Accepted Manuscript

On the comparison of group inequalities using subjective data

Xavier Fontaine, Luke Haywood

*Highlights (for review)

We study the validity of comparisons of subjective health across groups.

We use SHARE data on subjective health reports to rank (levels and variances) for individuals with different characteristics.

Using anchoring vignettes we show that differential reporting is much more problematic for comparisons (across groups) of inequality than for comparisons of levels.

This raises questions about recent contributions comparing levels of inequality in subjective well-being across groups.

On the comparison of group inequalities using subjective data

Xavier Fontaine*& Luke Haywood [†]

Abstract

Subjective data offer a simple way of comparing outcomes across populations. However, reporting behavior may differ across groups. We investigate the validity of studies comparing inequality based on such data. Comparing group inequalities appears substantially more problematic than comparing averages.

---

[*]Paris School of Economics
[†]Corresponding author. DIW Berlin, Mohrenstrasse 58, D-10117 Berlin, *LHaywood@diw.de*.

1 Introduction

Subjective data on health, job satisfaction, trust and other values are commonly used in economics, since objective data are difficult to collect in these areas. While comparisons across individuals based on subjective data (e.g. wellbeing and health) are frequent, their validity has also often been called into question. Gross differences in subjective data across groups indeed confound different reporting scales (what "good" health means) and differences in underlying latent values. We here focus on the validity of comparisons of inequality across groups. Such comparisons have become increasingly popular, in line with concerns about high levels of inequality. We use the standard "vignette" method to account for differences in reporting across groups. While we find that comparing mean levels of subjective values across groups is quite informative, comparing levels of inequality seems much less relevant. This raises questions for the recent literature on wellbeing inequality.

Comparisons of levels of inequality across populations based on subjective data is a burgeoning research area. The previous focus of the literature on subjective well-being on comparing means has indeed shifted to considering the whole distribution. Recent contributions that explicitly compare inequality using subjective measures include Fahey and Smyth (2004), Kalmijn and Veenhoven (2005), Ott (2005), Greve (2010), Delhey and Kohler (2011), OECD (2011), Clark et al. (2014), Jordá and Sarabia (2015) and John Helliwell and Wang (2016) [1]. In a prominent recent contribution, Goff et al. 2016 consider the influence of wellbeing inequality on life satisfaction, based inter alia on cross-country comparisons. Representative for this literature, they affirm "[i]n common with other SWL papers, we take interpersonal comparability for granted." The validity of this comparability has however only been tested with respect to comparing group means. This paper is the first to test the validity of comparisons of variances of subjective data

---

[1]For policy documents, see New Economics Foundation (2015) and The Happiness Research Institute (2015). Inequality in wellbeing was analyzed as a factor in Britain's vote to leave the EU (Abdallah (2015)).

across populations.

Several studies show evidence of reporting bias when comparing group means based on subjective data. A first approach proxies latent health with an objective indicator, and focuses on the discrepancies between subjective and objective measures (Murray and Chen 1992, Kerkhofs and Lindeboom 1995, Mathers and Douglas 1998, Lindeboom and van Doorslaer 2004, Etilé and Milcent 2006, Jürges 2007). A second approach uses 'anchoring vignettes' to establish differences in scale interpretation across respondents. Respondents are asked to assess the situation of a hypothetical person on the same response scale they use to evaluate their own situation (Salomon et al. 2004, Kapteyn et al. 2007, d'Uva et al. 2008). This method has been applied to reporting poverty (Beegle et al. 2012), job satisfaction (Kristensen and Johansson 2008), life satisfaction (Kapteyn et al. 2010, Angelini et al. 2014) and income satisfaction (Kapteyn et al. 2013). All researchers find substantial reporting heterogeneity. Unless differences in reporting are uncorrelated with differences in latent values, this will bias comparisons across groups. These contributions consider the role of reporting heterogeneity for comparisons in levels. We here consider the specific issues involved in how differences in reporting may influence comparisons of inequality.

Section (2) extends the traditional anchoring vignette model to include group inequalities. Section (3) describes the subjective health data we use. Section (4) presents results. First, differences in reporting do not invalidate comparisons of group averages. Second, and by constrast, the bias caused by differential reporting makes comparisons of health inequality across populations largely meaningless. Section (5) concludes.

2 Model and Identification

Comparing subjective reports on health or wellbeing across groups raises a number of specific issues. Assume we are interested in comparing underlying latent levels of H, but only observe reports R. Inference from differences in R to differences in H may be difficult: Individuals' subjective reports may be relative to group-

level averages. I may rate my health less positively if I am part of a very healthy group. Group-level reference points thus influence the way I translate H into R.[2] If reporting scales are interpreted differently, reported mean differences in health may not be informative of latent differences across groups. Similarly, health inequality (captured as the variance) may appear larger in a given group only because the health levels associated with the response categories are more distant from each other in one group than another. To reveal how differences in reporting influence comparisons of variance we include heteroskedasticity in the vignette model.

2.1 Group comparisons with reporting heterogeneity

We allow individuals with characteristics X_i to have both different health outcomes and reporting styles[3]. We are interested in the validity of group comparisons. A group is defined in the most flexible possible way: Any set of individuals with the same vector of personal characteristics X_i. Our main focus is the quality of comparisons of mean health, and health inequality, so that our objects of interest are $\mu_{H_s}(X_i) \equiv E(H_{i,s}|X_i)$ and $\sigma^2_{H_s}(X_i) \equiv V(H_{i,s}|X_i)$, where the dependence of $V(.)$ on X_i indicates heteroskedasticity.

Respondent i evaluates her health status on a K-point scale. Let $R_{s,i}$ denote i's evaluation of health in domain s (pain, breathing problems, sleep disorders etc.). This evaluation depends on the genuine and latent health level $H_{s,i}$. Statement k indicates that the severity of i's condition falls in the interval $[\tau_i^{k-1}; \tau_i^k[$ that defines i's interpretation of this category

$$(1) \qquad R_{s,i} = k \Leftrightarrow \tau_i^{k-i} \leq H_{s,i} < \tau_i^k, \quad \forall k \in \{1, \dots, K\}.$$

Our vignette study follows the methodology outlined in King et al. (2004).

---

[2]Note that other group members may also affect the latent variable. For subjective well-being, the importance of these relative effects is well-documented (Kahneman 2000, see also Clark et al. 2008).

[3]In the empirical section, we consider the following individual characteristics: income, age, number of children, gender, education, job status and country.

Importantly, however, and contrary to previous vignette models, respondents' characteristics are allowed to affect not only the expectation, but also the variance of health, which leads to the following model:

$$(2) \quad \begin{aligned} H_{s,i} &= X_i\beta + \sigma_i\varepsilon_{s,i} \\ \sigma_i &= \exp(X_i\kappa) \\ \varepsilon_{s,i} &\hookrightarrow \mathcal{N}(0,1), \quad \varepsilon_{s,i} \perp X_i. \end{aligned}$$

Previous vignette studies implicitly assume $\kappa = 0$. In line with previous work we allow the reporting thresholds to also depend on the respondent's characteristics, using the standard exponential form guaranteeing that the thresholds remain ordered:

$$(3) \quad \begin{cases} \tau_i^0 = -\infty, \ \tau_i^K = +\infty \\ \tau_i^1 = \gamma_0^1 + X_i\gamma^1 \\ \tau_i^k = \tau_i^{k-1} + \exp(\gamma_0^k + X_i\gamma^k), \ \ \forall k \in \{2, \dots, K-1\}. \end{cases}$$

The condition for differences in reporting scale to bias comparisons across groups is fairly simple: Whenever the reporting scale varies systematically across groups with different latent health outcomes - different means for comparisons of the group means, and in terms of variance for the comparison of inequality levels across groups.

2.2 Identification using Vignettes

Individual characteristics X_i may influence both actual health outcomes and the reporting scale. The identification challenge consists in separating two channels: underlying differences in latent health across groups, via β and κ, from differences across groups in the reporting scale, given by γ. We achieve identification through anchoring vignettes. This technique combines subjective self-assessments of health with a subjective report of the health of a hypothetical person (the vignette) described in the questionnaire (King et al. 2004). The vignette model can eliminate response bias under two hypotheses: *Response consistency* asserts that respondents use the K-point response scale in the same way in their self-reports

and for vignette evaluation. *Vignette equivalence* states that only reporting, but not perception of the vignette varies across respondents.

As an illustration, assume individuals from groups A and B rate the level of respiratory distress of a hypothetical person. On average, the scores are higher in A than B. Vignette equivalence implies that this difference only stems from reporting heterogeneity. Response consistency implies that the same bias affects self-reports: Members of A should overrate their own health the same way they overrate the vignette.

Since this is to the best of our knowledge the first paper to introduce heteroskedasticity into a vignette model, we focus on the specific assumptions related to this feature. The perception of the v-th vignette is labeled $H_{v,i}$, where as usual $R_{v,i} = k \Leftrightarrow \tau_i^{k-i} \leq H_{v,i} < \tau_i^k$, $\forall k \in \{1, \dots, K\}$. Respondents share the same understanding of the vignette, up to a random perturbation

$$
\begin{aligned}
H_{v,i} &= \theta_v + \sigma_v \varepsilon_{v,i} \\
\varepsilon_{v,i} &\hookrightarrow \mathcal{N}(0,1), \quad \varepsilon_{v,i} \perp X_i
\end{aligned}
\tag{4}
$$

Estimation of the whole model (1)-(4) requires normalizations of location and scale. The location normalization is standard for vignette models, but the scale normalization is related to our introduction of heteroskedasticity. The econometric literature on heteroskedastic ordered response models usually achieves both at the same time by excluding the constant from X_i (Greene and Hensher (2010)). Note that if there is a constant in the equation of the conditional variance, we can write

$$
\sigma_i = \exp(\kappa_0 + X_i \kappa) = \exp(\kappa_0) \exp(X_i \kappa),
$$

and $\exp(\kappa_0)$ cannot be separately identified from the other coefficients in the reporting model. All the model parameters can then be estimated consistently by maximizing the conditional log-likelihood l_i^c with respect to parameter vectors β, κ, γ, θ_v and σ_v

6

$$(5) \quad l_i^c = \sum_{k=1}^{K} \mathrm{I}(R_{i,s} = k) \times \ln \left[\Phi \left(\frac{1}{exp(X_i\kappa)} \left(\tau_i^k - X_i\beta \right) \right) - \Phi \left(\frac{1}{exp(X_i\kappa)} \left(\tau_i^{k-1} - X_i\beta \right) \right) \right]$$
$$+ \sum_{v=1}^{V} \sum_{k=1}^{K} \mathrm{I}(R_{i,v} = k) \times \ln \left[\Phi \left(\frac{1}{\sigma_v} \left(\tau_i^k - \theta_v \right) \right) - \Phi \left(\frac{1}{\sigma_v} \left(\tau_i^{k-1} - \theta_v \right) \right) \right]$$

with I(.) the indicator function.

3 Data and Specification

This paper uses the Survey of Health, Ageing and Retirement in Europe (SHARE) which asks a large number of respondents about their health outcomes and their assessment of health vignettes. Our sample consists of individuals born before 1954 and of their spouses. Evaluations of own and hypothetical health were collected during the first two waves of the survey which we pool for our analysis. Respondents rate the severity of their health problems on a scale from 1 ('none') to 5 ('extreme') for seven health problems (see table (1)). The dataset contains at least three vignettes for each health domain.

The model is estimated separately on each health domain. We have approximately 12,000 observations for each domain (see table (2) for details). All estimations include sociodemographic control variables: household income per capita (in logs), age, number of children, gender, education, labor market status and country. Due to the large number of estimated parameters, complete estimation results are relegated to an online appendix.

4 Results: Reliability of comparing group means & variances

We now report to what extent rankings of group means and variances are affected by reporting heterogeneity. To do this we compare ranking across groups with and without correcting for reporting heterogeneity. Rankings are a natural way

Table 1: Self-evaluation questions

Domain	Self-evaluation question
Breathing	In the last 30 days, how much of a problem did you have because of shortness of breath?
Sleeping	In the last 30 days, how much difficulty did you have with sleeping such as falling asleep, waking up frequently during the night or waking up too early in the morning?
Pain	Overall in the last 30 days, how much of bodily aches or pains did you have?
Sadness	Overall in the last 30 days, how much of a problem did you have with feeling sad, low, or depressed?
Concentration	Overall in the last 30 days how much difficulty did you have with concentrating or remembering things?
Mobility	Overall in the last 30 days, how much of a problem did you have with moving around?
Work disability	Do you have any impairment or health problem that limits the kind or amount of work you can do?

Table 2: Number of individual contributions to partial likelihoods

	Self-eval	Vignette 1	2	3	4	5	6	7	8	9
Breathing	11932	11833	4407	4415	-	-	-	-	-	-
Sleeping	11954	4412	4409	11850	-	-	-	-	-	-
Pain	11953	11846	4407	4414	-	-	-	-	-	-
Sadness	11927	4417	4404	11822	-	-	-	-	-	-
Concentration	11948	11853	4413	4416	-	-	-	-	-	-
Mobility	11943	4416	4413	11848	-	-	-	-	-	-
Work disability	8902	4395	8820	4390	4377	4376	8820	8818	4396	4397

of testing the validity of interpersonal comparisons - there is no other obvious way to compare ordinal data. Intuitively, we are testing how often one group is actually healthier than another, yet reports lower health levels, viz. actually has more equal health outcomes but reports a higher level of inequality. Formally, we compare

$$\rho_{\text{Spearman}}\left[\mu_{R_s}(X_i), \mu_{H_s}(X_i)\right]$$
$$\rho_{\text{Spearman}}\left[\sigma^2_{R_s}(X_i), \sigma^2_{H_s}(X_i)\right],$$

where $\rho_{\text{Spearman}}(.)$ denotes the Spearman rank correlation coefficient and μ_{R_s} and $\sigma^2_{R_s}$ are group means and variances of health declarations. Both the corrected and the uncorrected means and variances are predicted after estimation of the complete model. The expressions for the predicted latent group mean and variances are simply: $\hat{\mu}_{H_s}(X_i) = X_i\hat{\beta}$ and $\hat{\sigma}^2_{H_s}(X_i) = \exp^2(X_i\hat{\kappa})$. The uncorrected values are predicted parametrically as follows[4]:

$$(b) \qquad \hat{\mu}_{R_s}(X_i) = K - \sum_{k=1}^{K-1} \Phi\left(\frac{\hat{\tau}_i^k - X_i\hat{\beta}}{exp(X_i\hat{\kappa})}\right)$$

$$\hat{\sigma}^2_{R_s}(X_i) = \sum_1^K k^2 \times \ln\left[\Phi\left(\frac{1}{\exp(X_i\hat{\kappa})}\left(\hat{\tau}_i^k - X\hat{\beta}\right)\right) - \Phi\left(\frac{1}{\exp(X_i\hat{\kappa})}\left(\hat{\tau}_i^{k-1} - X\hat{\beta}\right)\right)\right]$$

$$(6) \qquad -\left[\sum_1^K k \times \ln\left[\Phi\left(\frac{1}{\exp(X_i\hat{\kappa})}\left(\hat{\tau}_i^k - X\hat{\beta}\right)\right) - \Phi\left(\frac{1}{\exp(X_i\hat{\kappa})}\left(\hat{\tau}_i^{k-1} - X\hat{\beta}\right)\right)\right]\right]^2$$

We find that comparing means of reported health is informative. Comparing levels of inequality across populations, by contrast, is problematic. The left column of table (3) shows correlations between rankings of group averages based on reported health values compared to rankings of the latent values (i.e. after correction via the vignettes). These correlations appear quite high - with an average

---

[4]For the uncorrected analogues, we may be tempted to directly use reported values. However, first, since X_i contains continuous variables, we would have very few observations in every group, and estimating variances would be infeasible. Second, since we must predict the corrected measures, a comparison between reported (uncorrected) and predicted (corrected) values would conflate prediction errors with the effect of differential reporting we are interested in.

Table 3: Correlations between the conditional expectations / variances of latent and reported health

	Correlations of corrected & uncorrected means	Correlations of corrected & uncorrected variances
Breathing	0.674	0.466
Sleeping	0.898	0.219
Pain	0.901	0.298
Sadness	0.903	0.402
Concentration	0.745	0.697
Mobility	0.963	-0.068
Work disability	0.946	0.282

ρ_{Spearman} of 0.861, close to 0.95 in three cases (sleep, mobility, work disability), and below 0.8 only for breathing (0.674) and concentration (0.754). Rankings based on group averages of reported health are mostly a good proxy of the true rankings and can be considered as a valuable tool for research.

By including heteroskedasticity in the vignette model, we can now also consider the impact of reporting heterogeneity on comparisons of variances. The second column correlates within-group variances in reported health with variances in latent health (i.e. after the vignette correction). The correlations are much lower - on average the correlation coefficient is 0.328, and close to zero for mobility (concentration is the only domain where the correlation is above 0.5). Variances in reported health are a weak and unreliable proxy for variances in true health.[5] Comparing variances of self-assessed health or well-being appears very challenging in light of these results.

---

[5]What is the cause of this lack of reliability? Additional analysis – not reported here in the interest of brevity – reveals significant correlations between the variance of latent health and the spaces between reporting cut-offs, giving rise to the current pattern.

5 Conclusion

Reporting heterogeneity is a pervasive issue in subjective data. Previous methodological work has focused on the validity of comparisons of means. This article tests the effect of including differences in reporting of variances, in line with a recent literature comparing inequality across groups. To do this, we add heteroskedasticity to a standard anchoring vignette approach: Different individuals' assessments of the health of common vignettes (health scenarios) are used to identify two types of reporting bias of subjective reports, covering the mean and the variance of the scale used to translate latent to reported health. The analysis is performed using answers to questions about seven health domains. Our results show that whilst reporting heterogeneity biases all comparisons across groups, the effect is small for comparisons of means but substantial for comparisons of health inequalities. This raises questions about the validity of comparisons of inequality based on subjective data which do not correct for individual differences in reporting.

References

ABDALLAH, S., *The one graph that could explain Brexit* (New Economics Foundation, 2015).

ANGELINI, V., D. CAVAPOZZI, L. CORAZZINI AND O. PACCAGNELLA, "Do Danes and Italians Rate Life Satisfaction in the Same Way? Using Vignettes to Correct for Individual-Specific Scale Biases," *Oxford Bulletin of Economics and Statistics* 76 (2014), 643–666.

BEEGLE, K., K. HIMELEIN AND M. RAVALLION, "Frame-of-reference bias in subjective welfare," *Journal of Economic Behavior & Organization* 81 (2012), 556–570.

CLARK, A., P. FRIJTERS AND M. SHIELDS, "Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles," *Journal of Economic Literature* 1 (2008), 95–144.

CLARK, A. E., S. FLÈCHE AND C. SENIK, "The great happiness moderation," in A. Clark and C. Senik, eds., *Happiness and economic Growth: Lessons from Developing Countries* (Oxford University Press, 2014).

DELHEY, J. AND U. KOHLER, "Is Happiness Inequality Immune to Income Inequality? New Evidence through Instrument-Effect-Corrected Standard Deviations," *Social Science Research* 40 (2011), 742–756.

D'UVA, T. B., E. V. DOORSLAER, M. LINDEBOOM AND O. O'DONNELL, "Does reporting heterogeneity bias the measurement of health disparities?," *Health Economics* 17 (2008), 351–375.

ETILÉ, F. AND C. MILCENT, "Income-related reporting heterogeneity in self-assessed health: evidence from France," *Health Economics* 15 (2006), 965–981.

FAHEY, T. AND E. SMYTH, "Do subjective indicators measure welfare? Evidence from 33 European Societies," *European Societies* 6 (2004).

12

GOFF, L., J. HELLIWELL AND G. MAYRAZ, "The Welfare Costs of Well-Being Inequality," *NBER Working Paper* 21900 (2016).

GREENE, W. AND D. HENSHER, *Modeling ordered choices: A primer* (Cambridge University Press, 2010).

GREVE, B., *Social Policy and Happiness in Europe* (Edward Elgar Publishing, 2010).

JOHN HELLIWELL, H. H. AND S. WANG, "The distribution of world happiness," in R. L. John Helliwell and J. Sachs, eds., *World Happiness Report Update 2016* (Sustainable Development Solutions Network, 2016).

JORDÁ, V. AND J. SARABIA, "Well-Being Distribution in the Globalization Era: 30 Years of Convergence," *Applied Research Quality Life* 10 (2015), 123–140.

JÜRGES, H., "True health vs response styles: exploring cross-country differences in self-reported health," *Health Economics* 16 (2007), 163–178.

KAHNEMAN, D., "Experienced Utility and Objective Happiness: a Moment-Based Approach," in E. Diener and N. Schwarz, eds., *Well-being: The foundations of hedonic psychology* (New York: Russell Sage Foundation, 2000).

KALMIJN, W. AND R. VEENHOVEN, "Measuring Inequality of Happiness in Nations: In Search for Proper Statistics," *Journal of Happiness Studies* 6 (2005), 357–396.

KAPTEYN, A., J. P. SMITH AND A. VAN SOEST, "Vignettes and Self-Reports of Work Disability in the United States and the Netherlands," *American Economic Review* 97 (March 2007), 461–473.

———, "Life Satisfaction," in E. Diener, J. Helliwell and D. Kahneman, eds., *International Differences in Well-Being* (Oxford University Press, 2010), 70–104.

———, "Are Americans Really Less Happy with Their Incomes?," *Review of Income and Wealth* 59 (2013), 44–65.

KERKHOFS, M. AND M. LINDEBOOM, "Subjective health measures and state dependent reporting errors," *Health Economics* 4 (1995), 221–235.

KING, G., C. MURRAY, J. SALOMON AND A. TANDON, "Enhancing the validity and cross-cultural comparability of measurement in survey research," *American Political Science Review* 98 (2004), 191–207.

KRISTENSEN, N. AND E. JOHANSSON, "New evidence on cross-country differences in job satisfaction using anchoring vignettes," *Labour Economics* 15 (February 2008), 96–117.

LINDEBOOM, M. AND E. VAN DOORSLAER, "Cut-point shift and index shift in self-reported health," *Journal of Health Economics* 23 (2004), 1083 – 1099.

MATHERS, C. AND R. DOUGLAS, "Measuring progress in population health and wellbeing," *Measuring progress: Is life getting better* (1998), 125–155.

MURRAY, C. J. AND L. C. CHEN, "Understanding morbidity change," *The Population and Development Review* (1992), 481–503.

NEW ECONOMICS FOUNDATION, *Inequalities in wellbeing - Challenges and opportunities for research and policy* (2015).

OECD, *How's Life? Measuring Well-Being* (OECD Publishing, 2011).

OTT, J., "Level and inequality of happiness in nations: Does greater happiness of a greater number imply greater inequality in happiness?," *Journal of Happiness Studies* 6 (2005), 397–420.

SALOMON, J. A., A. TANDON AND C. J. MURRAY, "Comparability of self rated health: cross sectional multi-country survey using anchoring vignettes," *British Medical Journal* 328 (2004), 258–261.

15

THE HAPPINESS RESEARCH INSTITUTE, *Happiness Equality Index Europe 2015* (mimeo, 2015).