline VT</td>
<td>93</td>
<td>3.3 (0.9)</td>
<td>4.7 (1.5)</td>
<td>6.8 (2.3)</td>
<td>5.8 (2.5)</td>
<td>3.6 (0.9)</td>
<td>4.1 (1.2)</td>
<td>11.3 (2.2)</td>
<td>3.4 (0.7)</td>
<td>7.1 (2.2)</td>
</tr>
<tr>
<td>Baseline WW</td>
<td>91</td>
<td>3.3 (0.9)</td>
<td>4.4 (1.4)</td>
<td>6.4 (2.0)</td>
<td>6.1 (2.8)</td>
<td>3.5 (0.8)</td>
<td>4.0 (1.1)</td>
<td>10.9 (2.2)</td>
<td>3.4 (0.6)</td>
<td>6.8 (2.1)</td>
</tr>
<tr>
<td>6 months follow up VT</td>
<td>87</td>
<td>3.3 (0.9)</td>
<td>5.0 (1.4)</td>
<td>6.7 (2.3)</td>
<td>4.4 (0.9)</td>
<td>3.5 (0.9)</td>
<td>4.3 (1.1)</td>
<td>11.9 (2.4)</td>
<td>3.3 (0.6)</td>
<td>6.8 (2.1)</td>
</tr>
<tr>
<td>6 months follow up WW</td>
<td>89</td>
<td>3.3 (1.0)</td>
<td>4.7 (1.6)</td>
<td>5.8 (2.1)</td>
<td>4.4 (1.1)</td>
<td>3.5 (0.9)</td>
<td>4.1 (1.0)</td>
<td>11.1 (2.0)</td>
<td>3.5 (0.8)</td>
<td>6.6 (1.9)</td>
</tr>
<tr>
<td>12 months follow up VT</td>
<td>84</td>
<td>3.1 (0.8)</td>
<td>5.3 (1.6)</td>
<td>5.9 (2.0)</td>
<td>4.2 (0.8)</td>
<td>3.5 (0.9)</td>
<td>4.6 (1.3)</td>
<td>11.8 (2.4)</td>
<td>3.3 (0.5)</td>
<td>6.4 (2.2)</td>
</tr>
<tr>
<td>12 months follow up WW</td>
<td>81</td>
<td>3.2 (0.8)</td>
<td>4.9 (1.4)</td>
<td>5.6 (1.9)</td>
<td>4.2 (1.0)</td>
<td>3.5 (0.9)</td>
<td>4.3 (1.1)</td>
<td>11.5 (2.0)</td>
<td>3.4 (0.6)</td>
<td>6.4 (1.9)</td>
</tr>
</tbody>
</table>

Results expressed as mean (SEM).

In the logistic models on the probability of improving in a specific domain at six or 12 months, we were able to adjust for possible confounders and to identify effect modifiers. No significant effect of treatment was found. In most domains only the score at baseline was a confounder that needed to be adjusted for.

The MANOVA on all the domains combined did not show any difference in treatment effect between the VT group and the WW group at six months or at 12 months follow up (p = 0.22 and p = 0.94 with the Hotelling–Lawley Trace test).

### Erickson scales

Tables 2 and 3 show the mean scores of the Erickson scales. Figure 2 shows the differences in child and parent scores on the Erickson scales between 12 and 0 months follow up. Some slight differences were found in favour of the WW group in most domains, but there was a slight difference in “compliance” and “hostility” in favour of the children VT group. None of these differences could be distinguished from chance. However, when the absolute mean scores at six and 12 months follow up were compared, some slight but significant differences were found in favour of the WW group, especially on the parent scales.

The logistic models on the probability of improving enabled confounder adjustment and a search for effect modification. We did not find any differences between the two groups. No association was found between improvement on all the domains combined and treatment (MANOVA) at six or 12 months follow up (p = 0.19 and p = 0.38 for the child scales, versus p = 0.21 and p = 0.95 for the parent scales, respectively).

### Maximum contrast

Spontaneous resolution of OME in the WW group and recurrence of OME in the VT group might have obscured the contrast between the two groups. To test our aetiological model...

[Figure 1: Difference scores on the subdomains of the TAIQOL between 12 and 0 months follow up. *Because of deviation in this scale (maximum score 16 instead of 12), we adjusted this difference score according to the other scales. †Because of deviation of this scale (maximum score 21 instead of 12), we adjusted this difference score according to the other scales. ‡A positive difference indicates an increase in complaints; a negative difference indicates a decrease. (Note that the difference scores between 6 and 0 months follow up did show similar results).]
(hearing loss caused by OME impairs quality of life, while functional ventilation tubes improve quality of life) we also performed an explanatory analysis in which we compared the children with effusion during the whole follow up period (n = 28) to those who were effusion free during follow up (n = 54). We did not find any difference in the 13 domains of the TAIQOL, while on the Erickson scales, there was a difference in only one (negativism) of the 10 domains. For negativism, we found greater increases in parent–child interaction in the children without effusion during follow up (these children showed less anger, dislike, or hostility towards their mother).

### Discussion

In contrast to earlier studies, our results did not show that ventilation tubes improve the quality of life of young children with persistent OME. These previous studies were uncontrolled, hospital based, and focused on older children, whereas our study was controlled, population based, and the children were 1–2 years of age. In addition, we used a generic QoL test, while Facione and Rosenfeld and Bhaya asked OME specific questions. Although disease specific QoL tests might be more sensitive, several domains of the TAIQOL and Erickson scales (for example, sleeping pattern, eating problems, avoidance) were a priori relevant for OME.

Formally, the Erickson scales do not measure quality of life, but parent–child interaction can be seen as a domain specific part of quality of life. The Erickson scales are well validated and more objective than the self completed questionnaires and are therefore suitable as an additional test.

No population norm is available for the TAIQOL. Therefore, mean scores cannot be compared to the equivalent scores of children without persistent OME. Some comparisons are possible with the Erickson scales. In our study, the parent–child interaction scores in the children with persistent OME were slightly poorer than those reported by Drenthen and Ricksen-Walraven in children without disease.

It is possible that OME does not have any great impact on QoL; we found no substantial differences between the children with effusion

---

**Table 2** Mean scores for the five subdomains of the Erickson (child) scales at baseline, six, and 12 months follow up

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Affection</th>
<th>Avoidance</th>
<th>Compliance</th>
<th>Negativism</th>
<th>Reliance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline VT</td>
<td>93</td>
<td>4.5 (1.4)</td>
<td>5.9 (1.5)</td>
<td>4.6 (1.6)</td>
<td>6.2 (1.5)</td>
<td>6.5 (0.8)</td>
</tr>
<tr>
<td>WW</td>
<td>91</td>
<td>4.7 (1.4)</td>
<td>6.2 (1.4)</td>
<td>5.1 (1.4)</td>
<td>6.5 (1.2)</td>
<td>6.5 (1.0)</td>
</tr>
<tr>
<td>6 months follow up VT</td>
<td>87</td>
<td>4.4 (1.3)</td>
<td>6.3 (1.1)</td>
<td>5.1 (1.1)</td>
<td>6.6 (1.0)</td>
<td>6.5 (0.9)</td>
</tr>
<tr>
<td>WW</td>
<td>89</td>
<td>4.6 (1.1)</td>
<td>6.5 (1.2)</td>
<td>5.2 (1.3)</td>
<td>6.7 (0.8)</td>
<td>6.7 (0.7)</td>
</tr>
<tr>
<td>12 months follow up VT</td>
<td>84</td>
<td>4.5 (1.4)</td>
<td>6.5 (0.9)</td>
<td>5.2 (1.1)</td>
<td>6.6 (1.1)</td>
<td>6.6 (0.7)</td>
</tr>
<tr>
<td>WW</td>
<td>81</td>
<td>4.9 (1.1)</td>
<td>6.9 (0.4)</td>
<td>5.6 (1.0)</td>
<td>6.9 (1.1)</td>
<td>6.8 (0.4)</td>
</tr>
</tbody>
</table>

Results expressed as mean (SEM).

**Table 3** Mean scores for the five subdomains of the Erickson (parent) scales at baseline, six, and 12 months follow up

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Hostility</th>
<th>Structure</th>
<th>Respect</th>
<th>Supportive</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline VT</td>
<td>93</td>
<td>6.9 (0.7)</td>
<td>4.6 (1.4)</td>
<td>4.9 (1.4)</td>
<td>4.9 (1.5)</td>
<td>4.5 (1.6)</td>
</tr>
<tr>
<td>WW</td>
<td>91</td>
<td>6.9 (0.3)</td>
<td>5.1 (1.2)</td>
<td>5.3 (1.3)</td>
<td>5.4 (1.3)</td>
<td>5.0 (1.4)</td>
</tr>
<tr>
<td>6 months follow up VT</td>
<td>87</td>
<td>7.0 (0.2)</td>
<td>4.8 (1.5)</td>
<td>4.9 (1.3)</td>
<td>4.9 (1.4)</td>
<td>4.4 (1.5)</td>
</tr>
<tr>
<td>WW</td>
<td>89</td>
<td>6.9 (0.5)</td>
<td>5.2 (1.2)</td>
<td>5.4 (1.3)</td>
<td>5.4 (1.4)</td>
<td>5.2 (1.3)</td>
</tr>
<tr>
<td>12 months follow up VT</td>
<td>84</td>
<td>7.0 (0.2)</td>
<td>4.8 (1.4)</td>
<td>5.0 (1.3)</td>
<td>5.0 (1.4)</td>
<td>4.7 (1.4)</td>
</tr>
<tr>
<td>WW</td>
<td>81</td>
<td>7.0 (0.2)</td>
<td>5.5 (1.0)</td>
<td>5.3 (1.2)</td>
<td>5.5 (1.2)</td>
<td>5.3 (1.3)</td>
</tr>
</tbody>
</table>

Results expressed as mean (SEM).
The effect of ventilation tubes (grommets) on quality of life

We tried to identify such subgroups in this study, but found no relevant differences. The power of studying such subgroups was low and type II errors can therefore not be excluded. Further research is needed to study possible subgroups as well as to investigate OME specific questionnaires that can signal small but relevant effects.

We thank the parents and children who took part in this study, the ENT surgeons who provided medical management, Joost Engel for his valuable suggestions concerning this paper, and Bonia von Oosterhout for trial management and data entry. This project was funded by the Dutch Investigative Medicine Fund of the National Health Insurance Board. ENT surgeons were: J van Leeuwen, Rinjsante Hospital, Arnhem; G Plummers, H Vencker, J Mol, and A Frima van Aarem, Reinland Hospital, Amersfoort; J Engel and SJ de Vries, Canisius Wilhelmina Hospital, Nijmegen; K Ingels, University Hospital St Radboud, Nijmegen; M Nis, F Gerritsma, and P van den Van, Slingeland Hospital, Doetinchem; SJ Retema, Hospital Sildere Valen, Ede; E Teurissen and G van de Meerakker, Carolus Laduina Hospital, ’s Hertogenbosh; F Kwadijk and A de Voschier, GGZ, ’s Hertogenbosch; H Cats, Willem Alexander Hospital, ’s Hertogenbosch; T Brugman and S Holstaf, Hospital Rivieren- land, Tiel; G Peeters and D Elprana, St Maartensgasthuis, Venlo; J Alvarado van Os, Beatris Hospital, Winterwijk; and F Mud, Ch Sempjez, Hospital Het Nieuwe Spitaal, Zutphen.

10 Rosenfeld RM, Bhaya M. Impact of tympanostomy tubes on child quality of life. 7th International symposium on recent advances in otitis media. 97. 1999.
13 TNO Prevention and Health, Paediatric Department, and University of Leiden. The TNO-ADL Infant Quality of Life Questionnaire (in Dutch). Leiden: TNO, 1996.