

A Comparative Study of Charity Rating Metrics across Europe and America

Abstract

Donors are increasingly invited to give wisely, and a range of charity rating systems have emerged to help donors make better giving choices. Yet, reliability of charity ratings and prudence in their use should also be highlighted. Study of the charity raters' performance, especially outside the US, is largely missing in the nonprofit literature and this paper strives to make a contribution to filling this gap. Via qualitative research, eight case studies are explored, and the metrics they use to rate charities are compared. The mostly used metrics as well as the advantages and challenges of each are underlined. The findings indicate the metrics are diverse, yet not exhaustive. Some metrics for future consideration are suggested. The concluding point is that donors should not rely on charity ratings as an indicator of overall performance of charities in all aspects. They should use them vigilantly in light of the selected metrics.

Keywords: nonprofit/NGO/charity raters, charity watchdogs, evaluative metrics, rating and ranking

Introduction

External ratings and rankings of organizations, including the nonprofits, have constantly increased in recent years. Charity evaluation started in the US around the 1920's by the National Charities Information Bureau (NCIB) and the national office of the Better Business Bureau, which later merged to form the BBB Wise Giving Alliance. (Wise Giving Guide, spring 2016) Then, charity evaluators started to proliferate in the 1990's in the US—e.g. Charity watch, formerly known as the American Institute of Philanthropy (AIP), was founded in 1992—and the sector has been growing ever since, leading to the formation of several American third-party charity raters, mostly nonprofits themselves. Nevertheless, it seems that private and nonprofit efforts for vetting charities form a more recent trend in Europe, and still lacking in most parts of the world. UK's first independent charity evaluator, Charity Clarity, was founded in 2014. A Geneva-based independent media organization that ranks the world charities, NGO Advisor, was formed in 2009. So, the sector appears to be at its early times in Europe, and inception is awaited in other continents.

The goal of the charity evaluators is often reported as guiding the donors so they can easily choose and trust charities based on evidence-based information and make sure their contribution is used effectively and in line with their values. In addition to impacting donors' choice of charities, the charity watchdogs have also sought to promote best practices among nonprofits, which were for centuries presumably less accountable than the private sector or government entities. Despite similar goals, every charity rater has developed a unique rating scheme, evaluating different measures. So, it is interesting to study the different metrics and methods the charity watchdogs are actually using to rate nonprofits, so that we learn the degree of consistency among them. It is necessary for the donors as well as all the stakeholders to know how reliable the charity ratings are and with which cautions and considerations they should be viewed and used.

Literature

To investigate the prior research on charity rating systems or charity watchdogs, the key words charity/NGO/nonprofit rating/ranking/watchdogs were typed in googlescholar.com and google.com. (Despite the differences in meaning, the three terms charity, NGO and nonprofit will be used interchangeably here because they are used by the rating services.) The literature on charity rating systems has mainly focused on the impact of ratings on the donors' behavior. Yet, there is little consistency among the findings. On the one hand, there are studies that indicate changes in rating of a charity, provided by third parties, has no significant effect on the amount of donations the charity receives. (Szper and Prakash, 2011) On the other hand, others like Brown *et al.* (2017) and Gordon *et al.* (2019) suggest that charity evaluations impact donors' choice and the amount of contributions. A third category suggests a mixed conclusion based on the size of the charity. Yörük (2016) argues that in general, third-party ratings have an insignificant impact on the contributions, unless in the case of relatively small charities, when a better evaluation leads to higher donations. The impact of charity ratings on the nonprofits' behavior has also been studied. Szper, R. (2013) indicates that the charity watchdogs influence the way nonprofits report their financial information.

It seems that the functions of charity evaluation systems have not been fully investigated by the academia, especially outside the US, where it appears to be relatively new (Szper 2012). Few studies have already focused on comparing the watchdogs and the metrics they utilize for evaluating nonprofits on a large scale. Stork *et al.* (2008) describe three different rating schemes created by three American watchdogs and call for further research in the area; but it seems that the research was not much extended. Therefore, to contribute to filling this gap, this article compares major charity raters in different countries.

Method

The present study strives to identify numerous charity/NGO/nonprofit rating systems and offer a comparative perspective on them. For this purpose, the qualitative method of case study was adopted.

Eight rating services, that use a number of metrics to rate or rank charities, were chosen in the US, Canada, the UK and Switzerland: Charity Navigator, Charity Watch, BBB Wise Giving Alliance, Impact Matters and Give Well in the US; Charity Intelligence in Canada; Charity Clarity in the UK; and NGO Advisor based in Geneva. The study focuses on the metrics the selected charity raters report on their websites to use to evaluate charities, in order to understand if there is a consistent and comprehensive pattern among the rating schemes. By comparing the metrics, the widely used factors and less commonly used elements will be identified and discussed. Some of the missing measures among rating schemes are also suggested.

Data on the Case Studies:

NGO Advisor

NGO Advisor is a Geneva-based independent media organization that ranks the NGOs worldwide. Unlike most charity evaluators, NGO Advisor is not a non-profit itself, and does not provide the NGO rankings to the public for free. It is also unique in that it evaluates the NGOs across the globe, ranking about 12 million organizations worldwide —whereas, charity watchdogs are usually focusing on the nonprofits registered in one single country, though they could engage in international development and relief. NGO Advisor started its activity in 2009 and has published the Top NGOs World ranking annually since 2012, revising its metrics and methods of evaluation in 2013. NGO Advisor's research on world NGOs is organized into rankings, and leads to lists of top NGOs: 500 world, 500 USA, and 100 world NGOs.

For its rankings, NGO Advisor gathers information through the websites of the NGOs and also, if available, the questionnaire that is provided by NGO Advisor. It evaluates NGOs based on three macro measures of innovation, impact, and governance, via 165 criteria grouped into four main categories: economics and finance; marketing and communications; governance and human

resources; and an overview section. In addition, NGO Advisor rewards NGOs that manifest excellent levels of transparency and accountability or independence by giving bonus points.

Charity Clarity

Charity Clarity was launched in 2014 as a UK-based registered charity that describes its core service as charity assessment. It is said to be the first independent charity watchdog of the UK. It evaluates the UK-based charities registered with a Charity Commission in the UK (with some exceptions) that have been active for at least two years.

Charity Clarity does not rank charities; it rates charities by scoring them against 18 key metrics under the three major categories of financial health; accountability and transparency; and accessibility of charities. The numerical rating for every charity is illustrated by the number of stars, ranging from zero to five stars, which represent very poor to excellent performance.

Charity Intelligence Canada (Ci)

Charity Intelligence, a non-profit Canada-based organization formed in 2006, provides free reports on more than 750 Canadian charities. It states that its aim is to assist Canada's dynamic charitable sector in being more transparent, accountable and focused on results. Charity Intelligence uses publicly-available information, which includes financial statements, annual reports, CRA T3010 filings, and websites. It attributes zero to four stars to every charity evaluated based on five factors:

1. Results Reporting (the public reporting of the charity's activities, outputs, and outcomes: accountability)
2. Financial Transparency (accessibility of audited financial statements)
3. Need for Funding (the ratio of the funding reserves to the costs of programming)

4. Cents to the Cause (overhead spending is regarded reasonable between 5 to 35%)

5. Social Impact Rating (the social impact produced by the charity for each dollar donated plus the quality of the data available)

The first factor, results reporting, is weighted twice as the three others that follow for the star ratings. So, it accounts for 40 percent, and numbers two to four account each for 20 percent of the star rating. However, the new fifth factor is not still evaluated for all the charities and is, thus, not considered in the ratings in the same manner. High demonstrated impact will increase the star rating by 1 and low demonstrated impact can reduce it by 1 star.

Charity Navigator

Charity Navigator is the most referenced Charity evaluator in the media and the academic articles. The American nonprofit charity rater was founded in 2001, and claims to be the largest evaluator of charities in the US. It provides ratings for over 9,000 of America's charities to help donors make better informed decisions and help nonprofits improve their performance.

Charity Navigator only assesses the nonprofits that meet a number of rigorous criteria. This has led to the selection of 9000 nonprofits out of the 1.6 million nonprofits registered in the U.S. In order to be eligible for evaluation by Charity Navigator, the organization needs to be based in the U.S. and registered with the IRS, not too young or small, and funded by public support, at least in part.

In the process of evaluation, Charity Navigator compares charities with those that have similar activities and financial functionality. For this reason, it classifies charities in to categories and causes. For gathering information on charities, Charity Navigator uses self-reported information accessible via the website of the nonprofit as well as its financial statements via the IRS Form 990. (1) Financial Health and (2) Accountability & Transparency are the two major categories that Charity Navigator

considers in evaluating the charities. Charity Navigator believes the ratings shed light on a charity's cost-efficiency, sustainability, governance, best practices and openness with information. In addition, a charity's reporting of results will be a future metric in Charity Navigator's assessments in the future. Charity Navigator does not provide rankings; it designates ratings based on the score of a charity in the two areas of evaluation. Based on the overall score of a charity, an overall star rating, ranging from zero to 4 stars, is used to show the performance quality of a charity in comparison with others that are categorized in the same cause. In addition to the five levels of quality illustrated by the star rating, if a charity is found to raise serious concerns, it receives a CN Advisory mark.

Charity Watch

Charity Watch is another major charity evaluator of America that was formed in 1992, initially called the American Institute of Philanthropy (AIP). It claims to have the most stringent ratings in the sector due to careful analysis of the charities' finances and the use of various financial documents, including audited financial statements, tax forms, annual reports, state filings, and other documents. So, Charity Watch does not solely rely on self-reported information by the charities; it can also make adjustments to the reported figures based on the overall information it gains from diverse sources. It evaluates over 670 American charities and issues A+ to F letter grade ratings mainly to guide donors' decisions. It generally focuses "on evaluating large charities that receive \$1 million or more of public support annually, are of interest to donors nationally, and have been in existence for at least three years".

Charity Watch evaluates a charity's financial efficiency. The two main categories of investigation that Charity watch studies include (1) Program % (the percent of total expenses a charity spent on its programs in the year analyzed) and (2) Cost to Raise \$100 (the amount of money the charity spends to bring in \$100). Program % of 75% or greater and a Cost to Raise \$100 of \$25 or less make a charity financially efficient in the opinion of CharityWatch. "In CharityWatch's view, a Program Percentage

of 60% or greater and a Cost to Raise \$100 of \$35 or less are the minimum efficiency standards reasonable for most charities. Ratios in this range typically indicate a "satisfactory" or "C range" rating."

BBB Wise Giving Alliance (WGA)

The BBB Wise Giving Alliance is an American charity evaluator that was formed in 2001 by the merger of two charity monitoring organizations, the National Charities Information Bureau (NCIB) and the Philanthropic Advisory Service (PAS) of the Council of Better Business Bureaus. It assesses 1,300 nationally soliciting American charities that request to be evaluated or that the public has often asked about, and it publishes the results on its website, give.org. It states its aim as to assist donors in their choice of charities as well as to improve the nonprofits' conduct. Charity Accountability is the main focus of evaluations by WGA. In 2003, it issued twenty standards against which the charities' accountability is measured. The twenty standards are grouped into four categories of (1) governance and oversight—to check if the governing board is active, independent and free of self-dealing; (2) effectiveness—to check if the charity has a policy for setting objectives, evaluating and reporting the results; (3) finances—to check if a charity is financially transparent; and (4) solicitations and informational materials—to check if a charity's representations to the public are accurate, complete and respectful. So, the twenty standards cover a wide range of metrics including governance, results reporting, program expenses, fund raising expenses, transparency and accountability, among others. Accordingly, the metrics considered by BBB Wise Giving Alliance seem more comprehensive than the others in the sector.

While reporting on charities' performance based on the standards, the Wise Giving Alliance does not assign a score on a range—like a star rating or a letter rating. The charity either meets the criteria or not. Following evaluation, a charity can receive one of the following labels:

Meets Standards; Standards Not Met; Did Not Disclose; Review in Progress; Unable to Verify.

Then, for every charity evaluated, WGA provides a report in which every standard is ticked if it is met and crossed if not. If a charity meets all the standards, it is regarded as accredited by BBB WGA. In addition, accredited charities can use a BBB Accredited Charity Seal on their websites and in their fund-raising materials by paying a fee.

Impact Matters

Impact Matters is a new American charity rating nonprofit that was launched in 2015. (Howgego, 2019) It states its aim as helping the donors find high impact nonprofits. To this aim, Impact Matters has created a rating system that provides customized impact reviews and takes “explicit account of how much good the nonprofit achieves per dollar of cost.” The agency rates “service delivery nonprofits, i.e., nonprofits that deliver a program directly to people to achieve a specific health, anti-poverty, education or similar outcome”. Moreover, the nonprofit should rely on public support to provide at least some of its resources to be rated by Impact Matters. Over 1,000 American nonprofits have been rated by Impact Matters. It uses publicly available information on nonprofits, including the websites, GuideStar, tax forms, annual reports, audited financial statements as well as academic research, for the evaluations.

In terms of measures of evaluation, Impact Matters states cost-effectiveness or actual impact is the major metric based on which the nonprofits are evaluated. In the first stage, the nonprofits are categorized into groups based on the type of service they offer, so that the impact a nonprofit creates is measured in comparison to others that have similar missions. So, for every type of program, a specific methodology is developed to measure impact. The agency has eight categories based on program type so far, which include, for example food distribution and cataract surgery. Depending on the type of service, a methodology is then developed to evaluate impact. For instance, Impact

Matters measures the success of a food distribution program as its cost to provide a meal to a person in need. “The cost to prevent a person from going blind” is the metric for measuring impact of nonprofits focused on curing cataract. Therefore, the unique tool that it provides to donors is the ability to measure the net impact the donors’ contribution can have in every evaluated charity.

In addition to estimating impact for a program, it assigns 1 to 5 stars to the charities. A charity that receives one star, shows financial improprieties. “Excessive overhead, paid non-staff directors or no financial audit (for large nonprofits) or excess benefit transactions, material diversion of assets or a moderate or high Charity Navigator advisory” are the improprieties that will make a charity receive a one-star rating. Lack of sufficient public information to estimate the impact of a charity’s programs will result in a two-star rating. So, transparency and accountability are also part of the metrics used in Impact Matters’ evaluations. Three to four stars show the degree of self-effectiveness of a charity, which is measured by comparing the estimated impact of a given charity with the threshold determined based on alternative ways of providing the same service. So, “Cost-effective programs are making good use of resources, given the alternatives.” A three-star charity is not cost-effective: it uses more resources than the common cost of the service they provide. Four- and five-star charities are reportedly cost-effective and highly cost-effective.

GiveWell

GiveWell is another American nonprofit dedicated to introducing the top American charities. It was founded in 2007 with the mission of finding the most outstanding charities to help donors make sure their donations are used in the most effective manner. So, GiveWell introduces only the best charities and focuses on a limited number of charities that work in the developing world and offer direct aid, instead of providing ratings for a vast range of charities, low and high performing included. In 2019, Give Well provided a list of eight top charities that conduct “health and poverty alleviation programs,

serving people in the poorest parts of the world”. For choosing the top charities, GiveWell considers four measures of evaluation: evidence of effectiveness; cost-effectiveness; room for more funding; and transparency.

[Table Here]

Analysis

The overview of the metrics of evaluations used by the charity raters (as illustrated in the table) shows three metrics were widely used by charity evaluators. (1) A charity’s financial information is used by five of the selected agencies. Even if not stated as a major metric, financial health is considered by impact-focused raters like Impact Matters too; (2) transparency and accountability are considered by six agencies; and (3) impact and cost-effectiveness represent the focus of four charity evaluators studied. Here, we will study every one of these common metrics of evaluations to see if the charity evaluators have used them in the same way or if they mean different information in different ratings. The advantages and challenges of using every one of the three common metrics are then highlighted. Afterwards, some other metrics that are less commonly used by charity evaluators are discussed, which indicate the diversity of the metrics of evaluation. Finally, we will identify some missing metrics among the existing services and discuss the potential for alternative perspectives through which we can rate the performance of charities.

Finances

Finances or financial health is a common metric of evaluation among the charity evaluating systems we studied. NGO Advisor’s first set of criteria falls under the category of economics and finance. It does not disclose the specific criteria it uses, so we don’t know if by finance, it means financial

transparency or efficiency. But most probably financial efficiency is evaluated in this category, as transparency is stated to receive bonus points elsewhere.

The term 'financial health' is used by two evaluators, Charity Clarity UK and Charity Navigator. Under the category of financial health, Charity Navigator measures financial efficiency and capacity performance which include program, administrative, fundraising expense percentages; fundraising efficiency (money spent for generating \$1); program expenses growth; working capital ratio; and liabilities to assets ratio. Charity Clarity considers 5 criteria under the category of financial health which include: late in submitting accounts; working capital; expense growth; total net expenditure / total income; employees over 3-year average income.

Charity Intelligence Canada uses 'Cents to the Cause' factor to "evaluate how many cents are available to go towards the charity's programs for each dollar donated" as opposed to the overhead spending.

Financial Efficiency is the term used by Charity Watch, and it is the sole metric it considers for evaluating charities. To measure financial efficiency, it considers two factors: (1) the percent of total expenses a charity has spent on its programs as opposed to overhead (fundraising, and management & general) and (2) the cost to raise \$100.

BBB Wise Giving Alliance also checks if the governing body is free of self-dealing, under the category of governance and oversight. It also defines standards for the fundraising and program expenses, so it considers administrative expenses or overhead.

Therefore, five of the selected charity evaluators evaluate charities based on overhead (administrative and fundraising expenses percentage) and fundraising efficiency. At least two charity evaluators consider financial capacity, including working capital and program expense growth, as well. Yet, as can be noticed, every charity collects different information under the category of finance. So, the fact

that financial information is used in evaluations does not mean they are always used in the same way, with the same metrics and criteria in mind.

Overhead ratio is one common criterion under the category of financial health, which has been the most controversial factor for rating charities. Many, like Charity Watch, argue that overhead ratio is an essential factor for deciding about donating to a charity because it is important to pay attention to how charities spend the donations. Based on the fact that the amount of annual giving has remained steady for many years in the US, Charity Watch contends that charitable money is limited and ignoring fundraising ratios can lead to increased overhead costs in the sector and therefore billions of dollars shifted out of charitable programs and spent on fundraising instead. (charitywatch.org) In addition, Charity Watch mentions overhead costs ratio is easy to calculate, in comparison with some other metrics such as impact. Therefore, it is reportedly a realistic and unbiased metric.

Research has indicated that donors care about overhead and that high overhead costs turn off donors. (Okten et al. 2000; Meer, J. 2014; Bowman, W. 2006) BBB Wise Giving Alliance reports that a charity's finances represent the most influential factor on donors' decisions, compared to results and ethics. (BBB Wise Giving Alliance: Shaping the Future of Charities. Annual Report 2014. Online at: <http://www.give.org/about-bbb-wga/annual-reports/>)

Nevertheless, there is an opposing view that holds overhead costs ratio is not a good measure of charity evaluation. Meer, J. (2017) indicates that overhead ratio is not indicative of a charity's effectiveness: "A charity with a low overhead cost ratio that fails in its stated mission should not be judged more highly than one with a higher ratio that succeeds." (Meer, J. 2017 p.5) Highly effective charities, the writer argues, like the top charities of GiveWell, can manifest higher overhead costs ratio. Meer concludes that under the pressure of keeping the overhead costs low to compete for donations, charities have to sacrifice important functions like attracting and retaining skilled staff.

The expectation that nonprofits should keep overhead costs very low is referred to as ‘Nonprofit Starvation Cycle’ (Gregory, A. G., & Howard, D. 2009), described as “a vicious cycle [that] is leaving nonprofits so hungry for decent infrastructure that they can barely function as organizations—let alone serve their beneficiaries.” (Gregory, A. G., & Howard, D. 2009, p.49) This movement of countering reliance on nonprofits’ low overhead costs for rating them has taken momentum recently, and several articles were dedicated to it. Therefore, some major charity rating agencies had to take action in support of the movement. In 2013 and 2014 GuideStar, Charity Navigator and BBB Wise Giving Alliance have started a campaign called "The Overhead Myth". In their two open letters to the donors and the nonprofits, they assert that the disadvantages of relying solely on overhead costs ratios outweigh its advantages. They justify their use of overhead ratios as one of their evaluative metrics by saying that it is only used “for rooting out fraud and poor financial management”.

Gneezy et al. (2014) have suggested a solution for attracting donations despite funders and donors’ reluctance to pay for overhead costs. They find that if donors are told that the overhead costs are covered by a major funder and that their money is used directly for the program expenses, donations will increase substantially. As the target of charity raters are mainly private donors, the overhead costs ratios factor can be adjusted in a way to more accurately inform the donors. The charity raters might want to use a new metric: the percentage of new donations that is spent on overhead, rather than the overall overhead costs ratio.

Transparency and Accountability

Transparency and Accountability are other common metrics of evaluation among charity evaluators. Charity Navigator offers definitions for accountability and transparency in assessing charities: “Accountability is an obligation or willingness by a charity to explain its actions to its stakeholders.

Transparency is an obligation or willingness by a charity to publish and make available critical data about the organization.” (www.charitynavigator.org)

Charity Clarity UK measures transparency and accountability as its second major category of metrics. Under this category, it considers seven criteria which follow: Clarity over Trusteeship Process; Number of Trustees; Female Representation; Number of Additional Boards the Trustees Serve On; Social Impact Reporting; Public Relations; Revenue from Trading Activities.

Charity Intelligence Canada’s five factors of measurement consider transparency and accountability in the two first factors: (1) Results Reporting and (2) Financial Transparency.

“The results reporting grade is an evaluation of the charity's reporting levels. This evaluation takes into account the public reporting of the charity's activities, outputs, and outcomes, without assessing the strength or quality of these elements.” A charity's financial transparency is evaluated based on how easily its audited financial statements can be accessed.

Charity Navigator’s second and last set of criteria is grouped under the category of accountability and transparency too. It uses seventeen different metrics to assess “whether the charity follows best practices of governance and ethics, and whether the charity makes it easy for donors to find critical information about the organization”.

BBB Wise Giving Alliance also regards accountability as its main focus. It uses 20 standards in four categories, which include governance, effectiveness, finances, and the fourth category is about transparency. The governance-related metrics define standards for the board oversight, size, meetings and compensation as well as conflict of interest. Then, the requirements for transparency concern accurate representation in all the informational materials, an annual report that is available to all, as well as accessibility of information on the website of the charity. Transparency about donor privacy policy and cause-marketing are also among the standards of the WGA.

Givewell is another American charity evaluator that investigates the potential top charities while considering the level of transparency they exercise, as its fourth metric of evaluation. Nevertheless, it does not mention any specific metrics under the category of transparency. It states that to be given the top charity label, charities must be open to their investigation and agree to the disclosure of their track record, both good and bad.

NGO Advisor's three pillars of interest do not include transparency and accountability, yet it gives bonus points for manifestation of transparency and accountability in the NGOs.

So, at least six of the charity evaluators studied here consider transparency and/or accountability as a factor for evaluating charities. It seems plausible that charities be encouraged to disclose information about their finances and results to prevent misuse of the donations and guarantee achievements of the goals. Yet, it is interesting to note that instead of having an impact on how the charities actually behave, the pressure for being transparent might only influence how the charities report their financial activities or outcomes. Szper (2013) studied how third-party ratings impact the behavior of nonprofits and the study found that "such ratings are, in fact, having an effect on how nonprofits report financial information on the IRS Form 990".

Impact and Cost-effectiveness

Impact and Cost-effectiveness represent the third common factor of evaluation among charity evaluators, and it seems a more recent trend, used by younger charity evaluating agencies like Impact Matters. 'Impact' is a factor that NGO Advisor considers while ranking NGOs. By impact, it refers to an NGO's output, or how it transforms the lives of its beneficiaries. In addition to evidence that an NGO's work has added value to the community it serves, results reporting is also investigated under the category of impact. Charity Intelligence Canada uses the term 'Social Impact Rating' to indicate "the social impact produced by the charity for each dollar donated as well as a measure of the quality

of the data available to assess the charity's social impact.” It is a newly added factor by Charity Intelligence Canada, and it believes it is the most essential factor for determining if a charity is a good one. Impact Matters rates charities on the basis of their impact and cost-effectiveness only. It provides an impact estimate for the amount of money a donor intends to give. In addition, the estimated impact of a charity is compared to benchmarks to determine the degree of its cost-effectiveness. Give Well also rates potential top charities based on cost-effectiveness: the number of lives saved or improved per donation.

Therefore, four of the selected charity raters, rate or rank charities explicitly against the factor of impact and cost-effectiveness. Somewhat in response to the criticism on charity rating based on overhead and financial documents, a movement has emerged that focuses on impact and cost-effective donations. From this emerging perspective, a charity is to be rated higher if it uses resources effectively and creates more impact with the same amount of donation. It appears a logical factor of evaluation because nonprofits are there to save or improve lives, and their higher impact should mean their better performance. This must be enticing for donors to know their donation can do good more in a given charity than in others. Nonetheless, this factor has also its shortcomings and has been criticized.

The three charity evaluators that use cost-effectiveness to rate charities admit that they can use this metric to rate certain types of missions. The impact of not all missions or programs can be evaluated easily, if at all. Charity Intelligence has assessed charities primarily in the social services and education sectors as well as a number in the international aid sector. Give Well seeks out charities that manifest evidence of effectiveness, leading to a limited number of choices like the category of health intervention and cash transfer. Impact Matters’ eight categories of programs already rated include missions that produce tangible results, like direct aid and relief, and health intervention: Food Distribution; Emergency Shelter; Postsecondary Scholarships; Cataract Surgery; Water Purification;

Tree Planting; Financial Assistance for Patients with Medical Conditions; Veterans Disability Benefits.

Cost-effectiveness models only consider—and maybe unintentionally value better—low-cost interventions with immediate results. As a result, other missions, like research, that have less quantifiable results are placed in an unfavorable position when cost-effectiveness is to be measured. Thus, if impact/cost-effectiveness is to be used for rating charities, a substantial number of charities might fall out of the list of rated charities and thus receive less attention. For instance, human rights and equity might be excluded from charity-ranking assessments focused on impact, as a cost-effective charity provides services to those easiest to reach. (Cochrane & Thornton 2015, P.58)

In addition, measuring impact and cost-effectiveness is not easy and evident, compared with other information such as financial figures. For measuring impact, qualitative information should usually be transformed into quantitative information. So, it often needs a complicated and costly process to measure impact. Gugerty & Karlan (2018) argue measuring impact is only possible at the right time and with the right tool. They contend that at times it is not feasible or worth it at all. Furthermore, CharityWatch believes impact-based ratings are probably biased, because they rely on results as reported by the charities themselves, rather than objective facts. (charitywatch.org)

Other Existing Factors of Charity Evaluation

In addition to the three common metrics which were already discussed, some of the selected charity evaluators use interesting unique metrics. Give Well uses the factor of ‘room for more funding’ to assess if the charity has the potential to put extra funding into good use, effectively and quickly. It asks, "What will additional funds — beyond what a charity would raise without the recommendation — enable, and what is the value of these activities?" As mentioned, Give Well is an impact-focused charity evaluator. This measure is used toward their goal of selecting highly effective charities. This

metric appears a very functional one, because it investigates or predicts how the extra money will be used in the future, rather than how the charity used its entire resources in the past. Such a metric can solve the controversy over administrative and fundraising costs ratio, because normally the overhead costs are covered by funds raised without Give Well's recommendation. So, the donor will know how their donation will be used, instead of how the overall resources—possibly provided by other funders—are used. If the new money goes to expanding programs or increasing the number of the beneficiaries, then the overhead might not matter as it is already covered by other funds.

This interesting factor also poses challenges. Measuring impact was already mentioned to be a complex and expensive process. Now, predicting future impact must be still more complicated, especially because no one can tell how the future might change and affect everything. Therefore, even if in theory 'need for more funding' is a very attractive measure of evaluation, its reliability is questioned. Another criticism that this metric has received is that every charity has a limited room for more funding. If the donors donate all at once to one single or a few charities, to the point that the resulting donations exceed the maximum amount of estimated needed funding, the money can no longer be used effectively. (Peters, 2019)

Innovation is also an interesting metric that NGO Advisor is using to rank the world NGOs. By considering innovative or creative approaches the NGOs adopt, NGO Advisor gives more chances to newly launched NGOs to enter the top lists. Whereas, for most other metrics experienced charities have better chances. Reaching good levels of impact and cost-effectiveness, satisfactory financial results or competent transparency might require years of experience.

Governance also appears among the factors of evaluation, but it is a general term that is used to measure varied criteria. NGO Advisor and Give.org both refer to governance among their factors of evaluation. NGO Advisor uses governance to refer to the general approach an NGO adopts in dealing with employees, directors, and stakeholders. Give.org focuses on the governing board to ensure that

the volunteer board is active, independent and free of self-dealing. Independence is also valued by some charity watchdogs like NGO Advisor and Give.org, and affects the ratings. Finally, independence is also rewarded by some of the charity evaluators. “Dependence on corporations, governments, single funders, or other specified sources” is penalized by NGO Advisor; and independent board is among the standards required by Charity Navigator and BBB Wise Giving Alliance.

Discussion

As the data from eight leading charity watchdogs in Europe, Canada and the US revealed, the metrics used to evaluate charities are diverse. Three similar themes emerged as somehow widespread among charity evaluation metrics (financial health, transparency and impact). However, similar themes do not always reflect the same data or importance (weight in scoring); even when the metrics appear to be the same, the information gathered or the process through which the information is analyzed might differ. For instance, ‘governance’ is a term used among factors of evaluation by some of the selected charity evaluators, yet in one case it refers to how the charity applies its mission to all its staff (NGO Advisor), in another it refers to how well the volunteer board is performing (BBB Wise Giving Alliance). Therefore, charity evaluation metrics are diverse and inhomogeneous.

Every charity evaluator has developed its own system of evaluation based on some selected metrics. The metrics used are the choices made by every charity evaluator, and the resulting rating is determined by the choices of metrics and can differ when other metrics are considered. By comparing the ratings attributed to a few charities by different charity evaluators, research has indicated the ratings are not always consistent (Stork & Woodilla, 2008), because each one uses a certain set of metrics. So, while relying on charity watchdogs, the donors should be very vigilant and view the ratings in light of the choices the watchdog has made. Donors should know the ratings are the result

of the analysis based on selected metrics and methods, rather than objective and unanimous means that determine the position of a charity in all aspects.

The study of the charity evaluation metrics also revealed that the spectrum of the existing metrics is still limited, despite diversity. We can think of other interesting metrics that haven't been considered. The possible metrics that future studies and charity watchdogs can use to rate charities include for example the long-term effect of the programs of a charity. As seen, focus on impact has been limited to concrete and short-term results. So, new methods can be developed to offer broader impact-based evaluations, and study other implications of provided services. The extent to which programs of a charity favor human rights and equity can also be part of the impact-based evaluation, to see if the services reach men, women and children, or different social groups who need the service equally.

Another important feature that future investigations can consider is the extent to which charities favor democracy. Are the services and interventions of a charity desired and welcomed by the majority of the target communities, or are they representing the decisions and wants of the charity leaders or a small minority of the affected people only. Programs of a charity might not be aligned with democracy if they thrust western thinking and lifestyle to societies that wouldn't welcome them, affecting the values and norms of a society.

Future research on evaluative measures can also focus on the importance of relevance of the type of services provided and the evaluative metrics. Some charity evaluators in our study, such as Impact Matters and Charity Navigator, group charities into categories based on the causes they support and type of interventions, and rate charities in comparison with others with similar missions. This is important in the study of rating systems, as charities should have some common grounds so their comparison makes meaningful results: comparing a research-focused charity and an aid-providing charity is not appropriate, as they are inherently different. The same could also be true for charities that work in different social, political, cultural or even religious contexts. Comparing their

performance without noticing the environmental differences could be unfair or even misleading. Therefore, to solve this problem, the evaluative measures can be divided into two groups of general metrics that apply to nearly all charities and specific metrics, developed for every category of activity, size of organization, environment, etc.

Another point that can be noticed among charity evaluators, especially impact-focused ones, is that they mainly focus on rating charities that provide direct aid. This can be explained partly by the fact that measuring impact of such missions is easier. But it might not be the only reason. Undoubtedly, the preference of the donors is an important factor in their giving decisions, and giving to causes like fighting diseases, educating children, and other direct aid interventions is more attractive than giving to causes like policy analysis and research with speculative results. Thus, the charity evaluators might choose to focus on the causes that the donors are more inclined to, which can be unfair to other causes. In order to better represent all the organizations and causes in the third sector, new metrics should be developed to evaluate different engagements, including research-based charities.

Debate on charity ratings is still controversial. Some have even criticized the use of charity performance metrics because they might negatively affect giving, since they provide excuses not to give (Exley, 2020) or because focus on a selection of metrics can miss other important information and create a set of unintended outcomes (Cochrane & Thornton, 2015). More research is needed in this arena to feed the arguments and lead to better judgments about charity evaluation. As was noticed in the study, other countries have recently started to adopt the American model of charity ratings. Before more countries embrace the concept of rating charities, they should carefully consider the challenges and the debate rating systems have created.

Conclusion

In conclusion, analysis of evaluative metrics used by eight case studies indicates diversity in the metrics and methods and shows there is no comprehensive pattern of evaluation. Instead, every charity evaluator uses a selected number of metrics and rates the charities only based on the limited number of metrics, at times one or two factors only, which portray charities partially instead of giving a complete picture of them. Yet, far from deciding that the charity evaluators are misleading, we suggest the donors should use the ratings in this light. Donors should not rely on ratings as an indicator of overall performance of charities in all aspects. They should use them prudently while considering the metrics and methods they use.

References:

- Bowman, W. (2006) "Should donors care about overhead costs? Do they care?" *Nonprofit and Voluntary Sector Quarterly* 35:2: 288–310.
- Brown, A. L., Meer, J., & Williams, J. F. (2017). Social distance and quality ratings in charity choice. *Journal of Behavioral and Experimental Economics*, 66, 9-15.
- Cochrane, L., & Thornton, A. (2015). Charity Rankings: Delivering Development or Dehumanising Aid? *Journal of International Development*, 28(1), 57–73.
- Exley, C. L. (2020). Using charity performance metrics as an excuse not to give. *Management Science*, 66(2), 553-563.
- Gordon, T. P., Knock, C. L., & Neely, D. G. (2009). The role of rating agencies in the market for charitable contributions: An empirical test. *Journal of accounting and public policy*, 28(6), 469-484.
- Gregory, A. G., & Howard, D. (2009). The nonprofit starvation cycle. *Stanford Social Innovation Review*, 7(4), 49-53.
- Gugerty, M. K., & Karlan, D. (2018). Ten reasons not to measure impact—And what to do instead. *Stanford Social Innovation Review*. 41-47
- Howgego, J. (2019). How to do good. *New Scientist*, 244(3259), 42–46
- Meer, J. (2017). Are overhead costs a good guide for charitable giving?. *iza World of Labor*.
- Meer, J. (2014) "Effects of the price of charitable giving: Evidence from an online crowdfunding platform." *Journal of Economic Behavior & Organization* 103: 113–124.

Okten, C., and B. Weisbrod. (2000) “Determinants of donations in private nonprofit markets.” *Journal of Public Economics* 75:2 255–272.

Overhead Ratios are Essential for Informed Giving (2014): <https://www.charitywatch.org/charity-donating-articles/overhead-ratios-are-essential-for-informed-giving>

overheadmyth.com

Peters, D. (2019). *Economic Design for Effective Altruism. The Future of Economic Design* (pp. 381-388). Springer, Cham.

Stork, D., & Woodilla, J. (2008). Nonprofit organizations: An introduction to charity rating sources and cautions in their use. *International Journal of Applied Management and Technology*, 6(4), 1.

Szper, R., & Prakash, A. (2011). Charity watchdogs and the limits of information-based regulation. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 22(1), 112-141.

Szper, R. (2013). Playing to the test: Organizational responses to third party ratings. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 24(4), 935-952.

Wise Giving Guide (spring 2016) Celebrating Our 15th Anniversary as the BBB Wise Giving Alliance, p.3. <http://www.give.org/globalassets/wga/wise-giving-guides/spring-2016-guide-article.pdf>; Accessed April 6, 2020

Yörük, B. K. (2016). Charity ratings. *Journal of Economics & Management Strategy*, 25(1), 195-219.

Charity	Evaluated Charities	Evaluation scheme	Evaluation Metrics
Evaluator			

NGO Advisor	About 12 million organizations worldwide	ranking	1. Innovation 2. Impact 3. Governance (transparency and accountability and dependence will receive bonus points.)
Charity Clarity	UK registered charities	zero to five star rating	1. Financial Health 2. Accountability and Transparency 3. Accessibility
Charity Intelligence Canada	More than 750 Canadian charities	Zero to four-star rating	1. Results Reporting 2. Financial Transparency 3. Need for Funding 4. Cents to the Cause 5. Social Impact Rating
Charity Navigator	More than 8000 American Charities	Zero to four star rating	1. Financial Health 2. Accountability & Transparency
Charity Watch	Over 670 American charities	A+ to F letter grade ratings	Financial Efficiency

BBB Wise Giving Alliance	1,300 American charities	Either Accredited or not	Charity Accountability: 20 standards in four categories: (1) governance and oversight; (2) effectiveness policy and report; (3) finances; and (4) solicitations and informational materials
Impact Matters	Over 1,000 American “service delivery” nonprofits	Estimated impact per donated dollar; 1 to 5 star rating	Impact (Cost-Effectiveness)
Give Well	potential top charities that work in the developing world and focus on direct aid	Provides two lists of eight “Top Charities” and “Outstanding Charities”	1. Effectiveness 2. cost-effectiveness 3. room for more funding 4. transparency