

## NNT Literature Review

By Jennifer Gilbert

Questions addressed in the literature review that follows:

- A. Is the paper examined in the [Akl et al.](#) systematic review? If not, why not?
- B. Is the paper examined in the review by [Zipkin et al.](#)? If not why not?
- C. Does the study objectively assess s comprehension of NNT? How? Does it do so compared to other measures of relative risk?
- D. What are the findings regarding comprehension of NNT and other measures of relative risk, if any?

(Note that Austin examined Akl et al. and Zipkin et al. [in a prior post.](#))

The chart below summarizes findings based on studies suggested in [a post by Hilda Bastian](#) and other studies discussed in [prior posts](#) by [Austin](#) and [Hilda](#). Depending on how one wants to count three studies (Christensen 2003, Gyrd-Hansen 2011, and Cuite 2008—see footnotes of the chart) then out of four to seven studies, two found NNTs no more difficult to comprehend. Many more studies are just not relevant to the question of comprehension, as explained below.

Paper	Examined by Akl and/or Zipkin?	Objectively tested comprehension of NNT relative to other metrics (not just self-reports)?	Found NNT is more difficult to comprehend
<a href="#">Lyndal Trevena et al. (2013)</a>	No	No	
<a href="#">Selinger (2013)</a>	No	No	
<a href="#">Sheridan (2002)</a>	No	Yes	Yes
<a href="#">Christensen (2003)</a>	No	Yes <sup>(b)</sup>	Yes <sup>(b)</sup>
<a href="#">Gyrd-Hansen (2011)</a>	No	Yes <sup>(c)</sup>	Yes <sup>(c)</sup>
<a href="#">Dahm et al. (2009)<sup>(a)</sup></a>	No	Yes	No
<a href="#">Chao (2003)</a>	Yes (Only Akl)	No	
<a href="#">Kristiansen (2002)</a>	Yes (Only Akl)	No	
<a href="#">Halvorsen (2007)</a>	Yes (Only Zipkin)	No	
<a href="#">Stovring (2008)</a>	Yes (Only Zipkin)	No	
<a href="#">Halvorsen (2005)</a>	Yes (Only Zipkin)	No	
<a href="#">Berry et al. (2006)<sup>(a)</sup></a>	Yes (Only Zipkin)	Yes	No
<a href="#">Cuite (2008)</a>	Yes (Only Zipkin) <sup>(d)</sup>	Yes <sup>(d)</sup>	Yes <sup>(d)</sup>
<a href="#">Grimes (1999)</a>	Yes (Both)	No	
<a href="#">Carling (2009)</a>	Yes (Both)	No	
<a href="#">Sheridan et al. (2003)<sup>(a)</sup></a>	Yes (Both)	Yes	Yes

(a) Included in prior discussion (see text).  
 (b) The other metric was postponement of hip fracture, not a generalizable measure of risk.  
 (c) The study measured sensitivity to risk information, which is not exactly the same as comprehension.  
 (d) This study examined “1 in n” which is awfully close to NNTs. Even though it’s mentioned in Zipkin, it’s cited in that paper among the NNT studies.

The following pages include details on studies cited in Hilda’s post [here](#), excluding ones marked with footnote (a) in the chart.

[Lyndal Trevena et al. \(2013\)](#): Not included by Akl or Zipkin (post-date). An expert research group discussed eleven major issues of risk communication and came up with eleven components: 1) Presenting the chance an event will occur; 2) Presenting changes in numeric outcomes; 3) Outcome estimates for test and screening decisions; 4) Numeric estimates in context and with evaluative labels; 5) Conveying uncertainty; 6) Visual formats; 7) Tailoring estimates; 8) Formats for understanding outcomes over time; 9) Narrative methods for conveying the chance of an event; 10) Important skills for understanding numerical estimates; and 11) Interactive web-based formats.

**Does not explicitly mention NNT.**

Hilda organized other papers in four batches, each of which is examined below.

Batch 1, with key findings related to interpretation/understanding of risk:

1. [Grimes \(1999\)](#): Included by both Akl and Zipkin. Assessed comprehensions of rates versus proportions by having women identify the larger of two risks. Women understood rates better than proportions.  
**Does not explicitly mention NNT.**
2. [Cuite \(2008\)](#): Included by Zipkin. Visitors to a cancer-related webpage were recruited and tested on their understanding of risk when communicated by percentages, frequencies, and 1 in n (which is NNT, basically, but Zipkin didn't list this study among those that examined NNTs). Many magnitudes of risk were tested in these three formats. **Percentage and frequency questions had higher accuracy rates than the 1 in n format (57% vs 55% vs 45% respectively).** This had a 36.1% response rate and was online, so there might be some selection bias.
3. [Carling \(2009\)](#): Included by both Akl and Zipkin. Adult volunteers rated the importance of outcomes with visual analogue scales (VAS) and regressions were used to associate the participants' RIS with their choice. Showing the benefits of taking statins as a relative risk reduction increased the probability of people accepting treatment compared to presenting taking statins as an absolute summary statistics. This amounts to an NNT of 5 (for every five participants given the relative risk reduction, one chose to take statins compared to other summary statistics). The study participants' preferred presentation was natural frequencies (31%) closely followed by RRR (30%). Event rates were preferred by 20%, NNT by 10%, ARR by 5%, and TNT by 3%. When asked to evaluate understanding, they reported that they understood natural frequencies best (86%), compared to 76% reporting high understanding of NNTs. **Comprehension was not specifically tested – it was self-reported.**
4. [Selinger \(2013\)](#): Not included by either (post-date for both). Maintenance therapy benefits were presented to patients with ulcerative colitis (UC) in remission by four methods: RRR, ARR, NNT, and CP (optical representation via Cates plot). Patient preference and understanding were evaluated by asking patients to state the minimum threshold near relapse prevention and colorectal cancer risk reduction that would make them adhere to medication for UC. Of 50 participants, 48% preferred data presentation by RR over CP (28%), AR (20%) and NNT (4%). 94% found RR easy to understand, better than AR (88%), CP (74%), or NNT (48%). All in all, patients preferred data presented by RR and were more likely to adhere to data by CP. **The authors note that comprehension of statistical information was not formally studied in this paper.**

Batch 2 includes studies in medical students, who are counted as lay people by the Akl review. Two studies with 265 students:

1. [Sheridan \(2002\)](#): Not included by Akl (“Does not report the necessary data”) or Zipkin (because studies of medical students, health professionals, and public health or mass media campaigns were excluded). 62 medical students were given baseline info about an imaginary disease and learned about it through RRR, ARR, NNT, or a combination of two of the three formats. Data interpretation was judged by a comparative task (stating which drug is more beneficial) or a quantitative task (calculating how much one of the drugs reduces disease risk). **Accurate quantitative interpretation was lower with the NNT format versus the other three formats (25% versus 75%).**
2. [Chao \(2003\)](#), also reported by [Studts \(2005\)](#): Included by Akl but not Zipkin (because it used students). Preclinical medical students received hypothetical scenarios regarding their mother to decide if she should undergo chemotherapy. The scenarios were presents with RRR, ARR, absolute survival benefit, or NNT. Participants who heard the info as RRRs were more likely to endorse chemotherapy survival method than other methods if it was the only one they heard. If they heard all four measures, there were no treatment decision differences, although they were higher ratings of confusion. **Comprehension was not tested.**

Batch 3 includes studies that compare NNT for one outcome with a completely different outcome (but from the same clinical trial data). Three studies, 4,890 people:

1. [Christensen \(2003\)](#): Not included by Akl (“Not a comparison of interest (NNT to avoid one hip fracture compared with the duration of postponement of hip fracture).”) nor by Zipkin. A cross-sectional RCT was used to contact people at home in Denmark for face-to-face interviews to receive information about a hypothetical intervention in terms of NNT magnitudes or different durations of hip fracture postponement. Based on NNT of 10, 50, 100, and 400, the proportions of respondents who said they would consent were 65%, 63%, and 57% respectively. Increasing NNT had no significant association with lower consent. 43% of participants said NNT were difficult to understand, 38% interpreted them incorrectly. **Postponement of hip fracture was easier for them to understand than NNT**, and they were less likely to consent when the information was presented as postponement of fracture by 1 month, 6 months, 1 year, and 4 years, with 25%, 40%, 39%, and 53% consenting respectively.
2. [Halvorsen \(2007\)](#): Not included by Akl (“Not a comparison of interest.”), included by Zipkin. A survey presented a hypothetical drug to reduce the risks of heart attacks or hip fractures from clinical trials. Respondents were assigned to a scenario with one of three different outcomes after treatment. When presented with the NNT outcome, 93% of respondents would consent to drug treatment, while 82% would when consent to treatment when presented with a large postponement for some patients, and 69% presented with short postponement for all patients consented to therapy for heart attacks. Similar rates were observed for hip fractures: 74%, 56%, and 34% respectively. Expressing the treatment risk in NNT led to higher consent rates than other postponements, suggesting the

description of the expected outcome might influence the patient's consent for interventions. **No assessment of comprehension was conducted.**

3. [Stovring \(2008\)](#): Not included by Akl (most likely post-date), included by Zipkin. RCT in Denmark randomized people between the ages of 40-59 to receive info on cardiovascular disease treatment effectiveness in terms of ARR, RRR, NNT, or POL (prolongation of life) without heart attack. ARR led to the highest concordance between the initial and final decision (94%), but was not "statistically superior" to the other formats. The initial decision was made without knowing how to interpret ARR, RRR, NNT, or POL and then the final decision was using the same measure but after being taught how it can be interpreted by the research staff. **There was no test of comprehension.**

Batch 4 includes studies that look at whether giving people NNT of different sizes affects their (hypothetical) decision making. We get a repeat study here from batch 3, plus 3 new ones, an extra 2,771 people:

1. [Kristiansen \(2002\)](#): Included by Akl but not Zipkin (why?) A sample of Danish people were interviewed in person and asked if they would consent to a hypothetical drug for a lower risk of heart attacks. The benefit was expressed in NNT and randomly set at 10, 25, 50, 100, 200, and 400. Irrespective of NNT, 80% consented, and some who did not consent misinterpreted the NNT's meaning. **However, no systematic assessment of comprehension was reported.**
2. [Halvorsen \(2005\)](#): Not included by Akl (Information presented was different in comparison groups), included by Zipkin. 1201 Norwegians selected from a representative cross-section were given scenarios of random disease combinations to be prevented, costs of drug treatment, and NNT effect size. They were interviewed about their hypothetical consent, then randomized to different interpretations of NNT and asked if they would change their original responses based on this. The consent proportion varied from 76%-67% when the NNT was 50 and 1600, respectively. Respondents' decisions were influenced by the type of disease (they were more likely to consent if the drug would potentially help avoid lethal disease) and cost (consent increased at lower cost), but not by NNT. When taught how to interpret NNT, 24% of participants changed their decisions, 93% of which switched from positive to negative decisions, regardless of how high the NNT magnitude was. This suggests that NNT is difficult to understand for average patients. **NNT was not compared to other metrics, however.**
3. [Gyrd-Hansen \(2011\)](#): Not included by either (post-date for Akl, not sure why for Zipkin?). 895 people in Denmark were randomized to receive info expressed as ARR, RRR, NNT, or POL with respect to heart attacks, and asked if they would be willing to receive a specific treatment. Respondents were asked to imagine having a heart attack and being offered a drug that cost \$100 USD. They were given information on the hypothetical drug's side effects, and also given treatment effectiveness of either 1 or 2 interventions in terms of a single risk information formation and asked how willing they would be to take the medicine. Hearing more than one option improved sensitivity to the risk reduction magnitude. For NNT, joint evaluation had very little effect on sensitivity to risk reduction – so hearing more than one type of measure did not help people better understand NNTs, but did help them

better understand the other 3 measures. The researchers also did a 3-way interaction analysis to test the combined impact of risk information format, risk reduction magnitude, and separate versus joint evaluation setting, NNT was the only statistically significant outlier in this test. Respondents could randomly allocate the valuing interventions either separately (either great effect or small effect) or jointly (small effect and large effect). Participants were least sensitive to the scale when given info as NNT compared to all other formats. Evaluability bias was most evident in those presented with POL and ARR. Joint evaluation reduced the likelihood of consenting to treatment that offered the smallest effect. Basically, participants were more likely to differentiate sensitivity of risk when given multiple measures (other than NNT) than when given a single measure. Having NNT as their only measure made them least sensitive to the differences in risk, and adding NNT with another measure did not help people become more sensitive to the risk differences compared to adding one of the other measures. **Perhaps this qualifies as a comprehension test of a sort – whether they were more likely to differentiate higher and lower risk.**