A flexible formula for incorporating distributive concerns into cost-effectiveness analyses: priority weights

Ø.A. Haaland\textsuperscript{a} PhD, F. Lindemark\textsuperscript{a} MD, K.A. Johansson\textsuperscript{a,b,c} MD, PhD

\textsuperscript{a}Bergen center for ethics and priority setting (BCEPS), Department of global public health and primary care, University of Bergen, Norway

\textsuperscript{b}Department of addiction medicine, Haukeland university hospital, Bergen, Norway

\textsuperscript{c}Department of international collaboration, Haukeland university hospital, Bergen, Norway

Corresponding author:

Øystein Ariansen Haaland, PhD, oystein.haaland@igs.uib.no, (+47) 55586124, PO Box 7804, NO-5018 Bergen

Other authors:

Frode Lindemark: frode.lindemark@igs.uib.no

Kjell Arne Johansson: kjell.johansson@igs.uib.no

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
ABSTRACT

Cost effectiveness analyses (CEAs) are widely used to evaluate the opportunity cost of health care investments. However, few functions that take equity concerns into account are available for such CEA methods, and these concerns are therefore at risk of being disregarded. Among the functions that have been developed, most focus on the distribution of health gains, as opposed to the distribution of lifetime health. This is despite the fact that there are good reasons to give higher priority to individuals and groups with a low health adjusted life expectancy. Also, an even distribution of health gains may imply an uneven distribution of lifetime health.

We develop a systematic and explicit approach that allows for the inclusion of lifetime health concerns in CEAs, by creating a new priority weight function, \( PW = \alpha + (t - \gamma) \cdot C \cdot e^{-\beta \cdot (t - \gamma)} \), where \( t \) is the health measure. \( PW \) has several desirable properties. First, it is continuous and smooth, ensuring that people with similar health characteristics are treated alike. Second, it is flexible regarding shape and outcome measure, so that a broad range of values may be modelled. Third, the coefficients have distinct roles. This allows for the easy manipulation of the PW’s shape. In order to demonstrate how PW may be applied, we use data from a previous study and estimated the coefficients of PW based on two approaches.

The first considers the mean weights assigned by the respondents:

\[ PW_{\text{mean}} = -0.42 + (t + 22.2) \cdot 0.27 \cdot e^{-0.031 \cdot (t + 22.2)} . \]

The second considers the median weights:

\[ PW_{\text{median}} = 0.79 + (t + 8.85) \cdot 0.17 \cdot e^{-0.053 \cdot (t + 8.85)} . \]

This illustrates that our framework allows for the estimation of PWs based on empirical data.
INTRODUCTION

Prioritisation of limited health care resources involves saying no and yes, hard ethical problems and reasonable disagreements. In health economics, cost-effectiveness analysis (CEA) is used to identify the most efficient allocation of health care resources. Within such a framework, informed decisions based on explicit CEA rankings can be made. However, CEAs have been extensively criticised for not being sensitive to fair distribution of health benefits, and there are other important equity concerns which may be considered, and few standardised methods are available to directly adjust incremental cost-effectiveness ratios (ICERs) [1-3]. CEAs are used to inform policy decisions on the introduction or reimbursement of new technologies or pharmaceuticals in countries such as the UK, Norway, Sweden, the Netherlands, New Zealand and Australia, in some cases including considerations beyond CEA through concepts like severity or burden of disease [4-14].

Typical health metrics in CEAs are quality-adjusted life years (QALYs) gained and disability-adjusted life years (DALYs) averted [15]. These aggregate measures take into consideration not only the life expectancy (LE) of a group, but also the quality of life. Interventions and health programs are typically ranked by ICERs, where those that maximize QALYs gained or DALYs averted are favoured. However, such rank-orders of interventions based on ICERs have been criticized because they disregard concerns for those who are worst off [16]. There are good reasons to give higher priority to individuals and groups with impaired health [16]. As opposed to health maximisation, one could aim for equal distribution of health across all individuals in society. This would require that resources should always be directed at people with the most impaired health, so as to minimize inequality in health outcomes, even if this means that the total health in the population will be reduced. People’s intuition about what is fair may lie somewhere between strict health maximisation and strict equality. A compromise between these perspectives is that health maximization is an important goal, but health gains to people with impaired health are assigned more weight than similar gains to people with better initial health [17]. This “in between” position is called prioritarianism and has been defended by the philosophers like Derek Parfit [18] and Matthew Adler [19]. According to prioritarianism, a cheap and effective intervention directed at the healthiest may be preferred to an expensive and ineffective intervention directed at those who have worse health if left untreated. Still, if cost and effect were equal, the allocation of resources should be directed towards people with more impaired health if left untreated. Finally, when comparing competing health programs, studies have found that people favor priority to worse off, and they favor interventions that benefit those that are worse off over slightly less cost-effective interventions that benefit better off groups [20, 21]. Such equity concerns are too important to leave out from standard CEAs.

One may also argue that non-health concerns, like age, indirect benefits, productivity, and financial risk protection, are also important regarding equity. However, this paper focuses on health related concerns. At the time of a prospective health intervention, concerns for health distribution can be roughly divided into two perspectives: The first perspective is forward looking and is only concerned with the distribution of the health gains, typically measured in QALYs or DALYs. E.g., one may think that one QALY gained for 50 people is better, equal
or worse than 50 QALYs gained for one individual. The second perspective is dealing with
the distribution of health over a lifetime, typically measured as quality adjusted life
expectancy from birth (QALE) [22, 23]. E.g., consider two individuals who may benefit from
some treatment. One has QALE=80 without treatment and 81 with, and the other has
QALE=20 without treatment and 21 with. Their gains would be equal, but the second is
expected to get much less lifetime health. Hence, if an equal distribution of lifetime health in
the population is a priority, one would value the set \{80, 21\} above \{81, 20\}. The view that
expected lifetime health is what matters draws support from the literature on health priorities
and some empirical studies [24-28]. Note that although QALE often increases with age, a
high age does not necessarily correspond to a high QALE. E.g., the QALE of an old person
who has experienced a lifetime of chronic illness may be far less than that of a healthy child.
In this paper we focus on applications of the lifetime health perspective.

Our aim is the development of a systematic and explicit approach to empirically include these
equity concerns in the framework of CEAs, by creating a new priority weight function. The
function should be smooth and continuous in order to treat people with similar characteristics
alike. Further, it should be flexible regarding shape and outcome measure to be able to
encompass a broad range of values. Finally, coefficients should have distinct roles, to ensure
that different functions can be easily compared, and allow for easy manipulation of the
function’s shape.

METHODS

Moving from utility curves to priority weights

A social welfare function (SWF) is a function that represents the sacrifices that society is
willing to make to promote a more equitable health distribution. The input should be a health
distribution (individual QALYs, DALYs, etc.), and the output should be a real number, so
that different distributions can be easily ranked. An SWF incorporates trade-offs between
total health gain and health inequality in a population. Well-known SWFs are the Gini index
and the Atkinson index [29, 30]. Both include a parameter that indicates the degree of
inequality aversion, ranging from 0 to infinite. Such SWFs are data hungry, because complete
data about the health or health gains for an entire population are needed. Often the data
required does not even exist. A different set of SWFs is the one based on the aggregated
QALY model. These utilitarian SWFs have the form

\[ W(t_1, ..., t_n) = \sum_{i=1}^{n} u(t_i) \]  

where \( t_i \) is the number of QALYs received by individual \( i (i=1, ..., n) \), and the utility function
\( u(t_i) \), defined over \( t_i \), is positive and monotonously increasing (\( u(t) \geq 0, du/dt > 0 \)). If \( u(t) \) is
concave (\( d^2 u/dt^2 < 0 \)), \( W \) represents societal preferences for equality. Equality indifference
yields \( u(t) = t \), and reduces \( W \) to be the sum of the QALYs received across all individuals in
society. Finally, if \( u(t) \) is convex (\( d^2 u/dt^2 > 0 \)), society is assumed to prefer inequality. The
function \( u(t) \) has been the focus of several papers [29, 31-34]. As opposed to the Gini and
Atkinson indexes, calculating \( u(t) \) does not require complete information about the health or
health gains of an entire population \((t_1, \ldots, t_n)\). We refer to Rodriguez and Pinto for a more
detailed discussion on \(u(t)\) [34].

However, \(u(t)\) does not account for the distribution of lifetime health. We approach this
problem by considering priority weight functions (PWs) instead. These are functions that
adjust the weight of health gains according to health related equity concerns, such as health
attainment in a lifetime perspective. A general form of an SFW based on PWs is

\[
W(s_1, \ldots, s_n, t_1, \ldots, t_n) = \sum_{i=1}^{n} F(s_i, t_i),
\]

where \(t_i\) is the health level of individual \(i\) at the time of intervention, \(s_i\) is the health gain, and

\[
F(s_i, t_i) = \left| \int_{t_i}^{t_i+s_i} \text{PW}(x) \, dx \right|.
\]

The vertical bars denote that the absolute value is to be considered, and \(x\) is a dummy variable
used for integration. As an example of how (6) may be calculated, consider a sick individual
with a health level without intervention of QALE=60, and an intervention with a gain of 3
QALYs. The priority weighted gain of the intervention is now \(F(3,60) = \int_{60}^{63} \text{PW}(x) \, dx\).

Although the literature is lacking, one candidate for PW is the age-weight function of the
Global burden of disease (GBD),

\[
\text{PW}_{\text{GBD}}(t_i) = C_{t_i} e^{-\beta t_i}.
\]

Here \(t_i\) is the age of individual \(i\) [35]. As mentioned, age weights are beyond the scope of this
paper. Still, by letting \(t_i\) denote the health gain or lifetime health of individual \(i\), (5) becomes
a PW with quite an appealing form, although lacking in flexibility.

Flexibility is an important property of a PW, as it should be able to reflect a wide range of
health related equity concerns. Further, if the coefficients of the PW have distinct roles, they
can easily be adjusted to accommodate such concerns. E.g., one may wish to set an upper
limit for the maximum weight, or set the maximum weight at a fixed value of \(t\). Also, if the
PW is fit to empirical data on people’s preferences, the roles of the coefficients will help to
disentangle different concerns of society. One may, e.g., wish to estimate both the \(t\) for which
the PW reaches its maximum, and what the maximum is. This is not straightforward in (5). A
PW should also be twice differentiable with respect to \(t\), so that it is continuous (no jumps)
and smooth (no breaking points), ensuring the approximately equal treatment of individuals
with similar values of \(t\).

A priority weight function

In this section we present a new flexible PW with coefficients that have distinct roles, which
is suitable in a lifetime health framework via (6). The formula is as follows:

\[
\text{PW}_1 = \alpha + (t - \gamma) \cdot C \cdot e^{-\beta \cdot (t - \gamma)}. 
\]
Setting $\alpha = \gamma = 0$ reduces (7) to (5). Setting $C = 0.1658$ and $\beta = 0.04$ we get the weights shown in Figure 1A. Panels B and C in Figure 1 show how $PW_1$ can be altered by manipulating $C$ and $\beta$. As seen, $C$ has the effect of inflating or deflating $PW_1$ (Figure 1B). For $\beta > 0$, it can be shown that (7) ultimately approaches $\alpha$ with increasing $t$. We see that for larger $\beta$'s this happens more quickly than for smaller $\beta$'s (Figure 1C). When $\beta \leq 0$, (7) increases with $t$, and never approaches $\alpha$ at all.

[Figure 1 here]

The intercept $\alpha$ of (7) shifts the curve up or down (Figure 2A), whereas $\gamma$ provides a right-left shift (Figure 2B). As is shown in Figure 2B, the maximum can be moved beyond $t = 0$, making (7) strictly decreasing. Note that $C$ is interpreted slightly differently for (7) than for (5). For (7), changing $C$ will affect the relative weights between different values of $t$ unless $\alpha = 0$. Also, $C \leq 0$ is not a problem for (7), because $\alpha$ can be adjusted to make sure that $PW_1 > 0$ for all relevant $t$'s. A negative $C$ simply causes (7) to be reflected through the line $PW_1 = \alpha$, or in other words, turned up-side-down. If $C = 0$, (7) reduces to the line $PW_1 = \alpha$, corresponding to inequality indifference. This is also true if $\beta$ approaches infinity.

[Figure 2 here]

As demonstrated in Figure 1 and Figure 2, manipulating the different coefficients of (7) allows us to flexibly change its shape. E.g., if we want to ensure that people with a certain value of $t$ are given a certain weight (anchoring the curve), we may simply change $\alpha$. Letting $\alpha = \alpha_0 - (A - \gamma) \cdot C \cdot e^{-\beta(A-\gamma)}$, (8)

anchors the curve to $\alpha_0$ at $t = A$. Further, if we want to set a maximum weight, so that $\max (PW_1) = w_m$, we let $C = \beta \cdot (\alpha - w_m) \cdot \frac{e^{\beta(A-\gamma)}-1}{\beta(A-\gamma)-e^{\beta(A-\gamma)}-1}$.

(9)

If $\alpha_0$, $A$, $\gamma$, $w_m$ and $\beta$ are given, we can treat (8) and (9) as two linear equations with two unknowns, and solve for $C$ and $\alpha$. Figure 3A shows (7) anchored to 1 at $t = 80$ for different values of $w_m$.

As mentioned, manipulating $\gamma$ will shift (7) right or left. It may be of particular interest to identify the shift needed to ensure that the maximum is moved to $t = 0$. This means that the slope is flat at $t = 0$, and gradually becomes steeper, so that gains for most individuals with poor lifetime health are given weights close to $w_m$. Differentiating (7) with respect to $t$, we find that the maximum is when $t = 1/\beta$. Hence, letting $\gamma = -1/\beta$ ensures that the maximum is at $t = 0$ (Figure 3B). Differentiating (7) once more with respect to $t$, the $t$ for which the slope is the steepest can be identified as $2/\beta$. Therefore, letting $\gamma = -2/\beta$ causes (7) to be at
its steepest when \( t = 0 \). In such a scenario, only gains to people with very poor health are
given weights close to \( w_m \) (Figure 3B).

So far the mathematical properties of the coefficients in (7) have been presented, but the
policy implications and conceptual meaning of these properties may not be obvious. The next
paragraphs are dedicated to the discussion of these implications. First, we consider \( \gamma \). If
\( \gamma > -1/\beta \), it is possible that \( PW(t_1) < PW(t_2) \) for \( t_1 < t_2 \). This makes sense if one thinks
that small gains to individuals with almost no lifetime health should be worth less than a
similar gain to an individual with more lifetime health. This could pertain to treatment of
children born with very severe conditions. In Figure 1 and Figure 2 the default value of \( \gamma \) was
0.04, so that the maximum was at \( t = 1/0.04 = 25 \) QALYs, which is probably too high in
most scenarios. When \( \gamma \leq -1/\beta \), a gain of, e.g., 1 QALY will always be weighted higher for
the individual with the least lifetime health. From now on we shall assume that \( \gamma \leq -1/\beta \).

Next we consider \( \alpha \). The particular value of \( \alpha \) is not necessarily easy to interpret, but because
this coefficient allows for the anchoring of the PW at a certain weight for a certain value of \( t \),
it is useful when selecting a reference value of \( t \) for which \( PW=1 \) (e.g., one may anchor \( PW \) at
1 when \( QALE=70 \)). Once a reference is chosen, the interpretation of the other coefficients is
easier. A large \( \beta \) corresponds to a narrow peak of the PW, ensuring a steep decline of the
curve. Now, gains to individuals with little lifetime health are given much higher weights than
gains to individuals with more lifetime health, whereas gains to individuals with intermediate
and high levels of lifetime health will be weighted similarly. A small \( \beta \) yields a flat peak, so
that the weight drops steadily as the initial health level improves (\( t \) increases). Although \( \beta \)
also will affect the range of possible weights, this property is more closely connected to the C.
The choice of \( C \) decides how much value will be assigned to a given \( t \) relative to the
reference. E.g., one may think that the maximum weight should be no more than three times
the weight of the reference. A very small \( C \) would ensure that all health gains were treated
almost the same.

Application of the priority weight function to experimental data

In order to illustrate the potential use of (7) in a CEA within a lifetime health framework, we
fit the function to data from a discrete choice experiment. The procedure involves two
conceptually simple steps. First, data must be collected, and then the curve must be fit.
However, both these steps may be performed a number of different ways. We shall focus on
an approach that is suitable for a dataset presented in Ottersen et al. (2014).

Ottersen et al. used a computer based questionnaire to ask 96 students in Norway how they
valued QALYs gained for people at different initial health levels, measured in “initial QALE”
(QALE without intervention). The reference case was a group of healthy people with
QALE=70 who were subject to an intervention that would give them 10 extra QALYs (i.e.,
move from QALE=70 to QALE=80). Respondents were asked to assign QALYs to
individuals with worse initial health (QALEs in the set \{10, 25, 40, 55\}) in such a way that they were indifferent between the different scenarios. Now, priority weights could be calculated for each individual respondent based on the ratio between the QALYs assigned to the unhealthy groups and the reference group. E.g., consider a respondent who was indifferent between a situation where the group with initial QALE=25 gained 5 QALYs (i.e., moving from 25 to 30) and the reference group gained 10 QALYs (i.e., moving from 70 to 80). This would mean that gains in the group with a worse initial health, where initial QALE=25, are regarded as more important than gains in the healthy group, where initial QALE=70. Mathematically, this extra weight can be expressed as \( V(25) = \frac{10}{5} = 2 \) relative to the reference group (QALE=70). We refer to Ottersen et al. for a more thorough discussion of data and study design.

One may organize the data as follows. Two lists, \( V \) and \( T \), are constructed. \( V \) contains all the weights assigned by the respondents, and \( T \) contains the initial QALEs, so that the weight \( V_i \) corresponds to the QALE \( T_i \). For purposes of estimating the coefficients of (7), we treat the observations as independent. Applying a least squares approach, we get that the sum of squared errors (SSE) is

\[
SSE = \sum_i \left( V_i - \alpha + (T_i - \gamma) \cdot c \cdot e^{-\beta(T_i - \gamma)} \right)^2. \tag{10}
\]

This expression may be minimized using a numerical routine or by solving the set of equations

\[
\frac{\partial}{\partial \alpha} SSE = 0, \quad \frac{\partial}{\partial \gamma} SSE = 0, \quad \frac{\partial}{\partial c} SSE = 0, \quad \text{and} \quad \frac{\partial}{\partial \beta} SSE = 0.
\]

If one wishes to anchor the curve to a specific value (e.g., the data from Ottersen et al. has \( V(70)=1 \)), it is straightforward to use (8) in (10). Setting maximum weights would imply using (9) in (10). Also, any coefficient in (10) may be kept constant.

Minimizing SSE ensures that the curve is fit to the mean \( V \) for each \( T \) in \{10, 25, 40, 55\}. Using the R function \texttt{optim()} to minimize SSE, we get

\[
PW_{\text{mean}} = -0.42 + (t + 22.2) \cdot 0.27 \cdot e^{-0.031 \cdot (t+22.2)}.
\tag{11}
\]

However, minimizing instead the sum of absolute errors,

\[
SAE = \sum_i |V_i - \alpha + (T_i - \gamma) \cdot c \cdot e^{-\beta(T_i - \gamma)}|, \tag{12}
\]

the curve is fit to the median. Now \( PW \) becomes

\[
PW_{\text{median}} = 0.79 + (t + 8.85) \cdot 0.17 \cdot e^{-0.053 \cdot (t+8.85)}.
\tag{13}
\]

Note that when fitting the curves, \( \alpha \) and \( \gamma \) were set to ensure that \( PW(10) \) was the maximum (i.e., \( PW(10) = w_m \)) and \( PW(70) = 1 \). Figure 4 shows (11) and (13) plotted with the empirical data. Note that the \( PW_{\text{mean}} \) is more affected by extreme values than are \( PW_{\text{median}} \).

[Figure 4 here]
**Data availability statement**

All necessary data and code to reproduce Figure 4 are available at [https://github.com/oeh041/Priority_weights](https://github.com/oeh041/Priority_weights) and [https://doi.org/10.5281/zenodo.3256151](https://doi.org/10.5281/zenodo.3256151)

**DISCUSSION**

We have presented a new priority weight function (PW) with key characteristics that are important in consistent priority setting across patient groups and populations. As demonstrated, the PW was suitable for CEA in a lifetime health framework. The PW is flexible and easy to adjust to a broad range of equity considerations by altering a limited number of comprehensible parameters, each with its distinct role. We may set the range of priority weights by modifying $C$, we may shift the curve upwards or downwards by modifying $\alpha$, we may shift the curve left or right by modifying $\gamma$ and we may set the “pointiness” of the curve by modifying $\beta$. These characteristics make it easy, e.g., to anchor the curve or to set maximum weights. Considering (11) and (13) (Figure 4), we see that $\beta_{\text{median}} > \beta_{\text{mean}}$, indicating that $PW_{\text{median}}$ decreases at a faster rate than $PW_{\text{mean}}$ (i.e., $PW_{\text{median}}$ is pointier). Further, we have that $C_{\text{median}} < C_{\text{mean}}$, meaning that $PW_{\text{mean}}$ adds more weight to the worse off than does $PW_{\text{median}}$.

Drawing the PW by hand would of course be more flexible than using (7). However, we argue that setting some restrictions for the behaviour of the PW is useful. First, it offers consistency when comparing different curves. Second, the space of possible PWs is limited by (7), but the major restrictions caused by (7) seem reasonable (smooth and maximum one hump). Hence, a large number of nonsensical PWs are impossible (discontinuous and multiple humps).

We did not account for discounting of future health. It is not clear how this would affect rankings of interventions. A positive discount rate would put different weights on equally sized gains that were experienced at different times from the start of intervention (e.g., 0.5 QALYs gained immediately is valued higher than 0.1 QALYs gained each year in five years, or 0.5 QALYs five years from now). Hence, interventions with immediate effects would be valued over interventions where the benefit occurs in the future. Also, discounting health gains would give a relatively low priority to those with large gains, who are expected to have less lifetime health. This is counterintuitive in a lifetime health priority perspective.

Although we do give an example of how priority preferences can be inferred from data, future work includes developing methodology to do inference about parameter estimates based on data from discrete-choice experiments. This would allow for the testing of differences in preferences between groups. E.g., conditioning on the other variables, one could test if men and women preferred different values of $C$ (determining the maximum weight) or $\gamma$ (determining the $t$ for which the maximum weight was obtained). When an appropriate PW (or a range of PWs) has been selected, it could be applied to evaluate the equity impact of different interventions on reduction of mortality or morbidity. Also, work should be done to
include non-health concerns like socioeconomic status and productivity in the model, and to
explore the consequences of priority-weighted CEAs across settings.

In this paper we do not intend to determine which PW is the most appropriate, but rather to
illustrate what can be achieved by a flexible general PW. However, we would like to stress
the fact that a flat PW is also a PW. In other words, there is no such thing as “not using a PW”
in CEA. As illustrated, our framework allows for the estimation of PWs based on empirical
data.

ACKNOWLEDGMENTS

We would like to acknowledge Ole Frithjof Norheim and Trygve Ottersen for their useful
comments during the creation of this manuscript.
REFERENCES


FIGURE LEGENDS

Figure 1. Properties of the coefficients of (7) when $\alpha = \gamma = 0$. Panel A: Priority weights according to t. Panel B: Altering $C$ causes an inflation or deflation. Panel C: Altering $\beta$ changes the “pointyness” of the hump, causing the peak to become more or less pronounced.

Figure 2. Properties of the coefficients of (7) when $C = 0.1658$ and $\beta = 0.04$. Panel A: Altering $\alpha$ causes an upward or downward shift. Panel B: Altering $\gamma$ causes a left or right shift.

Figure 3. Anchoring to 1 for $t = 80$, with $\beta = 0.04$. Panel A: Varying max $t$ ($w_m$) when $y = 0$. Panel B: Varying $y$ when $w_m = 2.5$. For $y = 0$, there is no shift. For $y = -1/\beta$ the curve is shifted left so that the maximum is at $t = 0$. For $y = -2/\beta$ the curve is shifted so that the slope is the steepest at $t = 0$.

Figure 4. Estimating PW from data. The red line minimizes (10) to estimate a PW based on the means (black +). The coefficients are $\alpha = -0.42$, $\gamma = -22.2$, $C = 0.27$, and $\beta = 0.031$. The pink line minimizes (12) to estimate a PW based on the medians (blue $\times$). The coefficients are $\alpha = 0.79$, $\gamma = -8.85$, $C = 0.17$, and $\beta = 0.053$. In both scenarios $PW(10)=w_m$ (maximum weight at $T=10$) and PW is anchored at $PW(70)=1$. 