**Day 2, part 5, 11:30 - 12:30 (1 hr), Group assignment**

**Exploring heterogeneity: statistical heterogeneity measures**

Zaman et al. performed a study to investigate fiber and prebiotic supplementation of enteral nutrition (EN) for (amongst others) the risk of diarrhea.

Inclusion criteria for this review were:

(1) primary research of randomized controlled trials (RCT), non-RCT studies, and observational cohort study designs;

(2) studies conducted on adult patients of any health or nutritional status receiving EN;   
(3) studies assessing effects of fiber in EN on diarrhoea;

(4) studies conducted from January 1990 to June 2014.

Exclusion criteria included studies that:

(1) did not use enteral formula as the sole or main source of nutrients, either orally or through a tube;

(2) involved supplementation of synbiotics (prebiotics and probiotics) in the enteral formula;

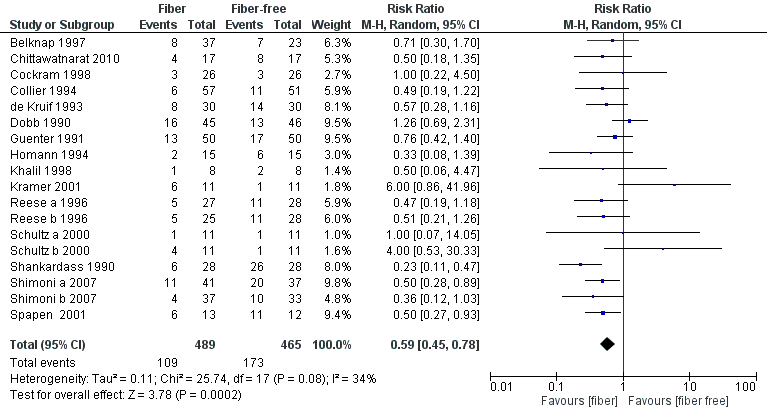
(3) involved animal or in vitro experiments or were case control or cross-sectional studies, review articles or dissertations.

The following data were extracted:

* population descriptions (location, inclusion and exclusion criteria, method of recruitment, and consent),
* methodology (aim, design, study duration, and ethical approval),
* risk of bias assessment,
* participants (number of randomized, withdrawals and exclusions, and characteristics of the
* study participants),
* interventions (timing and delivery of EN, formula used, and fiber dosage and type),
* diarrhoea incidence

Data selected from: Zaman MK, Chin KF, Rai V, Majid HA. Fiber and prebiotic supplementation in enteral nutrition: A systematic review and meta-analysis. World Journal of Gastroenterology: WJG. 2015 May 7;21(17):5372.

1. The authors included 18 studies, and conducted a random-effects meta-analysis, to compare the risk of diarrhoea in the groups with and without fiber supplementation. As an outcome parameter they used the Risk Ratio (RR).   
   One of the studies (Belknap 1997) treated 37 patients with fiber and 23 patients without fiber. Of the 37 fiber treated patients, 8 patients suffered from diarrhoea and of the 23 other patients, 7 suffered from diarrhoea.   
   Can you calculate the risk of diarrhoea for each group, and the RR for this study?
2. The forest plot below shows the results for all 18 studies.  
   The meta-analysis resulted in the following:   
   Risk Ratio (RR) = 0.59; 95% CI: 0.45-0.78; P =0.0002).   
   What can we conclude from this meta-analysis? Is the supplementation of the EN with fiber useful to prevent diarrhoea ? Give an interpretation of the Risk Ratio, the confidence interval, and the p-value.



**Figure 1: forest plot for the overall effect of fiber supplementation in enteral nutrition on   
the incidence of diarrhoea**

1. Eighteen studies were included that fulfilled the selection criteria. Which aspects of the review may be important sources for the heterogeneity? (Hint: Check the in- and exclusion criteria for the studies.)

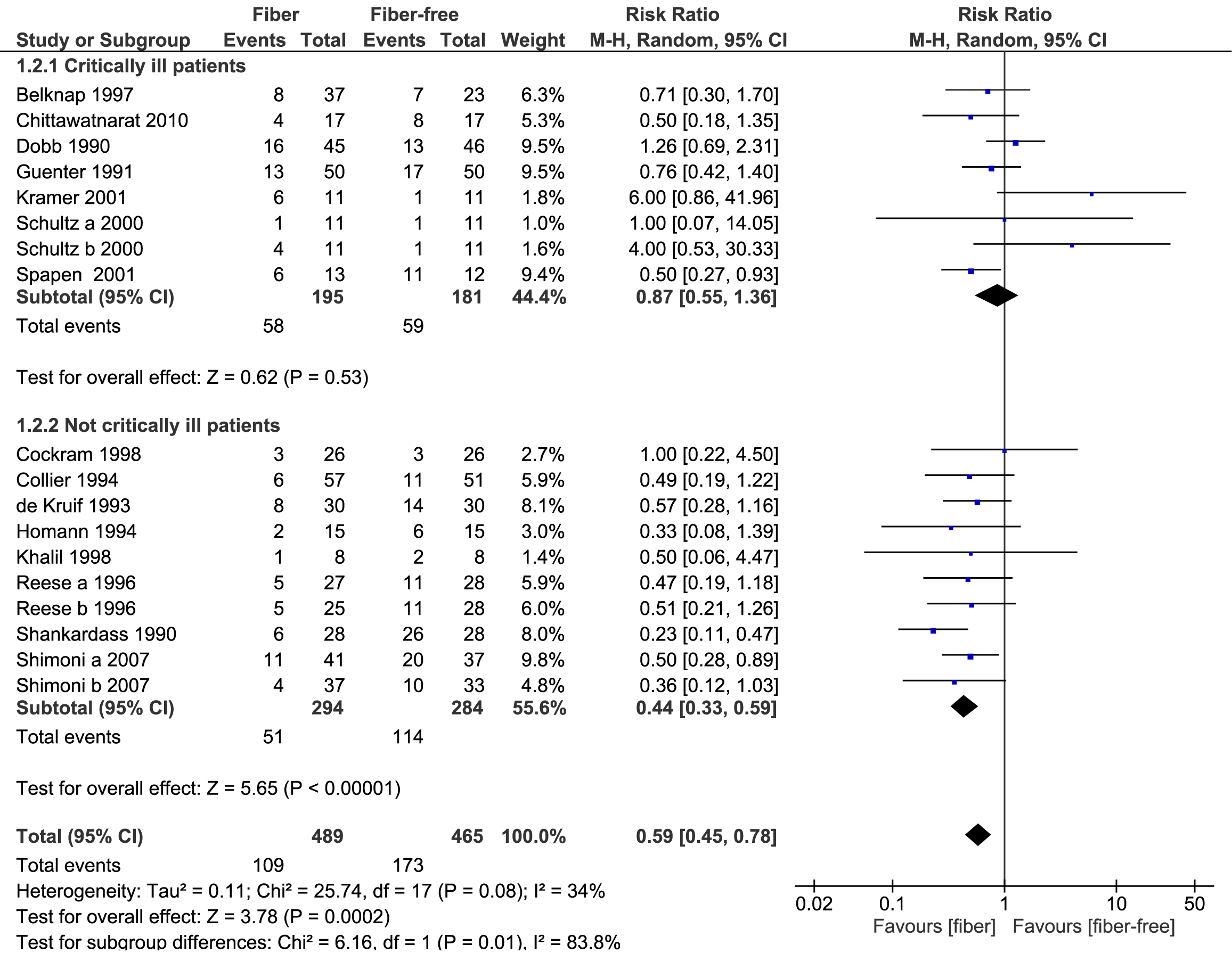
Check the in- and exclusion criteria for the studies. 🡪   
- Variation in trial designs (observational and RCTs)

* - Inclusion period: 1990-2015, i.e. using a 25 year period of inclusion for studies  
  - Patient population: any health condition and nutritional status allowed

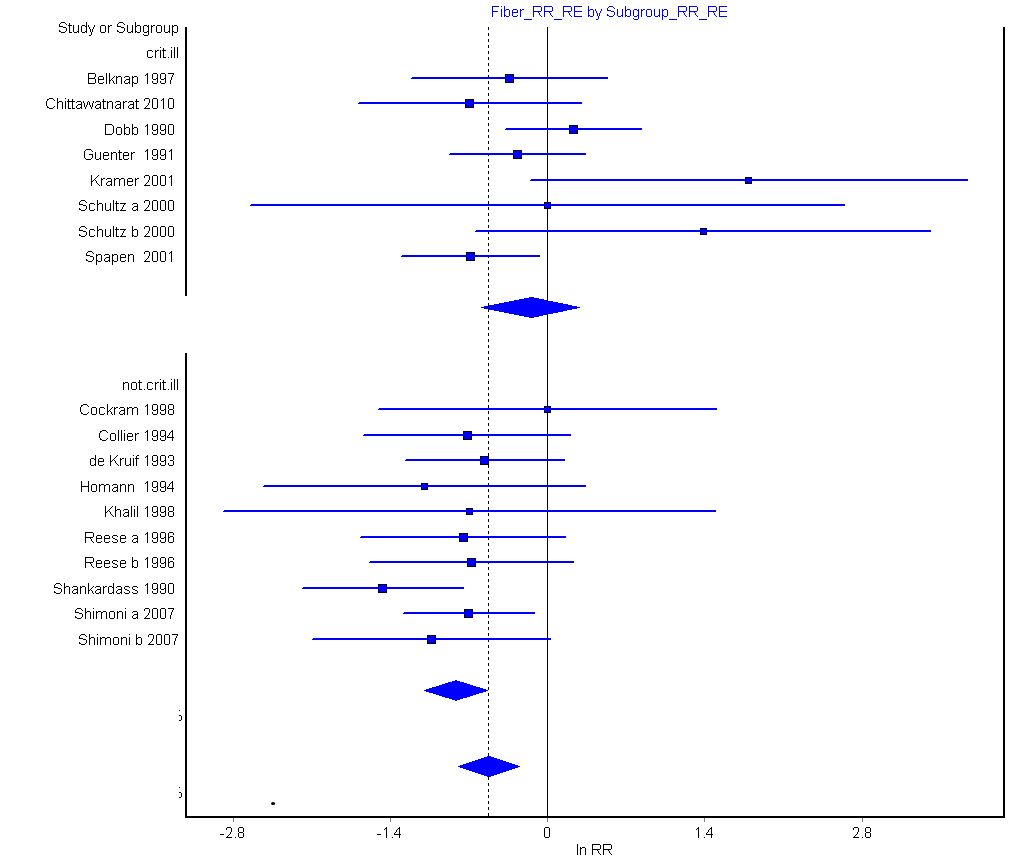
1. The forest plot shows the following results for the between-study heterogeneity : I2 of 34% , a Cochran’s Q of 25.74 (denoted as Chi2), a tau2 of 0.11, and a p-value for heterogeneity of 0.08.   
   In total this adds to four different measures to quantify the heterogeneity. Which one do you prefer? Why? Are there also disadvantages connected to this measure?   
   Also think of (dis-)advantages of the other measures.
2. Do you think that the confidence interval of the overall effect also reflects the between study heterogeneity?

|  |  |  |
| --- | --- | --- |
| **Measure** | **Advantage** | **Disadvantage** |
| **I2** | + I2 presents the inconsistency between the study results and quantifies the proportion of observed dispersion that is real, i.e. due to between-study differences and not due to random error.  + I2 represents the inconsistency always on a scale between 0 and 100%, therefore it might be compared with suggested limits for low or high inconsistency  + I2 reflects the lack of overlap of the confidence intervals of the study-effects  + does not depend on the number of studies included + not sensitive to the metric of the effect size | -A direct clinical interpretation of I2 is difficult.  - I2 is ambiguous because its size depends on sample size:  -with very large studies, even tiny between-study differences in effect size may result in a high I2,  -with small (imprecise) studies, very different treatment effects can yield an I2 of 0  🡪 Thresholds for interpretation can be misleading,  -whether it is serious heterogeneity also depends on the reasons for the heterogeneity |
| **τ2**(**Tau2)** | + τ2 is the estimate of the between-study variation and therefore useful in calculations + τ(the square root of τ2) is the standard deviation of the between-study variation, and therefore on the scale of the original outcome, useful to quantify the variation in individual study results + Level of τ is not sensitive to the number of studies | - a direct clinical interpretation based on τ2 is difficult, especially when τ2 belongs to outcomes that were analyzed on a transformed scale, e.g. log odds ratios  -When the τ2 estimate is based on only a few studies it will be imprecise |
| **Cochran’s Q** | + not sensitive to the metric of the effect size index | If you use it to statistically test the heterogeneity, there are issues with sample size and interpretation 🡪 low power if sample size is too small & always significant is sample size is large |
| **p-value (from test on Q)** | + Only one way interpretable  + not sensitive to the metric of the effect size index | -Issues with sample size 🡪 low power if sample size is too small & always significant if sample size is large |
| **Confidence interval (CI) of the RR** | * The CI in a random effects model contains highly probable values for the summary (mean) treatment effect this is **not** a heterogeneity measure! | * The CI gives no information on the range of true treatment effects that are likely to be seen in other settings, e.g. in the next study or in the patients a clinician wants to treat in her clinic |

1. The authors did several subgroup analyses, but did not prespecify these analyses. What is your opinion on this approach?   
     
   This approach has a high probability to lead to incorrect conclusions. The analysis should be pre-specified in order to prevent data dredging. The more subgroups you evaluate, the higher the probability that you will find somewhere a significant result, just by chance, which cannot be replicated in a new meta-analysis (suppose you would have 18 new studies).
2. Below you see the random-effects meta-analysis of the effect of fiber supplementation in enteral nutrition on the incidence of diarrhea (Figure 2). The analysis is presented by subgroup: studies with critically ill patients versus studies with not-critically ill patients. Do you think this a useful way to subgroup the studies?   
     
   We removed the estimates of the between study heterogeneity per subgroup.   
   What do you think of the heterogeneity in both subgroups? Is it increased or decreased compared to the analysis on all studies (i.e. compared to the forest plot above)?   
   Also consider the similar forest plot, but now on log-scale, i.e. the scale on which the analysis is actually performed (Figure 3).



**Figure 2: forest plot for the effect of fiber supplementation in enteral nutrition on   
the incidence of diarrhea by subgroup and overall**



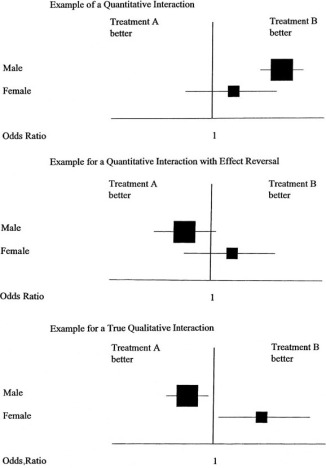
**Figure 3: forest plot for the effect of fiber supplementation in enteral nutrition on   
the incidence of diarrhea by subgroup and overall, on log-scale**

For the critically ill group there seems to be quite some heterogeneity, i.e. there is quite some variation in the size and direction of the effects. This will result in relatively high Cochran’s Q, I2, Tau2 and a low p-value).

- Heterogeneity: Tau² = 0.16; Chi² = 12.50, df = 7 (P = 0.09); I² = 44%

For the non-critically ill group there seems to be less heterogeneity, all estimated effects tend to be in favour of the treatment, and there is not so much variation in the effect sizes (so lower Cochran’s Q, I2, Tau2 and not so small p-value).  
Heterogeneity: Tau² = 0.00; Chi² = 5.43, df = 9 (P = 0.79); I² = 0%

1. The results in Figure 2 for the overall effect are exactly equal to the results above (Figure 1). However, the forest plot shows that the subgroup effect is statistically significant.
   1. What does this mean?  
        
      This means that the treatment effect is not the same in both subgroups, in one subgroup the treatment effect is stronger than in the other group.
   2. It is useful to distinguish between the notions of ‘qualitative interaction’ and ‘quantitative interaction’ (Yusuf 1991). Qualitative interaction exists if the direction of effect is reversed, that is if an intervention is beneficial in one subgroup but is harmful in another.   
        
      Qualitative interaction is rare and may be used as an argument that the most appropriate result of a meta-analysis is the overall effect across all subgroups.   
        
      Quantitative interaction exists when the size of the effect varies but not the direction, that is if an intervention is beneficial to different degrees in different subgroups.  
        
      Do you think that we see quantitative or qualitative heterogeneity?  
        
      Here we see an example of quantitative heterogeneity.

**[](http://www.sciencedirect.com/science/article/pii/S0002870300571269#gr1)**