Appendix B *Sample size calculation*

The required sample size of 360 patients (180 patients per group) was based on the following assumptions:

1) A clinical response (a clinically relevant improvement of at least 4 points[1] of 50% in the intervention group versus 30% in the control group[2 3] (implying an effect size *d* = 0.42 for the clinical response), and a power of 80% to detect a difference of the primary outcome between the intervention and control group with a two-tailed alpha of 5%. This assumption gave a sample size of 180 patients in total (90 patients per group), ignoring at first the design effect due to clustering of patients within physicians.
2) The number of participating GPs was about twice as large as the number of pulmonologists.
3) An estimated availability of 5 patients per GP and 8 patients per pulmonologist on average. This, together with assumptions 1 and 2, gave a total of 20 GPs and 10 pulmonologists. However, the following three steps (4-6) resulted in a sample size which was twice as large, that is 40 GPs and 20 pulmonologists.
4) An intraclass correlation coefficient (ICC) of 0.05, meaning that about 5% of the total outcome variation within each group is between GPs and between pulmonologists, instead of between patients of the same physician. Literature suggested that an ICC of 0.05 was a good default value for trials in primary care.[4-6] Combined with assumptions 2 and 3, and allowing for 10% more clusters (healthcare providers) to compensate the power loss due to variation in cluster size, that is, in number of patients included per healthcare provider, this ICC of 0.05 implied a design effect of 1.38.[7] The number of clusters was thus multiplied with 1.38.
5) A dropout rate of 25% of patients and/or clusters, was compensated by multiplying the number of clusters to be included by 1.33 (since 75% of 1.33 is 1). Dropouts were included into the analyses (intention to treat), but contributed less to the power due to missing data, hence the present correction.
6) Data analysis of the primary outcome with the recommended PQL2 (penalized quasi-likelihood) estimation method which required a further multiplication of the number of clusters with a factor of 1.10.[8]

Combining assumptions 4, 5 and 6 gave a multiplication factor of 1.38 * 1.33 * 1.10 = 2 for the number of GPs and pulmonologists as computed in steps 1 to 3, leading to the planned sample size of 40 GPs, 20 pulmonologists and 360 patients in total.

References

1. Jones PW. St. George's Respiratory Questionnaire: MCID. *Copd* 2005; 2: 75-79.
2. Tashkin D, Kesten S. Long-term treatment benefits with tiotropium in COPD patients with and without short-term bronchodilator responses. *Chest* 2003; 123: 1441-1449.
3. Casaburi R, Mahler DA, Jones PW, *et al*. A long-term evaluation of once-daily inhaled tiotropium in chronic obstructive pulmonary disease. *Eur Respir J* 2002; 19: 217-224.
4. Adams G, Gulliford MC, Ukoumunne OC, *et al*. Patterns of intra-cluster correlation from primary care research to inform study design and analysis. *J Clin Epidemiol* 2004; 57: 785-794.
5. Eldridge SM, Ashby D, Feder GS, *et al*. Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. *Clin Trials* 2004; 1: 80-90.

6. van Breukelen GJ, Candel MJ. Efficient design of cluster randomized and multicentre trials with unknown intraclass correlation. *Stat Methods Med Res* 2011:Epub ahead of print.
7. van Breukelen GJ, Candel MJ. Calculating sample sizes for cluster randomized trials: we can keep it simple and efficient! *J Clin Epidemiol* 2012; 65: 1212-1218.
8. Candel MJ, Van Breukelen GJ. Sample size adjustments for varying cluster sizes in cluster randomized trials with binary outcomes analyzed with second-order PQL mixed logistic regression. *Stat Med* 2010; 29: 1488-1501.