

## **A conversation with Dr. Jelte Wicherts on August 28, 2014**

### **Participants**

- Dr. Jelte Wicherts – Associate Professor, Department of Methodology and Statistics, Tilburg University
- Jake Marcus – Research Analyst, GiveWell

**Note:** These notes were compiled by GiveWell and give an overview of the major points made by Dr. Jelte Wicherts.

### **Summary**

GiveWell spoke with Dr. Jelte Wicherts about the validity of cognitive tests used in iodine supplementation randomized controlled trials (RCTs), potential reasons for concern in interpreting the results from the iodine RCTs as gains in IQ, and the value of a three to five point IQ gain.

### **Comparing standard IQ tests with the cognitive tests used in iodine supplementation RCTs**

An IQ score is an averaged summary statistic of typically ten to twenty subtests for fluid intelligence, short-term memory, spatial reasoning, and other measures of intelligence. The measurements from these subtests tend to correlate highly with each other in standard IQ tests.

In contexts where children may not have the same levels of formal education as children in developed countries, researchers should choose specific subtests to get the most accurate and fair scores for a group. For example, presenting questions with multiple-choice answers to a group that is not used to the format would not result in a fair comparison to groups that answer multiple-choice questions frequently. Instead, researchers in these contexts usually focus on subtests like Raven's Progressive Matrices, which can be understood without scholastic training.

#### *Evaluating the validity of cognitive tests used in the iodine supplementation RCTs*

The first step in determining whether the results from the subtests used in the iodine supplementation RCTs are likely to correlate with cognitive abilities tapped by IQ tests is to see how strongly the results of the subtests correlate with each other. If the results of the subtests do not correlate very highly with each other, they are likely not measuring what IQ tests measure.

Even simple measures of short-term memory, like reciting a list of animal names, will probably have some correlation with IQ. However, estimates of cognitive gains through iodine supplementation that only relied on a few very simple tests might not correlate very highly with the results of an official IQ test. The greater the variety and number of subtests used, the more likely that the results of the tests would correlate with a standard measure of IQ.

A motor coordination test, such as threading beads, which was used in some of the iodine supplementation RCTs, is likely to be at least somewhat predictive of IQ. John B. Carroll's research is a good source for more information on the connection between motor coordination, cognitive subtests, and IQ.

### **Concerns about using the results from the iodine supplementation RCTs**

For each of the other studies, GiveWell obtains an overall treatment effect by standardizing the treatment effects for the individual tests and averaging them. GiveWell then multiplies this standardized treatment effect by 15 to convert it to an IQ scale. Dr. Wicherts thought this approach for estimating the average treatment effect across the subtests in the iodine supplementation RCTs and converting them to an IQ scale seemed reasonable. However, if one of the subtests in the iodine supplementation RCTs used for GiveWell's estimates did not show an effect, it is difficult to determine if the lack of effect is because the subtest was not reflective of general intelligence, or because the iodine supplementation did not cause cognitive gains. If the problem was an unreliable subtest not reflective of general intelligence, the average treatment effect would be biased downwards.

Preregistration for trials would somewhat alleviate concerns about the choice of subtests, since it would be easier to see if it seemed like specific subtests were chosen simply for their likelihood of observing the desired effect.

Publication bias is a potential issue with the iodine supplementation RCTs, since it is much more interesting to find an effect from iodine supplementation than it would be to not find an effect. Studies with very small sample sizes are also of concern, since they may inflate the estimated treatment effect. If the studies with small sample sizes show the biggest treatment effects, there is reason to suspect that the combined results are not reliable.

### **The value of a three to five point IQ gain**

The two most recent iodine supplementation RCTs had an estimated average treatment effect of a gain of three to five IQ points.

The Flynn effect is the gain in average IQ scores by about three points per decade, observed throughout the twentieth century. Twentieth century improvements in nutrition, education, and vaccination in the developed world may explain the Flynn effect. In comparison, achieving the same IQ gain as a decade of the Flynn effect with a low-cost intervention like iodine supplementation should be considered very valuable.

Another reference point illustrative of the value of three to five point IQ gains is the US education program Head Start. Head Start, despite spending thousands of dollars per child in the more intensive programs, did not achieve average IQ gains as high as the iodine supplementation RCTs.

*All GiveWell conversations are available at <http://www.givewell.org/conversations>*