Health ROI as a measure of misalignment of biomedical needs and resources

Lixia Yao[1], **Ying Li**[2], **Soumitra Ghosh**[3], **James A Evans**[4,5], and **Andrey Rzhetsky**[5,6]

[1]Department of Software and Information Systems, the University of North Carolina at Charlotte, Charlotte, North Carolina, USA

[2]Department of Biomedical Informatics, Columbia University, New York, New York, USA

[3]Department of Genetics, GlaxoSmithKline R&D, King of Prussia, Pennsylvania, USA

[4]Department of Sociology, the University of Chicago, Chicago, Illinois, USA

[5]Computational Institute, the University of Chicago, Chicago, Illinois, USA

[6]Departments of Medicine and Human Genetics, Institute for Genomics and Systems Biology, the University of Chicago, Chicago, Illinois, USA

To the Editor

With a growing demand and a diversity of health problems on the one hand and finite resources on the other, scientists and funding agencies are faced with difficult choices about which conditions to study and cure. In this Correspondence, we liken investment in understanding disease and discovering remedies to trades in a financial market. Research attention and funding, as traced by articles, grants and clinical trials, constitute subjective 'prices' that scientists and society pay for research on disease-specific therapies. In finance, the Black-Scholes-Merton model[1,2] was introduced as a tool to estimate the intrinsic 'value' of stock options and guide capital investment. Here, we present a health research opportunity index (health ROI) to measure the misalignment of biomedical needs and resources. Our health ROIs suggest where greater returns on investment in health research could be obtained for society.

There are many reasons why a disease might remain under-researched, even if it poses a substantial health burden. If a disease is not perceived as scientifically interesting, or there seems little chance of finding a viable or marketable therapy, scientists and funders may neglect it, with the outcome that health burden becomes decoupled from research investment. We argue that society can improve its allocation of resources to R&D of new treatments by systematically considering the human benefits that would be gained from cure. Other studies have consistently identified an imbalance between health needs and biomedical investments, suggesting that we improve population health by bringing them

closer[3–7]. Here, we expand on that work by, first, validating a new, insurance-based measure of health burden that enables automatic evaluation of burden and research investment for many more diseases than have been previously assessed; and, second, by developing a transparent index (health ROI; Box 1) that measures imbalances between the optimal and actual resource allocation for individual diseases and for all diseases in aggregate. Using our health ROI, we uncover a substantial imbalance between US health needs and research investment.

Box 1

Method for calculating health ROI

To construct a disease-specific health research opportunity index (ROI), we first define a normalized disease-specific variable, X_{nd}. Let Y_{nd} be the raw measurement m for disease d where all measures constitute the set M and all diseases constitute the set D. For example, in this analysis we denote Y_{bd} the disease burden, Y_{rd} the research coverage in the scientific literature, Y_{cd} the clinical trial coverage and Y_{fd} the received funding. A normalized disease-specific X_{nd} is obtained by scaling the disease-specific value of the raw variable by the sum of the corresponding variable over all diseases:

$$X_{nd} = \frac{Y_{nd}}{\sum_{d \in D} Y_{nd}} \quad (1)$$

X_{nd} is a positive number between 0 and 1. Then a disease-specific health ROI can be defined as:

$$ROI_d = \log_{10}\left(\prod_{n \in \{M-b\}} \frac{X_{bd}}{X_{nd}} \right) \quad (2)$$

where X_{bd} is a relative disease burden defined as in equation (1). Thus ROI_d of 0 suggests perfect balance, and thus no opportunity. A ROI_d value of greater than 0 indicates a research opportunity (i.e., disease d received less than optimal allocation of attention and resources compared with other diseases). A ROI_d smaller than 0 indicates that resources devoted to the disease exceed its share of disease burden.

Finally, we define an overall health research opportunity index for all resource allocation as:

$$ROI = \sum_{d \in D} \left[X_{bd} \times \log_{10}\left| 10^{ROI_d} - 1 \right| \right] \quad (3)$$

> Like the disease-specific health ROI, larger overall health ROI values indicate less optimal allocation of resources and greater opportunity for research efficiency through reallocation. Additional details are provided in the Supplementary Methods.

Previous empirical investigations have linked disease burden to a particular resource (e.g., US National Institutes of Health (NIH) funding and publications) for a much narrower set of diseases than we investigate in this Correspondence. For example, Gross *et al.*[3] and Gillum *et al.*[4] examine the relationship between NIH funding and a range of disease burden measures for 29 common conditions in 1999 and 2011. Both studies demonstrate a moderate correlation between NIH funding and disease burden, as measured by disability-adjusted life years (DALYs). HIV/AIDS, dementia, diabetes mellitus, breast cancer and perinatal conditions all received more funding from NIH than would be expected based on US disease burden, whereas depression, injuries and chronic obstructive pulmonary disorder were relatively underfunded. Sampat *et al.*[5] later undertook an expanded study of 107 diseases and found a statistically significant positive relationship between clinical trial funding and deaths or days in hospital. In another study, Vanderelst *et al.*[6] specifically looked at data from low-income countries and found a weak association ($r=0.1$) between disease burden and funding. Work by Evans *et al.*[7], however, found research attention and disease burden aligned within countries, but not at the global scale. These studies disagree over the degree to which health need and investment align, but they examine it for only a small number of disorders, assessed infrequently.

Here, we use well-established health impact assessment measures to capture the burden of disease, including mortality rate, DALYs, years lived with disability (YLDs) and years of life lost (YLLs)[8]. The World Health Organization's (Geneva) most recent and complete global burden of disease estimates (2012) provide YLDs, YLLs and DALYs for only 128 distinct conditions (http://www.who.int/healthinfo/global_burden_disease/estimates/en/index2.html). These assessments are made inconsistently, on average twice per decade. Moreover, estimation of traditional health burden measures, such as DALY, is expensive. DALYs lost to disease equal the sum of that disorder's YLLs and YLDs, where YLDs for colon cancer represent the product of incident cases, their average duration and a qualitatively determined disability weight that reflects the severity of the disease, ranging from perfect health to major discomfort (e.g., life with a colostomy) to death.

Table 1 presents the prevalence and treatment cost of different diseases as measured by insurance claim records. It shows that both disease prevalence and treatment cost from insurance claims correlate positively and significantly with YLDs and DALYs, suggesting that they can be used as a large-scale proxy for DALYs. By using disease prevalence and treatment cost we are able to simultaneously compute and analyze the health ROI for almost 1,400 medical conditions every year over a 12-year time period (2000–2011) using claims data from >100 million patients aged 65 and under in the United States. To analyze the relationship between health ROI for populations over 65, we also calculate health ROI using >7 million inpatient stays in ~1,000 US hospitals and diagnosis data from a private US research hospital (Supplementary Data).

We sought to create an index that compares the balance of burden and resources for each disease and combines these factors into a transparent aggregate measure that allows comparison across thousands of conditions. All previously mentioned studies[3–7] use regression models to assess the overall balance between burden and resources. As a result, the balance for any one condition is a complex, indirect by-product of the total model (i.e., a sum of the average relationship between burden and resources plus a 'residual' or model error associated with that particular condition). Like regressions of health research on needs, interpretation of our health ROI measure is based on the assumption that to achieve maximal societal benefit, resources should be allocated across the full distribution of illness proportional to the costs those illnesses impose on society[9]. The health ROI, however, imposes fewer statistical assumptions on the data[10] to create an easily computable measure that we use here to integrate multiple factors on >100 million US patients under 65 across 1,400 medical conditions (see Supplementary Data and Box 1 for data and methods). Higher disease-level ROI values suggest that less research has been performed on an important illness and so a substantial research opportunity exists. Lower values suggest more research, smaller need and a diminishing opportunity to improve societal health. The overall, aggregate health ROI drops as research across all diseases comes to better align with the distribution of needs (Table 2).

The apparent imbalance between disease burden and disease-specific resource allocation during 2000–2011 is depicted in Figure 1, where burden is measured by overall treatment cost (the same goes for the rest of this article if not specified). We found that the ROI for "breast cancer" steadily diminished in value between 2000 and 2011 (Fig. 1b) as the disease attracted more resources while the suffering from the disease decreased. The ROI for "extremity atherosclerosis," by contrast, consistently increased between 2000 and 2011, suggesting a greater research opportunity as the disease prevalence grew and/or research diminished (Fig. 1d). A few conditions demonstrated fluctuations in ROI over the same period. For example, "palpitation" and "spondylosis without myelopathy" had substantial ROIs in 2000, which increased steadily until 2005 before collapsing in 2011 (Fig. 1d). This suggests that the burden and/or research associated with these disorders are changing rapidly. Table 3 lists the top ten over- and understudied conditions in 2011.

By calculating the overall health ROI, which measures the imbalance of resources across needs for all conditions in a given year, we observe that the measure consistently decreased from 2000–2011, indicating a trend toward overall improvement in resource allocation relative to disease burden (Table 2). The best and most recent alignment, however, was far from optimal—the smallest observed value was 1.6, where it has hovered between 2007 and 2011. Overall health ROI, however, can be negative infinity when all disease-specific ROIs become approximately zero. At the average rate of improvement (2000 to 2011, about 2.6% per year), it would take decades to shift resources to perfectly reflect health needs.

Like return on investment in accounting, the *h*-index in scientific citation and the Black-Scholes-Merton model in finance, our proposed health ROI enables us to easily interpret and compare resource allocations across many diseases with a single indicator. It also imposes fewer assumptions on the data than imbalance assessments derived from regression models[4,10]. As we retain original values for individual components (publications, burden and

clinical trials), we can trace the influence of each factor on the aggregate index. For example, the high ROI for breast cancer is mediated by the exceedingly large proportion of clinical trials devoted to breast cancer therapies (Fig. 2a).

When we examine the health ROI and explore the relationships between disease burden, publications and clinical trials that underlie it, we observe that each variable displays substantial, expected "inertia" (Supplementary Fig. 1); each year-specific value is tightly correlated with values of the same variable for earlier years, and correlations decline with the difference in time between the two measurements. Unexpectedly, disease burden is not correlated with publications or clinical trials in previous years. This suggests that the negative feedback between clinical trials, scientific studies and disease burden operates on a longer timescale than we are able to observe in this study. Other possibilities are either that no feedback loop exists to realign resource allocation with disease burden or that the relationship is complex and mediated by other variables. For example, if researchers' attention is influenced by exposure to health problems that appear in their hospitals and clinics, then the biased distribution of health problems seen at tertiary-care research hospitals might play a role in keeping need and research apart.

Scientific publications (Supplementary Table 2) show no correlation with past or present disease burden, but have a weak, positive correlation with prior clinical trials (Supplementary Fig. 1). Similarly, clinical trials correlate with past research publications and disease burden; both relationships are stronger for publications and burden in the more distant past. This last observation is likely to arise from the inertia of clinical trials, which require lengthy preparation, execution and substantial funding. As a result, outputs of present-day clinical trials appear to be motivated by scientific arguments a decade old or older (Supplementary Fig. 1).

We then examined the degree to which research funding aligned with disease burden, disease-specific publications and clinical trials. To investigate this, we selected 83 diseases with documented NIH funding data and explicit disease name mapping between dollars and disease. We then analyzed correlations between the relevant variables for those selected diseases during 2003 and 2011 (Supplementary Table 1). There is no significant linear or nonlinear correlation between disease burden and NIH funding. This is surprising and differs from previous research, like that of Gillum *et al.*[4], which examined the correlation between 2004 burden and 2006 NIH funding for just 29 disorders. Instead, shifts in the research literature seem to underpin funding allocated by the NIH, and associated shifts in funding recursively focus research. Year- and disease-specific NIH funding is highly correlated with corresponding clinical trials (see Supplementary Table 1). Strong, positive correlations between the diseases addressed by drugs in trial over time suggest that each clinical trial represents a long-term financial commitment, which increases the likelihood of follow-up trials. Finally, funding is significantly negatively correlated with disease-specific ROIs, but this correlation weakens and becomes nonstatistically significant over time. In summary, using the ROI measure, we observe an imbalance between US research funding for medical conditions and disease burden in the US population under age 65.

We next used the National Inpatient Samples (NIS) and University of Chicago Hospital data to evaluate whether these findings extend to US populations over 65. First, we rank-ordered diseases in terms of the research opportunity they represent (ROI) for younger (65 and under) and older (over 65) populations. There is a high, statistically significant correlation between these ROI ranks (Fig. 2b,c). Figure 2 and Supplementary Tables 3 and 4 illustrate the differences between those diseases that disproportionately afflict older and younger patients. Diseases that incur high costs to patients 65 and under, but virtually none to the elderly include disorders of reproduction and development (see the 'island' of isolated diseases beneath the positive diagonal in Fig. 2c). On the other hand, degenerative disorders like dementia and cancer are more common and costly for older patients, but still exist among those 65 and under (see the bulge above the positive diagonal in Fig. 2b,c).

We found that whereas differences between health conditions distinctive to younger and older populations exist, the overall pattern of disease burden is similar. When we evaluated the relationship between disease burden for NIS patients over 65 and publications or clinical trials, we found no significant correlation between disease burdens and articles or clinical trials. When we tested these relationships for NIS patients 65 and under, however, the Spearman rank correlations were slightly negative (−0.099, −0.092) and significant, suggesting that disorders more burdensome actually receive less research attention and clinical development. Collectively, this demonstrates that our core finding about the disconnect between disease burden and published research holds when patients over 65 are included in our analysis. It also suggests that imbalance is slightly worse for the health needs of those 65 and below.

Our analyses provide evidence that resource allocation dynamics are influenced by previous research and allocations far more than by current health needs, resulting in a massive imbalance between US health needs and research investments. This dynamic may arise partly owing to research trends in which a biomedical breakthrough or celebrity illness tips funding toward one disease at the expense of others. A trendy approach may not always perform worse than the proportional assumption underlying our health ROI, as trends could be more sensitive to currently advancing research areas. This raises the question of whether our proportional assumption is the best baseline against which to establish research opportunity. One could imagine, for example, a rational bottleneck approach, where the biomedical research establishment estimates costs and benefits associated with current studies and then concentrates funding toward 'bottleneck' areas that promise the most total benefit to society. Focusing funding on a few bottleneck conditions would support redundant research, but possibly facilitate a faster accumulation of advances. Although we are unable to test our proportional assumption in this Correspondence, if an alternative, more sophisticated cost function (e.g., the bottleneck) proves more effective than our proportional assumption for identifying research opportunity, the health ROI could incorporate it.

The ROI could be modified to account for different time intervals (2, 5 and 10 years), which would smooth trends, reducing the effect of random fluctuations. Additional measures of health need and research investment could easily be incorporated, just as we extended the frequency of observations and number of diseases available for consideration with DALYs by using health insurance records. We envision that we and others will compute and

compare health ROIs for different groups in society (e.g., wealthy and poor; urban and rural), for privately funded research and for data from different countries. We recognize that by restricting analyses to the United States, many globally burdensome conditions are disregarded, including neglected tropical diseases[11]. We hope that using the health ROI can assist biomedical researchers, private funders and governments to improve the value of health research for society.

Supplementary Material
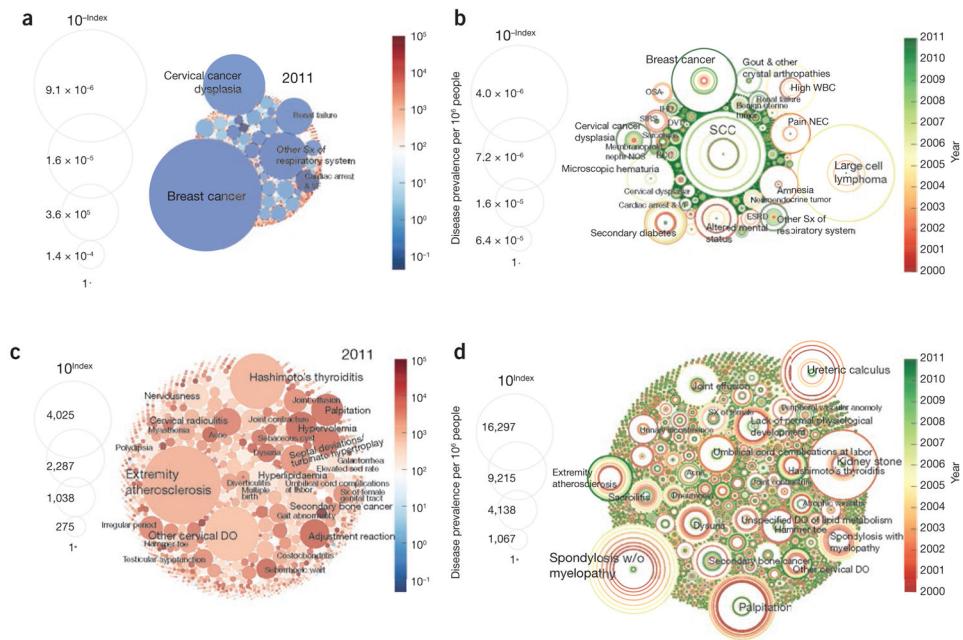
Refer to Web version on PubMed Central for supplementary material.

Acknowledgments
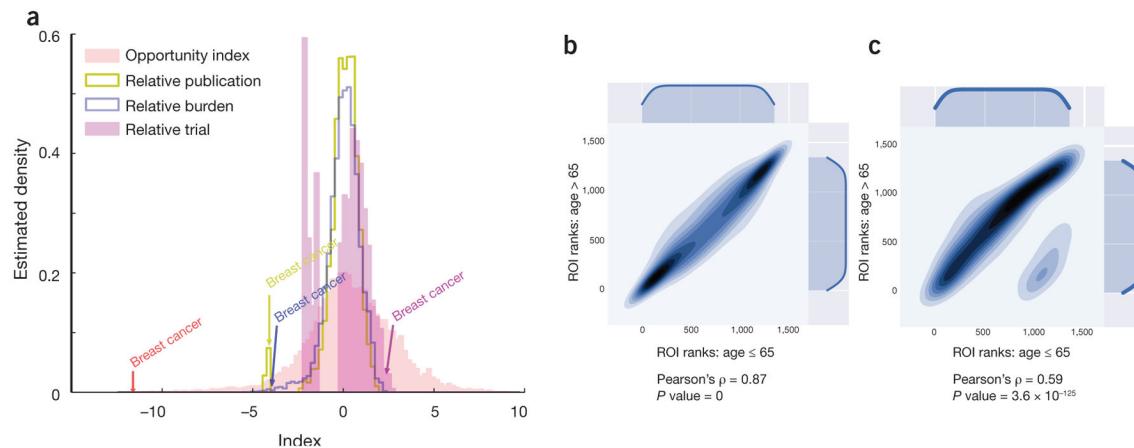
References

1. Merton RC. Bell J Econ. 1973; 4:141–183.

2. Black F, Scholes M. J Polit Econ. 1973; 81:637–654.

3. Gross CP, Anderson GF, Powe NR. N Engl J Med. 1999; 340:1881–1887. [PubMed: 10369852]

4. Gillum LA, et al. PLoS ONE. 2011; 6:e16837. [PubMed: 21383981]

5. Sampat BN, Buterbaugh K, Perl M. Milbank Q. 2013; 91:163–185. [PubMed: 23488714]

6. Vanderelst D, Speybroeck N. J Informetrics. 2013; 7:240–247.

7. Evans JA, Shim JM, Ioannidis JP. PLoS ONE. 2014; 9:e90147. [PubMed: 24691431]

8. Murray C. Bull World Health Organ. 1994; 72:429–445. [PubMed: 8062401]

9. Reiss J, Kitcher P. THEORIA. 2010; 24:263–282.

10. Kleinbaum, D., Kupper, L., Nizam, A., Rosenberg, E. Applied regression analysis and other multivariable methods (Cengage Learning). 2013.

11. Vanderelst D, Speybroeck N. PLoS Negl Trop Dis. 2010; 4:e576. [PubMed: 20126268]

Figure 1.

Health ROI for overstudied and understudied conditions (see Supplementary Table 5 for the top 50 over- and understudied conditions in 2011). (**a,b**) The area of each circle is inversely proportional to the opportunity index of the corresponding disease, indicating disproportionately overstudied conditions in 2011 (**a**) and in the period of 2000–2011 (**b**). (**c,d**) Circle area is directly proportional to opportunity index, indicating understudied, high-opportunity conditions in 2011 (**c**) and in the period of 2000–2011 (**d**). Sx, symptoms; VF, ventricular fibrillation; sed rate, sedimentation rate; DO, disorder; W/O, without; ESRD, end-stage renal disease; BCC, basal cell carcinoma; NOS, not otherwise specified/ unspecified; SIRS, systemic inflammatory response syndrome; DVT, deep vein thrombosis; IHD, ischemic heart disease; OSA, obstructive sleep apnea; WBC, white blood cell; NEC, not elsewhere classified.

Figure 2.
Disease-specific health ROIs. (**a**) Optum, (**b**) the University of Chicago Medical Center and (**c**) the National (Nationwide) Inpatient Samples data sets. (**a**) Distribution of ROI (red line) superimposed with distributions of relative disease burden (blue line), relative publications (yellow line), and relative clinical trials (magenta). The disease-specific ROIs follow an approximately normal distribution that is by definition zero-centered. For ease of comparison, relative disease burdens, relative publications and relative clinical trials are log-transformed and shifted to center at zero—by subtracting the index sample mean from each index value. In the case of breast cancer, the relative disease burden is approximately aligned with the relative publication focus on breast cancer, but not with the relative clinical trial value. As a result, the ROI for breast cancer is negative, which means the resources devoted to it exceed the disease burden. Computed for Optum. (**b,c**) Comparison of disease-specific ROI ranks for patients of 65 and >65 years old in the University of Chicago Hospital data (**b**) and the National (Nationwide) Inpatient Samples (NIS). (**c**) Both plots show statistically significant correlation between the two populations: the Pearson correlation coefficients equal 0.87 for University of Chicago Hospital data and 0.59 for the NIS. Health ROIs can be computed and meaningfully compared for the same population at different time points or for different populations.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Correlation between six measures of disease burden among 93 diseases in 2010

Correlation	Prevalence	Treatment cost	Death	DALY	YLD	YLL
Prevalence	1.000	**0.712**	−0.063	0.149	**0.444**	−0.079
Treatment cost	**0.819**	1.000	0.101	**0.383**	**0.588**	0.120
Death	**−0.328**	0.019	1.000	**0.861**	0.121	**0.973**
DALY	**0.311**	**0.495**	**0.472**	1.000	0.572	**0.877**
YLD	**0.668**	**0.516**	−0.151	**0.649**	1.000	0.108
YLL	**−0.334**	0.024	**0.989**	**0.479**	−0.154	1.000

Bold: significant correlation coefficients, $P < 0.01$.

Table 2

Health ROI 2000–2011

Year	ROI score
2000	2.254
2001	2.259
2002	1.995
2003	1.929
2004	1.853
2005	1.801
2006	1.728
2007	1.628
2008	1.608
2009	1.619
2010	1.649
2011	1.613

Table 3

Top over- and understudied conditions in 2011

Overstudied conditions	Understudied conditions
Breast cancer	Hashimoto's thyroiditis
Cervical cancer dysplasia	Other cervical disorders
Other symptoms of respiratory system	Palpitation
Renal failure	Secondary bone cancer
Cardiac arrest and ventricular fibrillation	Hyperlipidemia
Ischemic heart disease	Septal deviations/turbinate hypertrophy
Arrhythmia	Hypervolemia
Peptic ulcer	Adjustment reaction
Male infertility and abnormal sperm	Cervical radiculitis
Cholelithiasis and cholecystitis	Diarrhea

A more complete list of over- and understudied conditions is presented in Supplementary Table 5