

Effect of assessment method on the discrepancy between judgments  
of health disorders people have and do not have: A Web study.

Jonathan Baron<sup>1,2</sup>    David A. Asch<sup>2,3,4</sup>    Angela Fagerlin<sup>5</sup>    Christopher Jepson<sup>3</sup>  
George Loewenstein<sup>6</sup>    Jason Riis<sup>7</sup>    Margaret G. Stineman<sup>2,8</sup>    Peter A. Ubel<sup>5,9</sup>

July 7, 2002

---

<sup>1</sup>Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104-6196, e-mail [baron@psych.upenn.edu](mailto:baron@psych.upenn.edu), funded by N.I.H. Grant AG16258 (P. Ubel, Principal Investigator)

<sup>2</sup>Leonard Davis Institute of Health Economics, University of Pennsylvania

<sup>3</sup>Division of General Internal Medicine, University of Pennsylvania School of Medicine

<sup>4</sup>Veterans Affairs Medical Center, Philadelphia, Pennsylvania

<sup>5</sup>Division of General Internal Medicine, University of Michigan

<sup>6</sup>Department of Social and Decision Sciences, Carnegie-Mellon University

<sup>7</sup>Department of Psychology, University of Michigan

<sup>8</sup>Division of Rehabilitation Medicine, University of Pennsylvania School of Medicine

<sup>9</sup>Veterans Affairs Medical Center, Ann Arbor, Michigan

**Abstract**

1  
2 Three experiments on the World Wide Web asked subjects to rate the seriousness of common  
3 health disorders such as acne or arthritis. People who had a disorder (“Haves”) tended to rate  
4 it as less serious than people who did not have it (“Not-haves”). Two explanations of this Have  
5 vs. Not-have discrepancy were rejected. By one account, people change their reference point  
6 when they rate a disorder that they have. More precise reference points would, on this account,  
7 reduce the discrepancy, but, if anything, the discrepancy was larger. By another account,  
8 people who do not have the disorder focus on attributes that are most affected by it, and the  
9 discrepancy should decrease when people make ratings on several attributes. Again, if anything,  
10 the discrepancy increased when ratings were on separate attributes (combined by a weighted  
11 average). The discrepancy varied in size and direction across disorders. Subjects also thought  
12 that they would be less affected than others.

## 1 Introduction

Brickman and his colleagues argued that people adapt to their lives, so that the joy of positive changes and the despair of negative changes both wear off over time [1, 2]. Paraplegics, after a few years, are almost as “happy” as other people. Consistent with these observations, patients with chronic health disorders sometimes do not see their disorder as very bad, compared to judgments offered by those who have not experienced the disorder [3, 4]. In these studies, patients report that their quality of life is significantly better than the public estimates it would be. For example, Sackett and Torrance [4] found that the general public estimates the health related quality of life (HRQoL) of dialysis at a value of 0.39 (on a scale from 0 for conditions as bad as death to 1 for perfect health), whereas dialysis patients estimate their HRQoL at 0.56. Boyd et al. [3] found that patients without colostomies estimate the HRQoL of living with a colostomy at 0.80, while patients with colostomies rate their own HRQoL at 0.92. A similar discrepancy has been seen between rheumatoid arthritis patients and the general public.

Other studies, however, fail to find such discrepancies (e.g., Llewellyn-Thomas et al. [5]; O’Connor et al. [6]). These studies often examine short-term conditions, such as the experience of radiation therapy, evaluated by the same patients, before and during the experience.

The question we address here is whether the discrepancy between the ratings of people who have and do not have a disorder — the Have–Not-have discrepancy — can be found for common disorders in a broad sample of people. We ask this by using questionnaires on the World Wide Web, available to anyone. Although our sample is surely not representative of all human beings (the population of interest), it is diverse, and it includes many people with the health disorders of interest. If the discrepancy can be demonstrated by this sort of method, then future research on it is easier than would otherwise be the case.

We test here two explanations of the Have–Not-have discrepancy. In one, the meaning of the response scale changes when people evaluate disorders that they have. A term such as “good health,” which might be used for the top of a rating scale, can mean one thing to a normal healthy

39 person — being able to play tennis or ride a bike, for example — and quite another thing to a  
40 person who has just lost a leg, where it might mean being able to go back to the office. This can  
41 happen either because people compare themselves to others with similar disorders or because they  
42 adopt different aspiration levels. In either case, the discrepancy should be reduced by making the  
43 response scale more explicit, so that the terms used to name the ends of it are not so subject to  
44 variable interpretation.

45 This explanation is similar to the idea of “response shift” [7]. People undergoing cancer therapy,  
46 in some studies, retrospectively evaluate their pre-therapy condition as better than they rated that  
47 condition at the time they experienced it [8], or as worse, if their condition improved [9, 10].  
48 Analogous results are found for transplant patients, who rate their pre-transplant quality of life  
49 (QOL) as lower after the transplant than they did at the time [11, 12]. Jansen et al. [13], however,  
50 found no evidence for a response shift.

51 Here, in Experiments 1 and 2, we test this explanation by assessing the Have–Not-have discrep-  
52 ancy with different kinds of response scale, which differ in the clarity of their end point. If a scale  
53 is defined so that end points have a constant meaning, and the discrepancy is reduced, then the  
54 use of undefined scales is part of the explanation of it.

55 The other explanation holds that the discrepancy is the result of a focusing illusion. When  
56 imagining the situation of others who have a disorder, people focus disproportionately on what is  
57 affected by the disorder while ignoring those things that are not affected. In the original demon-  
58 stration of this focusing illusion, students in both California and the Midwest predict that they  
59 would be significantly happier living in California than in the Midwest [14]. This prediction cor-  
60 relates strongly with how important they feel weather is to their quality of life. Yet no significant  
61 difference in happiness is found between these two groups of students, suggesting that they focus  
62 disproportionately on the effect of weather on their quality of life when they compare themselves  
63 to those in a different climate.

64 In an earlier study, Ubel et al. [15] explored whether a focusing illusion contributes to general  
65 public estimates of the QOL associated with disabilities. Subjects estimated the QOL of either

66 paraplegia, below the knee amputation, or partial blindness. Then the experimenters attempted  
67 to “de-focus” subjects by having them reflect on the impact of these disabilities on a wide range  
68 of life domains. This de-focusing task should keep people from thinking too narrowly about the  
69 life domains affected by the disability. For example, subjects were asked to think about how such  
70 a disability would affect their family life, assuming that for most, it would have little effect. This  
71 manipulation showed no effect. The focusing hypothesis was not supported.

72 Ubel et al. [16] replicated this negative result with three other focusing tasks. One de-focusing  
73 task involved asking people to imagine how the disability in question would affect eight concrete  
74 life events, such as “paying bills and taxes” and “reading or watching TV or movies.” Another  
75 de-focusing task asked people to list those things that took up the largest amount of their time on  
76 the previous day and then indicate how much the disability in question would impact those things.  
77 A third de-focusing task asked people to specifically think not only about things that would be  
78 made worse about the disability in question, but also things that would be unchanged and things  
79 that would be better because of the disability. None of the tasks affected the magnitude of the  
80 discrepancy, measured after doing the task.

81 Here, we test the focusing illusion in a different way, in Experiment 3, by asking whether the  
82 Have–Not-have discrepancy is reduced when subjects provide ratings of the disorder attribute by  
83 attribute. This method insures that subjects attend to all the attributes we provide. It should thus  
84 prevent focusing, at least to some extent, and reduce the discrepancy, if the focusing explanation  
85 is at work.

86 Also relevant to the focusing illusion is a comparison of different kinds of scales in Experiments  
87 1 and 2. Scales that concern health should be less subject to a focusing illusion than those that are  
88 broader, such as those concerning quality of life, or happiness. In fact the Brickman study used a  
89 happiness scale. We might expect the happiness scale to show the smallest discrepancy.

90 In sum, our two main questions are whether the Have–Not-have discrepancy can be reduced by  
91 the use of well-defined scales (Experiments 1 and 2, which also examine different kinds of scales),  
92 and by the use of attribute-by-attribute ratings.

93 Experiments 1 and 2 also address a subsidiary question: whether people think that they can  
94 adapt to a disorder better than others can. In general, people tend to think they are “above  
95 average” on all good things [19]. Most studies comparing Have and Not-have are asking the Have  
96 group (those with the disorder) to rate themselves and the Not-have group to rate others. Thus,  
97 Have–Not-have is confounded with ratings of self and others. The Have–Not-have discrepancy could  
98 result from a belief that “I can adapt to this better than other people can,” whether or not the  
99 person answering has the disorder or not. We thus ask for ratings of self with the disorder, self  
100 without it, others with, and others without. One of the two Self ratings is necessarily hypothetical.  
101 A Self-Other difference would support this explanation.

102 It is also possible that this Self-Other difference is found mainly in those who have the disorder.  
103 If so, Self-Other would interact with Have–Not-have.

## 104 **2 Experiment 1**

105 The main purpose of this experiment was to look for a discrepancy in the ratings of common  
106 disorders, in which people who do not have the disorder rate it as more serious than those who  
107 have it (Not-have vs. Have). The ratings say how much dis-utility a disorder will create. So our  
108 general hypothesis is that Have’s don’t think things are as bad as Not-have’s think they would  
109 be. The justification for this hypothesis is that when people have found discrepancies, they have  
110 typically been in this direction.

111 Note that this discrepancy need not always go in this direction. For certain health disorders,  
112 especially those that primarily effect mood and subjective well being, we expect the discrepancy to  
113 go in the other direction. Classic examples of these types of disorders would be anxiety, depression  
114 and pain. And indeed in this study, one of the disorders that involves pain, migraines, showed  
115 a discrepancy in the opposite direction (Have rated worse than Not-have). Although we know of  
116 no previous findings of a reversed discrepancy of this sort, Adresen et al.[17] found that people  
117 who suffered pain as part of their disorder rated their pain as worse than did people who did not

118 have the disorder, and, in a different sort of study, Todorov and Kirchner[18] found that proxies  
119 under-reported symptoms of people with disabilities.

120 A second question was whether a discrepancy could result from vagueness in the judgment scale.  
121 In particular, if the scale has unclear anchors — that is, unclear standards of comparison — people  
122 with a health disorder might evaluate having the disorder by comparing themselves to others with  
123 the disorder, rather than to those without it. This possibility predicts that the discrepancy would  
124 be larger when the judgment scale is less clear.

125 We used three methods of eliciting judgments of undesirability, an anchored scale, a vague scale,  
126 and a happiness scale. The **anchored scale** is anchored at the bottom by death and at the top  
127 by the absence of the disorder being rated. The **vague scale** was anchored at the bottom by  
128 “extremely undesirable” and at the top by “not undesirable at all.” The **happiness scale** asked  
129 about the overall effect of the disorder on happiness. Both ends were vaguely described: “greatly  
130 increased happiness” and “greatly decreased happiness.”

131 The third main question concerns judgments of self vs. others. For each scale, subjects judged  
132 for themselves and for someone else. A Self-Other difference might help to explain past findings of  
133 a Have–Not-have discrepancy.

## 134 2.1 Method

135 Eighty subjects completed a questionnaire on the World Wide Web, at [http://www.psych.upenn.](http://www.psych.upenn.edu/~baron/qs.html)  
136 [edu/~baron/qs.html](http://www.psych.upenn.edu/~baron/qs.html). Their ages ranged from 16 to 60 (median 34); 27.6% were male; 15% were  
137 students. Most subjects had completed other studies on other topics at the same site. They had  
138 originally discovered it either through search engines, links from other sites (such as those listing  
139 ways to earn money on the web), or “word of mouth.” In general, the population of visitors to  
140 this site, while not all U.S. residents, has about the same median education and income as the U.S.  
141 adult population. Other than the subjects being web users, their most salient characteristic is that  
142 they are mostly women. The questionnaire began:

143 **Preference for health conditions**

144 The following study concerns judgments of chronic health conditions. . . .

145 One [type of] question concerns undesirability, which means the strength of preference  
146 for not having the condition.

147 [Another type of] question concerns the effect of the condition on overall happiness or  
148 unhappiness.

149 You will make the undesirability ratings from two points of view. One is your own point  
150 of view, if you had the condition. If you have or have had it, please pay attention to  
151 the description, and think about that rather than your own case.

152 The other is the point of view of the average person.

153 In each case, imagine that either you or the average person has had the condition for  
154 6 months. Also, the condition will not change in the foreseeable future. It will not  
155 get better, and it will not get worse. This is important. Do not suppose that it will  
156 improve.

157 A typical item (one screen) using the anchored scale for both Self and Other ratings is:

158 Item 1 out of 45:

159 Suppose that **you** had the condition: **bad knee — running is painful**.

160 **For yourself**, on a scale of undesirability in which 0 is 'not having' bad knee (with  
161 everything else the same) and

162 100 is 'imminent death', where would you put **bad knee**?

163 Now suppose that **the average person** had **bad knee (running is painful)**.

164 Where do you think that the average person would put **bad knee** on the same scale?

165



166 The description for the vague scale was, “on a scale of undesirability in which 0 is ‘not unde-  
 167 sirable at all’ and 100 is ‘extremely undesirable’.” The description for the happiness scale was “on  
 168 a scale of happiness/unhappiness in which 5 is ‘greatly increased happiness from the condition’,  
 169 0 is ‘no change in happiness from the condition’, and -5 is ‘greatly decreased happiness from the  
 170 condition.” (Formatting is omitted here.) Each subject used one scale type at a time, for all the  
 171 disorders. The order of the three types was randomized across subjects.

172 With each scale type, the subject rated the following 15 health conditions, shown here with  
 173 their brief descriptions:

asthma	attacks of breathing difficulty — 1 per week
chronic back pain	running is impossible and walking difficult
bad knee	running is painful
inability to walk	requires a wheelchair
insomnia	2 hours less sleep than desired on most nights
migraines	debilitating 2 hour headaches — 2 per week
short stature	6 inches shorter than average for sex
excessive weight	50% more than normal weight
nearsightedness	glasses required
partial deafness	hearing aid required
nightmares	frightening dreams most nights
acne	pimples all over face
smoking habit	pack a day of cigarettes
arthritis	pain in hips or shoulders with any movement
heart disease	chest pain from walking or other activity

174 After completing all the ratings, the subjects were asked, “Which of the following conditions  
 175 have you had yourself for at least a year? (You may include conditions that were more or less  
 176 severe than the descriptions used here.)” They saw a list of the conditions (disorders), with a letter  
 177

178 before each, and they were asked to type the letters.

## 179 2.2 Results

### 180 2.2.1 Not-have–Have discrepancy

181 To look for a Not-have–Have discrepancy effect, we first standardized the ratings for each disorder  
182 across subjects (separately for Self and Other ratings with each of the three scales). This removed  
183 the effect of differences in the seriousness of disorders. (Otherwise, we would expect that Have  
184 would seem less serious than Not-have simply because the less serious disorders are more frequent.)

185 Then we eliminated disorders that were rare in our sample. Because of the standardization,  
186 rare disorders could end up with extreme  $z$  scores. As it happened, four of the disorders (inability  
187 to walk, partial deafness, nightmares, and heart disease) occurred in 4 subjects or fewer, and all  
188 others occurred in 11 subjects or more, so we dropped these four health disorders from further  
189 analysis of the Not-have–Have discrepancy.

190 We then computed the mean of the standardized Have disorders for each subject, separately for  
191 Self and Other ratings and for the different scales, and the mean of the Not-have disorders. Table 1  
192 summarizes the main results by scale type and type of difference (with  $p$  levels based on two-tailed  
193  $t$  tests). Note that numbers refer to seriousness so that larger numbers are worse.

— Insert Table 1 —

194 The results showed a clear Not-have–Have discrepancy, when we combined the results from all  
195 three scales (which was possible, since all had been standardized). Have was considered less serious  
196 than Not-have. Across subjects, combining Self and Other, the mean  $z$ -score difference between  
197 Not-have and Have was .13 ( $t_{74} = 2.23$ ,  $p = .0291$ , across subjects — note that somewhat different  
198 subjects are involved in different analyses because of missing data). The effect was present for both  
199 Self (difference .11,  $t = 1.83$ ,  $p = .0728$ ) and Other (difference .14,  $t = 2.27$ ,  $p = .0260$ ). It is  
200 apparent that the Self-other by Not-have–Have interaction did not appear. The effects was at least  
201 as large for Other as for Self.

202 Our main hypothesis was that the Not-have–Have discrepancy would be larger for the vaguer  
203 scales. This was not supported. Table 1 shows the relevant results. Although the discrepancy was  
204 significant for the anchored and happiness scale and not for the vague scale, the three scales did  
205 not differ significantly in the size of the discrepancy. Importantly, the discrepancy was found for  
206 the anchored scale.

### 207 **2.2.2 Self-Other difference**

208 To examine the Self-Other difference, we used all health disorders (since this difference was between  
209 questions presented on the same screen and thus had less error). First, we looked for an overall Self-  
210 Other difference across all three measures. To do this, we used the actual ratings (not standardized),  
211 but we multiplied the happiness ratings by  $-20$  to roughly equate the scales (because most ratings  
212 were between 0 and  $-5$ ). Table 2 shows the results for each scale (again, with high numbers  
213 representing worse health).

— Insert Table 2 —

214 Combining the three measures, the overall difference amounted to a mean of 2.0 points on the  
215 100 point scale, with ratings more severe for other than self. As shown in Table 2, the difference  
216 was significant overall and for the anchored scale. Again, the differences among scales in the size  
217 of the effect were not significant.

218 It thus appears that the Self-Other discrepancy exists as hypothesized, and it therefore may  
219 account somewhat for the Have–Not-have discrepancy when Have’s are asked about themselves and  
220 Not-have’s are asked about others.

221 Although we had no particular hypothesis about the interaction between scale type and Other-  
222 Self, it appears, again, that the anchored scale is most sensitive to the difference.

### 223 **2.2.3 Differences among disorders**

224 The Not-have–Have discrepancy depended on the disorder. To show this, we asked whether dis-  
225 orders with a higher discrepancy in one half of the subjects were also higher in the other half.

226 We measured the mean discrepancy for the odd-numbered subjects for each disorder (combining  
227 all three measures) and the mean for the even-numbered subjects. The two sets of means were  
228 correlated across the 15 disorders ( $r = .74$ ,  $p = .0013$  one tailed). Table 3 shows, in the rightmost  
229 column, the Not-have–Have discrepancy for the different disorders. Although we made an effort to  
230 find all the usable common disorders, it seems that the overall result of a positive discrepancy was  
231 an artifact of our sample, and a different sample might have yielded even a reversed discrepancy.  
232 For example, migraines seem to be worse to those who have them than to those who do not. (The  
233 rare disorders are in parentheses. While these numbers are suggestive of similar results, recall that  
234 only four subjects or fewer had each of these disorders.)

235 We found the same kind of consistency across scale measures for the Self-Other difference  
236 ( $r = .89$ ). The means for the disorders are shown in Table 3. Table 3 also shows the mean severity  
237 ratings (averaging all three measures, with happiness multiplied by  $-20$  before averaging).

— Insert Table 3 —

### 238 **3 Experiment 2**

239 The main result of Experiment 1 was a failure to find a larger discrepancy between Haves and  
240 Not-haves in the vague scale or in the happiness scale than in the anchored scale. The happiness  
241 scale did, however, show a slightly larger discrepancy than the other two scales. Moreover, the use  
242 of “increased happiness” and “decreased happiness” in the description of the happiness scale might  
243 have tended to reduce the tendency to think of one’s own disorder as a reference point. Such a  
244 tendency is one of the mechanisms that could lead to a smaller discrepancy for the happiness scale.

245 It is possible that a scale focusing more explicitly on quality of life, rather than seriousness of  
246 a health state, might show a larger discrepancy. Especially when the reference points are vague,  
247 people with a disorder may tend to think about others with the same disorder when they evaluate  
248 their quality of life.

249 In Experiment 2 (actually done after Experiment 3), we asked four questions, with vague vs.

250 specific crossed with severity (of an disorder) vs. quality of life (QOL).

### 251 **3.1 Method**

252 Ninety-nine subjects completed a questionnaire on the World Wide Web. Their ages ranged from  
253 19 to 68 (median 36); 22.2% were male; 11% were students. The questionnaire began:

#### 254 **Health conditions**

255 The following study concerns judgments of chronic health conditions. There are 40  
256 questions (screens). Each question presents a short description of a health condition  
257 and asks one of two types of rating question.

258 One question concerns health. We ask you to rate the health condition on a 50-100 scale.  
259 Questions differ in how 50 and 100 are defined. Pay attention to these definitions.

260 The other question concerns the overall quality of life of those who have the condition.

261 Note: You can go below 50 if you feel that a condition is worse than the definition of  
262 "50".

263 You will make the ratings from two points of view. One is your own point of view,  
264 supposing that you have the condition. If you have really had the condition, please pay  
265 attention to the description of the condition, which may be more or less serious than  
266 your own case.

267 The other point of view is that of the average person who gets the condition described.

268 In each case, imagine that either you or the average person has had the condition for  
269 at least 6 months. Also, the condition will not change in the foreseeable future. It will  
270 not get better, and it will not get worse. Do not suppose that it will improve.

271 A typical question, in the QOL-specific condition, appeared as follows (with the definitions of  
272 100 and 50 in green and red, respectively):

273 Suppose that **you** had the condition: **acne — pimples all over face**.

274 For yourself, on a scale of overall **quality of life** in which

275 100 is as good as that of someone with a meaningful job, friends, family, and good  
276 health, and

277 50 is as bad as that of someone who cannot walk more than 10 feet because of partial  
278 paralysis, has a dull job, and no close family or friends,

279 where would you put **acne**?

280 Now suppose that **the average person** had **acne (pimples all over face)**.

281 Where do you think that the average person would put **acne** on the same scale?

282 The scale definitions for the QOL-vague condition were: “100 is a very good quality of life and  
283 50 is a very poor quality of life.” (We used 50 to make it easier for subjects to assign number below  
284 the bottom anchor.) For the health-specific condition they were: “100 is as healthy as a 20-year-old  
285 with no health conditions and 50 is as serious as being unable to walk more than 10 feet because of  
286 partial paralysis.” and for the health-vague condition they were, “100 is very good health and 50  
287 is very poor health.” The four conditions were blocked, and in a different random order for each  
288 subjects.

289 The health states were chosen on the basis of previous studies to be common. Their order  
290 randomized separately for each subject. Their definitions were:

Asthma: attacks of breathing difficulty — 1 per week

Back pain: pain from lifting objects

Insomnia: 2 hours less sleep than desired on most nights

Shortness: 6 inches shorter than average height for sex

Overweight: inability to lose excess weight

Nearsightedness: glasses required

Acne: pimples all over face

Smoking habit: uncontrollable addiction to cigarettes

Arthritis: pain in hips or shoulders with some movements

291 Heart disease: occasional chest pain from climbing stairs

292 At the end of the study, subjects again indicated which disorders they had for at least a year.

### 293 3.2 Results

294 Table 4 shows the mean ratings for the four conditions (after elimination of bad data, as we shall  
295 describe). Here, high numbers represent good health or good QOL. It is apparent that the four  
296 measures agreed closely on the relative seriousness of the health states, and the health states vary  
297 considerably in seriousness. Some subjects seemed to misunderstand the scale for some condition  
298 blocks. To assess misunderstanding, we correlated each subject's scores in each block with the ten  
299 means based on all four conditions. We eliminated blocks when the correlation was less than .25.  
300 This resulted in deletion of 16.3% of the data. Two subjects had all their data deleted, and 58 had  
301 no data deleted. All statistical analysis is based on whatever data were available.

— Insert Table 4 —

302 As is apparent in Table 4, Self ratings are consistently higher (less severe) than Other ratings,  
303 as found in Experiment 1 (mean difference of 2.01,  $t_{57} = 6.77$ ,  $p = 0.0000$ ). Specific ratings are  
304 also higher than vague ratings (mean difference 3.02,  $t_{57} = 5.68$ ,  $p = 0.0000$ ), but this is surely  
305 the result of the specific scale definitions, not a general result. Health and QOL ratings did not  
306 differ significantly. Self-Other difference was slightly greater for health than for QOL ( $t_{57} = 2.55$ ,  
307  $p = 0.0135$ ); this too does not seem to imply any general conclusion.

308 Figure 1 shows the results for the Not-have vs. Have discrepancy for the ten health states and  
309 the four conditions (for Self and Other combined); positive numbers indicate that Not-have's rate  
310 the disorder as worse than Have's. We did not find an overall Not-have–Have discrepancy, even for  
311 Self. This was because the health states differed in the direction of this effect, as is apparent from  
312 Figure 1. The correlation across disorders between the discrepancy measures for two halves of the  
313 sample (computed as in Experiment 1) was .49 ( $p = .0771$ , one tailed).

— Insert Figure 1 —

314 Although the overall discrepancy (Have less severe than Not-have) was not significant, we  
315 note, first, that it was greater for Self (0.063) than for Other (0.030, in contrast to Experiment 1,

316  $t_{54} = 2.22, p = 0.0303$ ).

317 It is also apparent (from Figure 1) that the discrepancy, to the extent to which we find it,  
318 was no larger for QOL than for health. In fact, it was nearly significant for health alone in the  
319 Self condition (mean 0.25,  $t_{69} = 1.92, p = 0.0596$ ) and in the opposite direction for QOL (-0.17),  
320 resulting in a significantly greater discrepancy (Have higher than Not-have) for health than for  
321 QOL ( $t_{63} = 3.03, p = 0.0036$ ). This result starkly contradicts the hypothesis that the discrepancy  
322 would be greater for QOL.

323 Similarly, the discrepancy was no larger for vague than specific. Again, we found the opposite  
324 ( $t_{63} = -2.23, p = 0.0291$ ): the Not-have–Have discrepancy was greater (higher ratings for Have)  
325 in the specific than in the vague condition.

## 326 4 Experiment 3

327 The purpose of Experiment 3 was to ask whether the discrepancy effect was the result of a focusing  
328 illusion. The idea of a focusing illusion is that, when people are asked about a difference between  
329 two disorders, they focus on the attributes of those disorders that are different, because of the way  
330 the question is asked. They thus exaggerate the magnitude of the difference. For example, people  
331 asked to compare life in California with life in the midwestern states of the U.S. focused on the  
332 weather, thus overestimating the benefits of living in California [14]. Likewise, people who do not  
333 have a disorder could focus on the attributes of life that are more affected by that disorder.

334 As we explained in the Introduction, a previous study [15] found no evidence for a focusing  
335 illusion as an explanation of the Have–Not-have discrepancy. The study tried to reduce such an  
336 illusion, if it existed, by calling subjects' attention to a variety of attributes that characterize the  
337 goodness of life. Asking subjects how a disorder affected each of these dimensions did not affect  
338 their subsequent rating of the same disorder.

339 Ubel et al. [15] used rare conditions, such as below-the-knee amputation and paraplegia, so  
340 they were unable to make the Not-have–Have comparison in their sample (jurors). We do not know



341 whether they would have found a discrepancy effect, if they had done so. Also, it is possible that  
342 subjects may have understood that the disorders did not affect all life attributes equally, yet, still,  
343 reverted to the focusing illusion even after they were forced to consider all the attributes.

344 In the present experiment, we followed the basic design of Ubel et al. In particular, we first  
345 asked for holistic ratings, then we asked for attribute-by-attribute ratings, and finally we asked for  
346 holistic ratings again. However, we asked about the attribute-by-attribute ratings in a way that  
347 allows us to do a rough calculation of utility based on multi-attribute utility theory [20]. We thus  
348 call this a MAU elicitation. In particular, we asked for numerical ratings on each of seven attributes,  
349 each with a clearly anchored endpoint. At the end of the experiment we asked for weights of these  
350 attributes. To calculate the MAU (multi-attribute utility) of each disorder for each subject, we  
351 multiplied each attribute rating by the weight of that attribute and then added up these products  
352 across the seven attributes.

353 If the Not-have–Have discrepancy results from a focusing illusion, it should largely disappear in  
354 the MAU ratings, because subjects were forced to rate attributes one at a time. Thus, even if the  
355 effect of this forced variety is limited to the MAU task itself, we should be able to detect it. We do  
356 not need to rely on transfer to the subsequent holistic rating task.

357 For the holistic question, we used paralysis from the waist down as the standard, rather than  
358 death, because we thought that people might be reluctant to say that a disorder they had was  
359 anywhere near as bad as death.

## 360 4.1 Method

361 Seventy-nine subjects completed a questionnaire on the World Wide Web. Their ages ranged from  
362 18 to 74 (median 36); 33% were male; 14% were students.

363 The questionnaire began:

### 364 Preference for health conditions

365 This study concerns judgments of chronic health conditions. It has 24 screens. Please

366 read all of these instructions carefully.

367 In each question, you will see a short description of a health condition, and you will  
368 answer one of two types of questions.

369 One question concerns undesirability, which means the strength of preference for not  
370 having the condition. This type of question will come first, and then it will be repeated  
371 at the end. Do not worry about whether your answer is the same or not. Just try to  
372 answer accurately both times.

373 You answer this question on a scale where 0 means “not having the condition” and 100  
374 means “as bad as being paralyzed from the waist down.” You can use numbers greater  
375 than 100 if necessary (but no greater than 200).

376 The other type of question concerns the negative effect of the condition on several  
377 domains of your life:

- 378 1. Pain and discomfort
- 379 2. Economic standard of living
- 380 3. Work
- 381 4. Love life
- 382 5. Family life other than love life
- 383 6. Spiritual life broadly defined
- 384 7. Leisure activities other than family life

385 Please try to interpret these descriptions so that they do not count the same effects  
386 twice. For example, if “spiritual life” includes communing with nature, do not also  
387 count this as part of “leisure”.

388 You answer these questions on a scale where 0 means “no negative effect” and the  
389 meaning of 100 is specified in the question. If you think that some health condition has

390 a **positive** effect, then use a **negative number** for your response. Remember, this  
 391 question is about negative effects.

392 At the end, you will be asked a few additional questions.

393 Each question asks you to “suppose that you had” the condition. Of course, you may  
 394 actually have it now. If so, please pay attention to the description, which may be more  
 395 or less severe than your own case. Rate that description, not your own case.

396 In each case, imagine that you have had the condition for 6 months. Also, the condition  
 397 will not change in the foreseeable future. It will not get better, and it will not get worse.

398 This is important. Do not suppose that it will improve.

399 We used the following disorders, which had showed a discrepancy in the hypothesized direction  
 400 in Experiment 1 (except for heart disease, which was modified to be less severe).

asthma	attacks of breathing difficulty — 1 per week
insomnia	2 hours less sleep than desired on most nights
short stature	6 inches shorter than average for sex
nearsightedness	glasses required
acne	pimples all over face
smoking habit	pack a day of cigarettes
arthritis	pain in hips or shoulders with any movement
heart disease	occasional chest pain from climbing stairs

401 For the holistic task, a typical question read.  
 402

403 Suppose that you had the condition: arthritis – pain in hips or shoulders with any  
 404 movement. On a scale of overall undesirability in which

405 0 is not having arthritis (with everything else the same) and

406 100 is waist-down paralysis,

407 where would you put  
408 arthritis?

409 For the MAU task, a typical question read:

410 Suppose that you had the condition: insomnia – 2 hours less sleep than desired on most  
411 nights. Rate this condition for its negative effect on each of the following domains of  
412 your life. 0 means 'no negative effect at all.' 100 is defined for each question. (Use a  
413 negative number for a positive effect.)

414 1. Pain, fatigue, and discomfort: 0 = no effect; 100 = as bad as death.

415 2. Economic standard of living: 0 = no effect; 100 = dire poverty.

416 3. Work: 0 = no effect; 100 = unable to do any work.

417 4. Love life: 0 = no effect; 100 = love life nonexistent.

418 5. Family life: 0 = no effect; 100 = family life nonexistent.

419 6. Spiritual life: 0 = no effect; 100 = spiritual life nonexistent.

420 7. Leisure activities: 0 = no effect; 100 = activities nonexistent.

421 The order of health conditions was randomized separately for each subject and then fixed for  
422 the three parts of the questionnaire.

423 The weight elicitation, at the end of the questionnaire, read.

424 Now please rate each of the following on a scale where 0 represents 'not bad at all' and  
425 100 represents 'as bad as paralysis from the waist down'. When you make these ratings,  
426 think only about the thing you are rating. Try to imagine that everything else is the  
427 same. (This is hard, but do your best.)

428 Then subjects saw a list of all the lower ends of each scale in the MAU task, e.g, "Pain, fatigue,  
429 and discomfort as bad as death."

## 4.2 Results

Seven subjects were dropped because they gave 0 responses to all or most of the holistic questions in one section (including the more serious disorders) or because they gave 0 to all the questions about weights. Also, some subjects were dropped from some analyses. In particular, subjects whose holistic ratings before and after the MAU ratings did not correlate positively with each other across disorders were dropped for analyses of the holistic ratings. When holistic ratings and MAU ratings correlated negatively, we dropped the one with the lowest correlation with the mean of all severity measures across disorders. By these criteria, we dropped one subject from analysis of the MAU data and 13 subjects from analysis of the holistic data. The first holistic task seemed particular prone to induce reversed judgments (high numbers for less serious disorders).

To compute MAU utilities, we first re-scaled the weights for each subject so that the maximum weight was 1. Likewise, we rescaled the ratings so that the highest rating given by each subject was 1. Then we multiplied the weights by the ratings for each attribute for each disorder for each subject. Notice that the weights were elicited using the same end points as those used in the rating task.

To assess the Not-have–Have discrepancy, we used the standardized responses as in Experiment 1. We also standardized the MAU utilities by disease.

The Not-have–Have discrepancy was significant overall, averaging the holistic (averaged over the two parts) and MAU ratings (mean difference .16 in terms of standardized scores,  $t_{61} = 2.19$ ,  $p = .0320$ , two tailed; positive numbers mean that Have is less serious than Not-have). It was also significant for the MAU ratings (mean 0.15,  $t = 2.51$ ,  $p = .0144$ ) but not for the holistic ratings (mean 0.14,  $t_{64} = 1.57$ ,  $p = .1213$ ). These results are a clear rejection of the focusing hypothesis, which predicted an effect for the holistic ratings but not for the MAU ratings.

Again, the discrepancy varied across disorders, as shown in Figure 2. The split half correlation across the eight disorders in the discrepancy (computed as in Experiments 1 and 2) was 0.70 ( $p = .0257$  one tailed, with 6 df).

— Insert Figure 2 —

456 The pattern of attribute ratings should vary by disorder. To test whether this was true, we  
457 performed an analysis of variance of the attribute ratings, using subject, attribute and disorder  
458 as factors. Importantly, the interaction between attribute and disorder was significant ( $F_{42,3276} =$   
459  $28.05$ ,  $p = .0000$ ). Main effects of attribute and disorder were also large and significant. Table 5  
460 shows the mean ratings, on a scale where the worse end of each attribute is 1 and “no effect” is  
461 0, and the Not-have–Have discrepancy on the same scale. The interaction between attribute and  
462 disorder is illustrated by the high numbers for the effect of arthritis on pain and the effect of acne  
463 (and shortness) on love life.

— Insert Table 5 —

464 We examined the Not-have–Have discrepancy for each attribute. The largest discrepancies were  
465 those for family life ( $t_{72} = 3.18$ ,  $p = .0022$ , two tailed), love life ( $t = 2.10$ ,  $p = .0396$ ), and spiritual  
466 life ( $t = 3.53$ ,  $p = .0007$ ). The effect for work was nearly significant ( $t = 1.75$ ). No other effect  
467 was significant by the usual criterion, although all were in the hypothesized direction, including  
468 pain and discomfort. These results suggest that those who do not have the disorders may fail to  
469 recognize how little effect they have on some domains.

## 470 5 Discussion

471 Our results indicate that the Have–Not-have discrepancy can be studied with in web respondents  
472 and common disorders. Although a minority of subjects had each disorder, most of them had at  
473 least one. For more general purposes the list could be expanded to include other aspects of life  
474 aside from health, such as being single, poor, unemployed — or good things such as being wealthy.  
475 Indeed, we may have done this by include shortness as a disorder. We have no reason to think that  
476 the general principles underlying the discrepancy would differ for non-health states.

477 We found consistent differences among disorders in the magnitude and direction of the dis-  
478 crepancy. One possible explanation of these differences — and there may well be more than one

479 — is that the usual discrepancy (Not-have worse than Have) is found for disorders that have an  
480 external manifestation, visible to others, while the opposite is found for disorders that involve pain  
481 or un-expressed emotion. Further research should examine these differences among disorders.

482 Our main conclusion is that vague scales are not the cause of the discrepancy. If anything, the  
483 discrepancy is larger when scales have more precise anchors. Thus, the discrepancy is not a simple  
484 artifact of the use of vague scales. Note, however, that this finding does not excuse the use of such  
485 scales. For other purposes, the tendency to chose an anchor close to one’s own condition may affect  
486 the conclusions drawn.

487 A limitation of this conclusion follows from our use of within-subject designs. It is possible that  
488 those who have a disorder recalibrate their entire scale for judgments of all disorders (or of their  
489 absence). If, for example, a paraplegic rated himself as “happy,” he might still believe that others  
490 are happier and rate a typical person without paraplegia as “ecstatic.” We do not think that such a  
491 general re-calibration is likely for the kinds of scales we used, for the kinds of minor disorders that  
492 we studied, but such recalibration may exist elsewhere. If it exists, a different design is required to  
493 detect it.

494 We also found further evidence that de-focusing manipulations do not reduce the discrepancy.  
495 Unlike earlier studies, these results did not rely on the carry-over from a de-focusing manipulation  
496 to another task. We found the discrepancy in the MAU task itself.

497 Also, if the focusing hypothesis were true, we might have found a larger discrepancy for health  
498 than for QOL in Experiments 1 and 2 (especially Experiment 2, which was better controlled), and  
499 we did not find this. Instead, we found a larger discrepancy for health than for quality of life in  
500 Experiment 2.

501 The focusing hypothesis is not quite dead, however. It is possible that, even within a life domain,  
502 people with a disorder focus on sub-attributes of that domain that are less affected by the disorder.  
503 It may be difficult to draw a line between “attributes” and “activities,” however. It may therefore  
504 be difficult to distinguish this form of the focusing hypothesis from other hypotheses that involve  
505 knowledge of specific adaptations (such as using a computer for reading when one is blind).

506 Finally, people think they will adapt better than others. In Experiment 2, this is more true  
507 when they have in fact adapted (Haves). This result could explain some of the discrepancy found in  
508 previous studies, if people who do not have the disorder think about others rather than themselves.

509 The main remaining explanations of the discrepancy are failure of Not-haves to predict adap-  
510 tation and self-deception. This is an important distinction. Adaptation is real, and self-deception  
511 is, in an important sense, false. We may think of people's values in terms of what Keeney calls  
512 fundamental values and means values [21]. People's good is in their fundamental values. (These  
513 may include wanting their means values honored.) Means values are connected to fundamental  
514 values through beliefs. If beliefs are false, then the means values lose their claim to represent a per-  
515 son's good [22]. Thus, when we help people make decisions, we do not do them any good when we  
516 honor their values based on self-deception and when these conflict with their fundamental values. If  
517 people with disabilities are deceiving themselves about how close to normal they are, then, to this  
518 extent, we should discount their judgments and work harder to cure and prevent their disorders  
519 than their own judgments would imply.

520 On the other side, mis-prediction of adaptation is a false belief held by Not-haves. To the  
521 extent to which adaptation is real and not predicted, then we should discount these judgments  
522 in evaluating the seriousness of disorders. Of course, both kinds of error could be true. And our  
523 results suggest a different kind of Not-have error as well, a failure to appreciate the seriousness of  
524 disorders that have no external manifestation.



## References

- 525
- 526 [1] Brickman P, Coates D, Janoff-Bulman R. Lottery winners and accident victims: Is happiness  
527 relative? *Journal of Personality and Social Psychology*, 1978; 36: 917–927.
- 528 [2] Brickman P, Campbell DT. Hedonic relativism and planning the good society. In M. H. Appley  
529 (Ed.), *Adaptation-level theory: A symposium*. New York: Academic Press, 1971.
- 530 [3] Boyd NF, Sutherland HJ, Heasman KZ, Tritchler DL, Cummings BJ. Whose utilities for  
531 decision analysis? *Medical Decision Making*, 1990; 10: 58–67.
- 532 [4] Sackett DL, Torrance GW. The utility of different health states as perceived by the general  
533 public. *Journal of Chronic Diseases*, 1978; 31: 697–704.
- 534 [5] Llewellyn-Thomas HA, Sutherland HJ, Theil EC. Do patients evaluations of a future health  
535 state change when they actually enter that state? *Medical Care*, 1993; 31: 1002–1012.
- 536 [6] O'Connor AMC, Boyd NF, Warde P, Stolbach L, Till JE. Eliciting preferences for alternative  
537 drug therapies in oncology: influence of treatment outcome description, elicitation technique  
538 and treatment experience on preferences. *Journal of Chronic Disease*, 1987; 40: 811–818.
- 539 [7] Sprangers MAG, Schwartz CE. Integrating response shift into health-related quality-of-life  
540 research: A theoretical model. *Social Science and Medicine*, 1999; 48: 1507–1515.
- 541 [8] Visser MR, Smets E.M, Sprangers MA, de Haes HJ. How response shift may affect the mea-  
542 surement of change in fatigue. *Journal of Pain and Symptom Management*, 2000; 20: 12–8.
- 543 [9] Bernhard J, Hürny C, Maibach R, Herrmann R, Laffer U. Quality of life as subjective experi-  
544 ence: reframing of perception in patients with colon cancer undergoing radical resection with  
545 or without adjuvant chemotherapy. *Annals of Oncology*, 1999; 10: 775–782.
- 546 [10] Bernhard J, Lowi A, Maibach R, Hürny C. Response shift in the perception of health for utility  
547 evaluation: an explorative investigation. *European Journal of Cancer*, 2001; 37: 1729–1735.
- 548 [11] Adang EMM, Kootstra G, Engel GL, van Hooff JP, Merckelbach HLGJ. Do retrospective  
549 and prospective quality of life assessments differ for pancreas-kidney transplant recipients.  
550 *Transplant International*, 1998; 11: 11–15.

- 551 [12] Postulart D, Adang EM. Response shift and adaptation in chronically ill patients. *Medical*  
552 *Decision Making*, 20; 186–93, 2000.
- 553 [13] Jansen SJ, Stiggelbout AM, Wakker PP, Nooij MA, Noordijk EM, Kievit J. Unstable prefer-  
554 ences: a shift in valuation or an effect of the elicitation procedure?? *Medical Decision Making*,  
555 2000; 20: 62–71.
- 556 [14] Schkade DA, Kahneman D. Does living in California make people happy? A focusing illusion  
557 in judgments of life satisfaction. *Psychological Science* 1998; 9: 340–346.
- 558 [15] Ubel PA, Loewenstein G, Hershey JC, Baron J, Mohr T, Asch DA, Jepson C. Do nonpatients  
559 underestimate the quality of life of chronic health states because of a focusing illusion? *Medical*  
560 *Decision Making*, 2001; 21: 190–199.
- 561 [16] Ubel PA, Loewenstein G, Jepson C. Quality of life ratings of unfamiliar disabilities: the con-  
562 tribution of a focusing illusion and failure to predict adaptation. Manuscript, 2002.
- 563 [17] Andresen EM, Vahle VJ, Lollar D. Proxy reliability: Health-related quality of life (HRQoL)  
564 measures for people with disability. *Quality of Life Research*, 2001; 10: 609–619.
- 565 [18] Todoroff A, Kirchner C. Bias in proxies' reports of disability: Data from the National Health  
566 Interview Survey on Disability. *American Journal of Public Health*, 2000; 90: 1248–1253.
- 567 [19] Weinstein ND. Unrealistic optimism about future life events. *Journal of Personality and Social*  
568 *Psychology*, 1980; 39: 806–820.
- 569 [20] Keeney RL, Raiffa H. *Decisions with multiple objectives: Preference and value tradeoffs*. New  
570 York: Cambridge University Press, 1993.
- 571 [21] Keeney RL. *Value-focused thinking: A path to creative decisionmaking*. Cambridge, MA:  
572 Harvard University Press, 1992.
- 573 [22] Baron J. Norm-endorsement utilitarianism and the nature of utility. *Economics and Philosophy*,  
574 1996; 12: 165–182.

Table 1: Relationship between scale type and discrepancy, in z scores of seriousness (mean of Self and Other, common disorders only), Experiment 1.

Scale type	Haves	Not-haves	p
Anchored	-0.08	0.04	.0430
Vague	-0.03	0.02	n.s.
Happiness	-0.13	0.04	.0479
Overall	-0.08	0.03	.0291

Table 2: Relationship between scale type and Self-Other difference, in seriousness scores on a rough 100 point scale (higher numbers representing worse health), all disorders, Experiment 1.

Scale type	Self	Other	p
Anchored	38.7	41.2	.0028
Vague	55.9	56.8	n.s.
Happiness	44.9	47.6	.0609
Overall	46.5	48.5	.0105

Table 3: Mean severity ratings for the 11 common disorders, and Self-Other (positive for Other worse than Self) and Not-have–Have differences (positive for Not-have worse than Have), Experiment 1 (ordered by discrepancy).

Health disorder	Number having	Mean severity (0–100 scale)	Self-Other difference (0–100 scale)	Not-have–Have difference (z scores)
Smoking	23	41.9	−14.0	0.41
Acne	13	43.6	4.2	0.39
Short	29	25.5	12.1	0.32
Insom	22	37.3	3.2	0.09
Asthma	11	49.2	3.2	0.09
Arthritis	18	56.5	0.8	0.05
Nearsht	43	21.7	6.5	0.04
Back	21	65.1	−1.6	−0.01
Weight	10	56.0	0.5	−0.02
Knee	12	38.1	4.0	−0.11
Migr	22	56.4	−0.8	−0.15

Table 4: Mean ratings in Experiment 2, for ten health states in the four conditions, for self and other (based on cleaned-up data); higher numbers represent better health.

		Health		QOL	
		specific	vague	specific	vague
Nearsighted	self	92.1	92.3	91.0	90.1
	<i>other</i>	<i>89.5</i>	<i>87.9</i>	<i>88.8</i>	<i>88.0</i>
Short	self	90.7	90.0	90.9	90.1
	<i>other</i>	<i>86.1</i>	<i>85.3</i>	<i>86.6</i>	<i>85.5</i>
Acne	self	87.1	83.4	88.3	84.2
	<i>other</i>	<i>82.3</i>	<i>78.4</i>	<i>82.9</i>	<i>80.8</i>
Insomnia	self	83.5	79.4	82.3	79.8
	<i>other</i>	<i>80.7</i>	<i>77.1</i>	<i>81.1</i>	<i>78.9</i>
Back	self	81.8	75.6	77.9	76.4
	<i>other</i>	<i>78.8</i>	<i>74.6</i>	<i>75.9</i>	<i>75.5</i>
Asthma	self	80.2	76.5	76.9	73.8
	<i>other</i>	<i>77.2</i>	<i>73.2</i>	<i>74.8</i>	<i>72.5</i>
Overweight	self	78.3	77.3	76.8	72.3
	<i>other</i>	<i>75.2</i>	<i>73.5</i>	<i>74.9</i>	<i>70.0</i>
Smoking	self	80.4	72.4	74.3	70.4
	<i>other</i>	<i>78.3</i>	<i>73.6</i>	<i>75.9</i>	<i>71.6</i>
Arthritis	self	77.8	74.4	73.8	72.5
	<i>other</i>	<i>74.8</i>	<i>71.4</i>	<i>72.2</i>	<i>71.6</i>
Heart	self	74.7	71.3	70.6	67.8
	<i>other</i>	<i>71.4</i>	<i>68.6</i>	<i>68.6</i>	<i>66.2</i>
MEAN	self	82.7	79.3	80.3	77.8
	<i>other</i>	<i>79.4</i>	<i>76.4</i>	<i>78.2</i>	<i>76.1</i>

Table 5: Mean ratings and discrepancies in these ratings as a function of attribute and disorder, Experiment 3; higher numbers represent worse health.

	Pain	Econ	Work	Love	Family	Spirit	Leisure
Mean disutility ratings							
Asthma	0.49	0.20	0.31	0.22	0.18	0.09	0.42
Insomnia	0.39	0.19	0.41	0.27	0.30	0.14	0.37
Short	0.10	0.13	0.13	0.32	0.12	0.07	0.15
Nearsighted	0.10	0.09	0.09	0.08	0.08	0.07	0.15
Acne	0.22	0.15	0.21	0.59	0.19	0.09	0.24
Smoking	0.33	0.37	0.21	0.31	0.28	0.16	0.29
Arthritis	0.74	0.41	0.60	0.44	0.31	0.11	0.62
Heart	0.64	0.29	0.48	0.38	0.30	0.12	0.52
Mean Not-have–Have discrepancy in disutility							
Asthma	0.08	−0.02	0.06	−0.03	−0.01	−0.07	0.01
Insomnia	−0.05	0.02	0.02	0.17	−0.04	−0.08	0.01
Short	0.03	0.09	−0.07	0.00	0.11	0.05	0.02
Nearsighted	−0.02	−0.04	0.06	0.00	0.00	0.06	−0.04
Acne	−0.01	0.02	0.02	−0.02	0.01	−0.08	0.03
Smoking	−0.08	−0.01	0.03	0.10	0.10	−0.01	−0.02
Arthritis	0.08	−0.02	−0.03	−0.08	−0.02	0.06	−0.14
Heart	0.06	0.11	0.05	0.03	−0.05	−0.01	−0.12

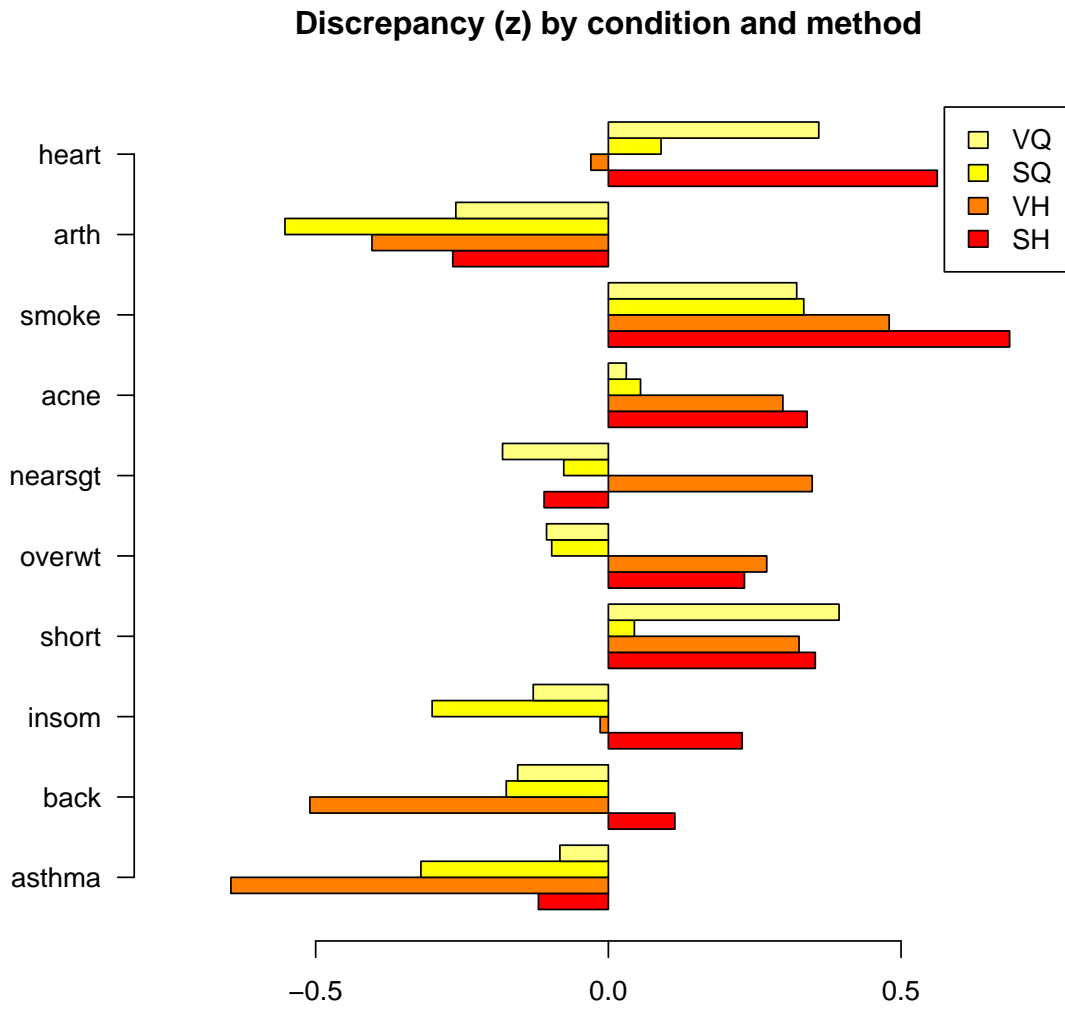


Figure 1: Have–Not-have discrepancy by disorder and method, Experiment 2. Positive numbers indicate that Not-have’s rate the disorder as worse than Have’s.

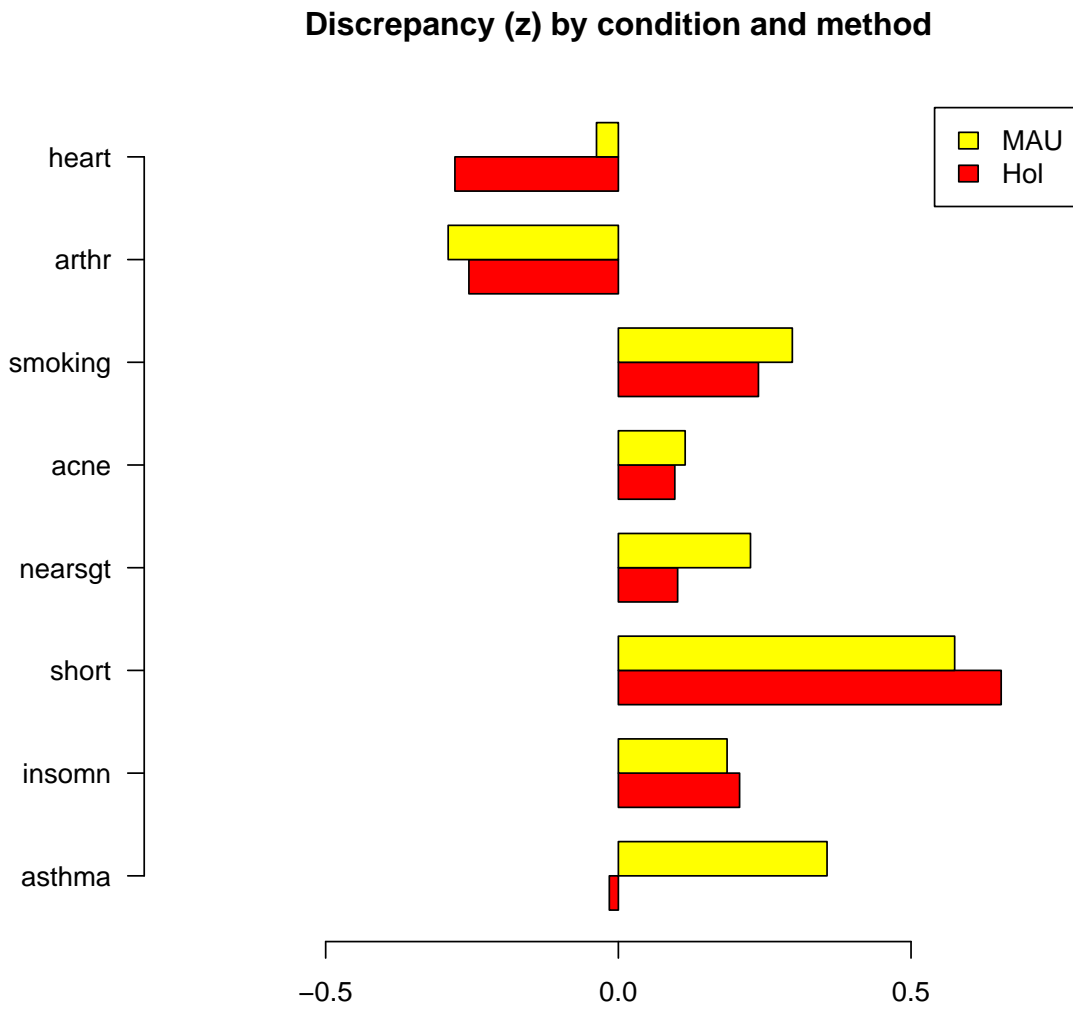


Figure 2: Have–Not-have discrepancy by disorder and method, Experiment 3. Positive numbers indicate that Not-have’s rate the disorder as worse than Have’s.