Coordinating Harmonized Phenotyping

David Cesarini
Olga Rostapshova

3rd SSGAC Workshop • May 19, 2012

Challenges and Approaches

- Challenge: Obtain large sample of GWAS + phenotype data
  - Solution #1: Genotype individuals already phenotyped.
  - Solution #2: New phenotyping of individuals already genotyped.
- New Phenotyping:
  - Low cost, esp. if new surveys planned.
  - Harness SSGAC social science expertise.
- For gene discovery, will take several years for large enough sample.

How SSGAC Can Help

- Have extensive experience designing social-science survey measures
  - e.g. WLS, SALTY, RS, AGES, HBCS, B2ESS.
  - Developed template measures of phenotypes.
  - Facilitate harmonization.
  - Can work with cohorts to prioritize, adjust, and implement these measures.
- Working on getting funds to be able to offset costs for new surveys.
- Efforts coordinated by Olga Rostapshova.

Criteria for an Ideal Phenotype

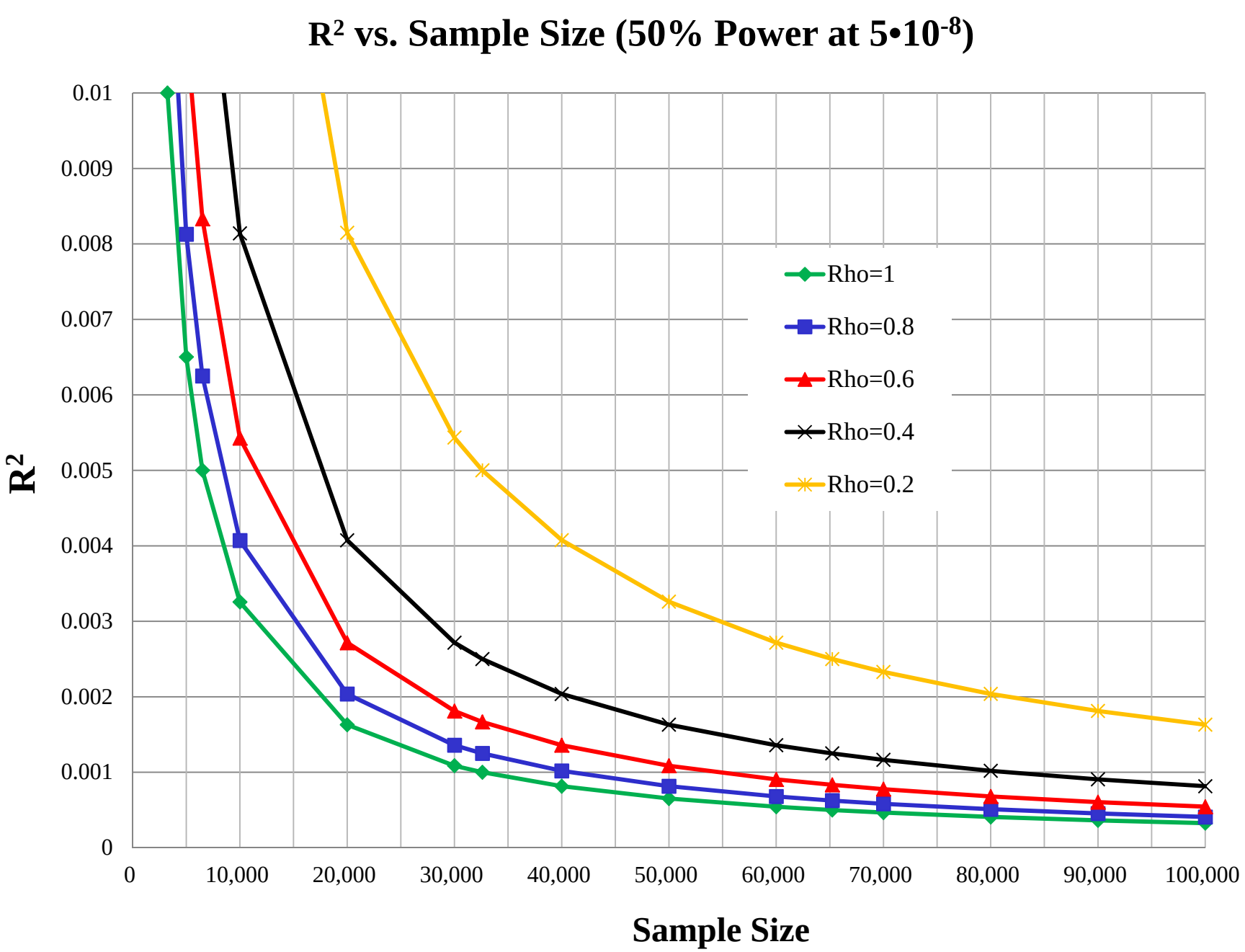
- Interesting to researchers in relevant domain.
- Biologically proximate.
- Meta-analysis friendly:
  - Widely-measured in GWAS cohorts.
  - Maximize overlap with other genotyped cohorts (STR, HRS, WLS).
  - Easy and inexpensive to administer for cohorts.
  - Questions must be easy and natural for the respondents.

Psychometric Properties of Ideal Phenotype

- Strong construct validity.
- Predictive of behavior in the relevant domains (medical outcomes, smoking, drinking).
- Reliably measured.

Tradeoff: Sample Size vs. Reliability

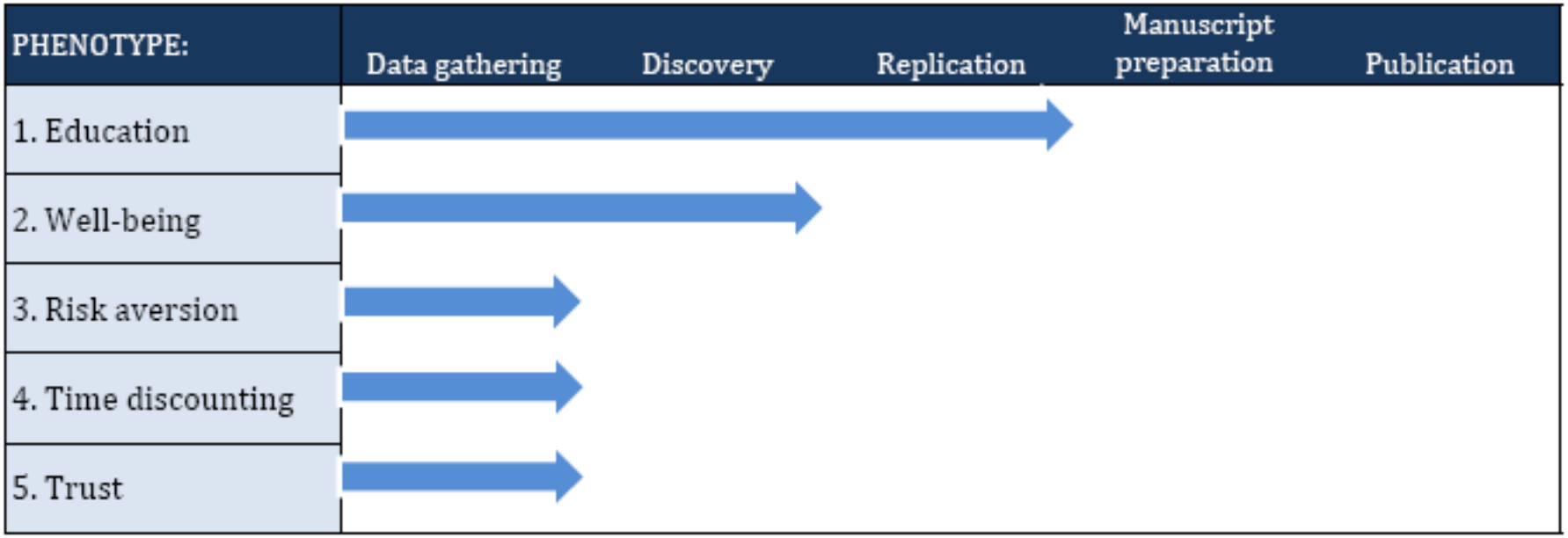
- In practice, tradeoff between sample size and reliability of measures.
- Strive for highest reliability measures conditional on survey instrument limitations.
- Calculations suggest that as a practical matter large samples should be the priority.


R² vs. Sample Size (50% Power at 5•10-8)
0.01
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0
R2
0
10,000
20,000
30,000
40,000
50,000
60,000
70,000
80,000
90,000
100,000
Sample Size
Rho=1
Rho=0.8
Rho=0.6
Rho=0.4
Rho=0.2

Prioritized Phenotypes

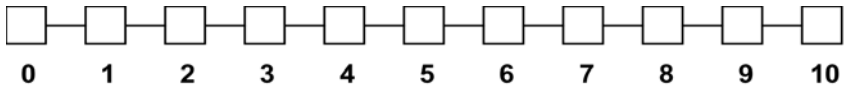
- Risk aversion
- Time discounting
- Trust
- Well-being
- Education

Timeline

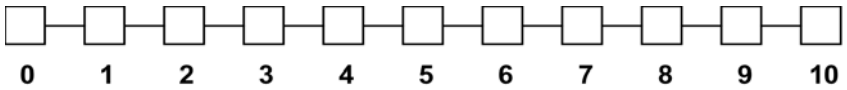

PHENOTYPE:
Data gathering
Discovery
Replication
Manuscript preparation
Publication
1. Education
2. Well-being
3. Risk aversion
4. Time discounting
5. Trust

Risk Preferences (I)

1. "How do you see yourself: are you generally a person that is fully prepared to take risks or do you try to avoid taking risks? Please tick on the scale below, where the value 0 means ”not at all willing to take risks” and the value 10 means ”fully prepared to take risks”.



2. Are you a person that is fully prepared to take financial risks or do you try to avoid taking financial risks? Please tick on the scale below, where the value 0 means ”unwilling to take risks” and the value 10 means ”fully prepared to take risks”.



Risk Preferences (II)

"Imagine the following hypothetical situation. You are the sole provider of your household and you have the choice between two equally good jobs. The incomes I will mention should be considered as net income (after all taxes and social security).

> **Job A** will with certainty give you EUR 2,800 per month for the rest of your life.
>
> **Job B** will give you a 50-50 chance of EUR 5,600 per month for the rest of your life, and a 50-50 chance of EUR 2,240 per month for the rest of your life. Which job do you choose?"

→Responses can be used to construct bounds on an individual's coefficient of relative risk aversion.

Well-Being Measures

- All things considered, how satisfied are you with your life as a whole nowadays?
- Overall, how satisfied are you with your life nowadays?
- Overall, how happy did you feel yesterday?

Time Preferences (“Patience”)

Now please consider the following hypothetical scenarios.

…

Imagine the following hypothetical situations. For each of the choices below, which option would you prefer?

Would you rather have:

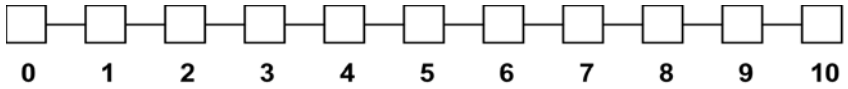
(i) 150,000 ISK now –or-

(p) 151,000 ISK a month from now?

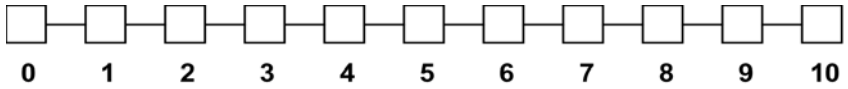
→ Follow-up questions of similar type with different money amounts, depending on the answer to above

Trust

1. Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people? Please tick on the scale below, where the value 1 means ‘need to be very careful’ and the value 10 means ‘most people can be trusted.’



2. Do you think that most people would try to take advantage of you if they got the chance, or would they try to be fair? Please tick on the scale below, where the value 0 means ‘would take advantage of me’ and the value 10 means ‘would treat me fairly.’



Evidence on Heritability

- Trust $h^2 \sim 0.30$ (Cesarini et al., 2008)
- SWB $h^2 \sim 0.50$ (Lykken & Tellegen, 1996; Bartels & Boomsma, 2009)
- Risk $h^2 \sim 0.50$ (Cesarini et al. 2009a, Beauchamp et al. 2011, Zhong et al. 2009)
- Discounting $h^2 \sim 0.40$ (Anokhin et al. 2011, Cesarini et al., 2011.)

Phenotypic Evidence

- Evidence that SWB is associated with immune function and health status (Marsland, Pressman, & Cohen, 2007).

- Discount parameters related to alcoholism risk, substance use, ADHD, BMI, frequency of exercise (Barkley et al. 2001, Chabris et al., 2010; Vuchinich and Simpson 1999)

- Risk parameters related to occupational choice, smoking, drinking, self-rated health, investment decisions (Barsky et al. 1997; Beauchamp et al. 2011)

Case Study on Risk Preferences (Beauchamp et al. 2012)

- Use data collected from a sample Swedish twins to study the psychometric properties of some widely used survey-based measures of economic risk attitudes.
  - Sample size is ~11,000.
  - Retest data available for 500 subjects.

- Throughout, we adopt a uniform and fairly general latent variable framework which easily accommodates measurement error. Study:

  - Reliability
  - Predictive validity in domains of health and finance
  - Characterize the empirical correlates of measured risk preferences

Risk Questions

Risk General and Risk Financial (Dohmen et al. 2006; forthcoming)

How do you see yourself: Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? Please tick a box on the scale, where the value 1 means: unwilling to take risks and the value 10 means: fully prepared to take risks.

Risk Financial is constructed from a similarly phrased question.

Risk HRS (Barsky et al. 1997, Sahm, 2007 and Kimball, Sahm and Shapiro, 2008).

Similar to the hypothetical gambles used in the Health and Retirement Survey. Individuals were asked to respond to a series of questions about a guaranteed monthly salary of SEK 25,000 for the rest of their lives or a gamble in which there is a 50-50 chance of earning either SEK 50,000 or SEK X for the rest of their lives.

Risky Behaviors

Smoking

- We classified as a smoker individuals who indicated that they at some point in their life smoked regularly or used to smoke regularly.

Alcohol Consumption

- We classified as drinkers those individuals who report that they drunk strong beer, wine or liquor more than twice during the last month.

Equity Share

- Self-reported data on fraction of assets held in equities.

Portfolio Risk.

- Riskiness of the individual's pension savings accounts introduced in 2000.

Own Business

- Response to the question "Have you ever started your own business?"

Basic Framework

For each of the three ordinal risk attitudes measures in our data, we model the permanent component of risk attitudes ρ^*as a continuous. We assume that depends on the vector of covariates x in the following way:

$$\rho' = x\beta + \alpha'.$$

ρ' is also subject to an additive disturbance m_t which is independent over time,

$$r_t{}' = \rho' + m_t.$$

Finally, $r_t{}'$ is not observed. Instead we observe r_t which is assumed to be monotonically related to $r_t{}'$ so that,

$$r_t = 0 \quad if\ r_t{}' < \tau_1$$

...

$$r_t = K \quad if\ r_t{}' < \tau_{K+1}$$

We assume that the disturbances have mean zero and are normally distributed and fix the variance of the latent variable. Estimation is by ML.

Using the Framework

- In the context of our latent variable model, a natural analogue to reliability with the variables X partialled out is,

$$R_X = corr(r_1', r_2'|x) = corr(\alpha' + m_1, \alpha' + m_2) = \frac{var(\alpha')}{var(\alpha') + 1}$$

- Also use our framework and an estimator proposed by Kimball, Sahm and Shapiro (2008) in order to
  - Consistently measure the ‘’effects’’ of the permanent component of risk attitudes, accounting for two types of measurement error.
  - Estimate the fraction of variation in risky behaviors accounted for by the permanent component of the risk variables.

Reliability of Best Risk Questions

No Covariates		s.e.	ln(L)	n_1	n_2
Risk General	0.633	0.029	-24617.69	11244	494
Risk Finance	0.670	0.026	-23423.38	11279	494
Risk HRS	0.587	0.038	-13594.69	10583	472
Covariates					
Risk General	0.610	0.027	-24485.90	11244	494
Risk Finance	0.640	0.024	-23191.82	11279	494
Risk HRS	0.572	0.036	-13419.21	10583	472

Predictive Utility of the Risk Variables

- One standard deviation increase in permanent risk attitudes is associated with increase in smoking and drinking probabilities of 3 percentage points.
- One standard deviation increase in permanent risk attitudes is associated with increase in the probability of having run a business by ten percentage points.
- Incremental R^2 increases by between 30 and 100% after accounting for measurement error.

Sibling Correlations

	Risk General		Risk Finance		Risk HRS	
Error?	No	Yes	No	Yes	No	Yes
$corr_{MZ}$	0.354***	0.513***	0.305***	0.433***	0.424***	0.698***
s.e.	(0.031)	(0.053)	(0.031)	(0.049)	(0.036)	(0.074)
# pairs	1128	1128	1134	1134	1000	1000
$corr_{DZ}$	0.141***	0.216***	0.176***	0.249***	0.192***	0.311***
s.e.	(0.028)	(0.042)	(0.030)	(0.045)	(0.037)	(0.064)
# pairs	1200	1200	1212	1212	1083	1083

* Significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Findings Relevant to Our Effort

- Reliabilities are around ~0.60 for preferred questions.
  - Implies power of around 50% to detect marker with $R^2 = 0.1\%$ in a sample of 50,000.
- Heritabilities in existing literature are probably significantly downward biased after
- Proportion of variance in risky behaviors explained rises significantly after measurement error adjustments

Cohorts Collecting New Phenotypes

- Austrian Stoke Prevention Study: January 2013
- Croatian National Bio-Bank: Nov 2011
- Finland HBCS: 2012
- Iceland AGES: February 2012
- Rotterdam study: January 2012
- Singapore, B2ESS: March 2012

Phenotyping Cohorts

- Already have phenotypes:
  - HRS (risk, some discounting)
  - SALTY (risk, discounting, trust)
- On track for a sample of: 30,000-40,000
- Our ambition is to conduct a GWAS with 50,000
  - Replication with data that is available subsequently