A novel approach used outcome distribution curves to estimate the population-level impact of a public health intervention

Anna Sarkadi, Filipa Sampaio, Michael P. Kelly, Inna Feldman

Department of Women's and Children's Health, Uppsala University, 751 85 Uppsala, Sweden
National Institute of Health and Care Excellence, Centre for Public Health, London, SW1A 2BU UK
Institute of Public Health, General Practice and Primary Care Research Unit, University of Cambridge, Cambridge, CB2 0SR UK

Accepted 18 December 2013; Published online 13 April 2014

Abstract

Objectives: To provide an analytical framework within which public health interventions can be evaluated, present its mathematical proof, and demonstrate its use using real trial data.

Study Design and Setting: This article describes a method to assess population-level effects by describing change using the distribution curve. The area between the two overlapping distribution curves at baseline and follow-up represents the impact of the intervention, that is, the proportion of the target population that benefited from the intervention.

Results: Using trial data from a parenting program, empirical proof of the idea is demonstrated on a measure of behavioral problems in 355 preschoolers using the Gaussian distribution curve. The intervention group had a 12% [9%—17%] health gain, whereas the control group had 3% [1%—7%]. In addition, for the subgroup of parents with lower education, the intervention produced a 15% [6%—25%] improvement, whereas for the group of parents with higher education the net health gain was 6% [4%—16%].

Conclusion: It is possible to calculate the impact of public health interventions by using the distribution curve of a variable, which requires knowing the distribution function. The method can be used to assess the differential impact of population interventions and their potential to improve health inequities. © 2014 The Authors. Published by Elsevier Inc. Open access under CC BY-NC-ND license.

Keywords: Public health; Intervention studies; Normal distribution; Area under the curve; Primary prevention; Parenting education

1. Introduction

Generating population health improvement in ways that produce more equitable health outcomes is difficult [1]. Interpreting the results of interventions to achieve these goals is equally challenging [2]. Using analytical methods developed for clinical randomized controlled trials (RCTs) to investigate naturally complex interventions in dynamic and complicated settings carries the danger of controlling out of the investigation the very factors that are of intrinsic interest and of focusing on individual and not population-level change. The essence of trial design is to isolate single causes and effects when, in reality, linear causal pathways are seldom relevant for public health practice [3]. Trials of public health interventions are often disappointing; effect sizes are small and sometimes they do not reach statistical significance [4—6], although publication bias probably conceals most trials that do not show significant effects. This article demonstrates a different technique for considering effects in public health trials by looking at the data through a population health lens.

We propose a method to look at population outcomes in their own right in the quest of understanding how public health interventions work. Rather than considering the effects of interventions on the individual level, we view the level and distribution of a certain outcome measure in the
population as the unit of interest. We do not apply the idea of representativeness on a subsample to then project these effects on a population: instead we aim to understand the underlying mechanisms [7] by which population effects play out when different interventions are applied.

2. Background

2.1. The population health approach

Rose [8] popularized the population health approach. Overall population change, in, for example, average hypertension, was the goal, rather than individual outcomes. Taking this approach seriously requires researchers to describe causal pathways that are not yet fully understood. Consider the large network study by Christakis and Fowler [9] where positive effects in the whole social network of those quitting smoking were registered, along with the marginalization of smokers in the network. The social-level explanations underpinning these results (such as the mechanisms whereby networks produce these effects) are much less well understood than the etiology of lung cancer for example, yet are critical if we are to advance population-level approaches [10].

2.1.1. The goal of modern public health interventions

There is an increasing body of evidence suggesting that health equity within a population benefits the population as a whole. A goal of many contemporary health policies is just that—closing the health inequity gap [11]. A common way of describing a variable on the population level is using the distribution function with the mean and standard deviation (SD) values of a certain relevant measure determining the shape of the curve. The most often used distribution function is the normal (Gaussian) distribution (Fig. 1A–C), but it is increasingly recognized that not all variables of importance in public health are normally distributed, but can be described by another function, for example, by a gamma-type curve (Fig. 1D).

The goal of a public health intervention is then—simply put—to move the population distribution curve of the targeted outcome or risk factor (the exposure) toward healthier levels and to decrease the distribution of the outcome, implying higher proportions of the population being within the healthy intervals (Fig. 1A–D), or being less exposed to a certain risk factor. Both the health state of the population and its exposure to risks can be expressed by indicating which percentage of the population is above or below any certain value of interest by calculating the area under the curve cutoff by a vertical line drawn at that value of the x-axis, exactly as it is done by the verticals indicating the SD values of a population mean. We propose a method that does exactly this: expresses health improvement in the percentage of the population improved.

2.1.2. Evaluating public health interventions

For most major public health issues, there is a cluster of indicators, rather than a single outcome that signifies health. However, one of the unintended consequences of the application of the principles of evidence-based medicine to public health has been a tendency to focus on determining effect size and significance levels in individual-level variables rather than thinking about the issue in population terms [10,12]. Also, in the quest for statistical significance, surrogate measures rather than direct patient outcomes are sometimes used [13].

Studies in the area of public health often collect information on variables that can be presented on a continuous scale. Although both the mean value and SD measures are necessary for the analyses presented in RCTs, it is rather unusual to actually draw the population distribution curves and estimate the outcomes based on these.

Let us suppose that Fig. 1A–D represent the distribution curves of outcome variables of hypothetical public health interventions. Fig. 1A describes a scenario where the population mean was effectively decreased without affecting the distribution of the variable. This could be possible through a universal program that actually manages to reach all the different segments of the population equally, but not proportionately depending on need. Thus, inequities would still remain, but the population health would improve (see the size of the gray area).

Fig. 1B represents a scenario where the SD of a variable has decreased, but the population mean is unaltered. This could be the result of a targeted intervention that has successfully addressed the needs of a population at risk for the studied outcome. Inequities in health for this outcome have then decreased, but the health gain for the population as such is less, as indicated by the smaller size of the gray area than in Fig. 1A. This scenario introduces some possible ethical concerns as decreases in values/areas on the right side of the curve necessarily imply increases on
the left side. Whether or not this becomes a real concern depends on the outcome we are aiming to address.

For individuals who already from start have a health outcome score better than normal and as a result of the intervention move closer to the mean value, this scenario does not necessarily mean health losses, as they are moving toward mean population values—which are normal per definition. If we analyze the effect of an intervention to normalize weight by height in adolescents in the population, then the right part of the spectrum corresponds to those with values higher than normal and the left part to values lower than normal. In the example in Fig. 1B, both changes at the lower end of the spectrum and at the higher end of the spectrum mean improvement and are desirable. However, there are other possible scenarios where interventions that benefit the population as a whole will not benefit the individual at all—a phenomenon Rose [8] called the “prevention paradox”.

Nevertheless, the possible dilemmas raised by Fig. 1B are primarily mathematical and theoretical in nature. In real-life settings, what tends to happen is that interventions are differentially distributed in the population with the lowest quintile receiving much less effective interventions than the higher quintiles [14]. Thus, health benefits from population interventions are often unequally distributed [15]. However, the better the overall coverage of interventions, the less these inequalities are present [14]. Therefore, looking at changes in outcome distribution is a powerful way to monitor the success of public health interventions.

Fig. 1C and D describe a situation where both the population mean and the SD has decreased, providing the largest health gain for the population of all three models. Such outcomes could be possible if a universal intervention proportionately reached different segments of the population based on need, that is, more intensive interventions to those at risk, while not failing to address the needs of the majority. Another option would be if the intervention in itself had the potential to differentially affect segments of the population based on their socioeconomic or risk profiles. That may sound like science fiction—and often enough the opposite, that interventions differentially affect the more advantaged groups is true [6,16], but is a possibility, as will be demonstrated in this article.

For all types of changes described previously, the shift can be depicted using the actual distribution curves before and after an intervention if the distribution function is known analytically. In other cases, a numerical method can be used, that is, calculating the actual distribution curves using real values of the outcome at hand.

2.1.3. Parenting programs as public health interventions

The example used to demonstrate our argument in this article is of parenting programs as public health interventions, especially those allowing self-selection into the intervention. The trials often report small effect sizes with low or no significance levels, when using conventional statistical analyses [5,17]. However, small effects in a single study could be of importance from a population health point of view. This is especially true if the intervention is considered as part of a “bigger picture” [2], where interventions overlap and build on one another as part of a broader public
distribution function, defined by the formula

\[ \phi = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\} \]

where \( \phi \) is the distribution of the variable in the target population, \( \mu \) is the mean value, \( \sigma \) is the SD of the distribution, and \( x \) is a random variable.

When trying to use the Gaussian distribution curve as a measure of effect of an intervention, account should be taken of the distribution curve of the outcome of interest at baseline (\( \phi_1 \)) and at follow-up (\( \phi_2 \)). The intervention is successful if there is a shift of the baseline distribution curve in the desirable (healthier) direction. This type of shift can assume three different forms as exemplified in Fig. 1A–C. The same approach can be used for other types of distribution, exemplified in Fig. 1D, with a different (\( \phi \)) function.

The difference between the distribution at baseline and the distribution at follow-up represents the impact of the intervention. Mathematically, it can be calculated as the area difference between the two overlapping curves. This area is represented by the space from the intersection point between the two curves \( 2 \) to infinity \( (\infty) \), the latter representing the decreasing probability of obtaining a certain value as one moves along the curve. This area is colored gray in Fig. 1A–C. The size of this area represents the part of the target population that benefited from the intervention. It can be used as a measurement of the impact of an intervention or total health gain for the population. To achieve a maximum impact of an intervention on population outcomes, the overlapping area between the two curves should be maximized. A calculation method used to estimate the gray area between the two curves is presented in Appendix A (see at www.jclinepi.com). Calculation method for the precision of the estimate (confidence interval [CI]) is presented in Appendix B (see at www.jclinepi.com).

In this article, we demonstrate the practical application of this approach using real data from a case example, a universal parenting program to prevent child externalizing behavior problems.

3. Methods

3.1. Using the distribution curve to analyze outcomes

Many of the measures used widely in population health research can be mathematically described by a distribution. The normal (Gaussian) distribution is often relevant and widely used to describe the prevalence of specific risk factors and health problems in populations. The area under a normal distribution curve represents the sum of probabilities of obtaining every possible value for a variable and is equal to one. Mathematically, the shape of the curve depends on two parameters, namely a mean value (\( \mu \)) and an SD value (\( \sigma \)) representing the dispersion of a certain problem from the mean level of the problem in that population. The distribution of a certain health variable is defined by its mean and its dispersion (SD) from the mean in the population. In mathematical terms, this dispersion variable can be used to describe the probability density for that health variable, creating what is called a normal distribution function, defined by the formula

\[ \phi = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\} \]

3.2. The case example

3.2.1. The intervention

The Triple P parenting program is an intervention based on social learning theory, evaluated in a number of RCTs [18]. The Triple P program has been evaluated at different intensity levels [19] with at-risk groups [20] and in different countries [21,22], finding that following the intervention parents develop better parenting skills, increased confidence in parenting and lessened depression [23], and that their children have fewer behavior problems [24]. At the population level, Triple P has shown effects in terms of reduced incidence of child abuse and foster care placements [25]. The Triple P program was found to have moderate evidence for preventing mental health problems [26] and has had difficulties demonstrating effect in population trials allowing self-selection and with individual-level outcomes [5,17].

A cluster RCT of Triple P was conducted in Uppsala, Sweden, with preschools as an arena and preschool teachers as practitioners delivering the program to parents of children aged 2–5 years. Levels 2 (open parenting seminars lasting 90 min) and 3 (individual consultation sessions of 20–30 min up to four times) of the program were offered to parents at 11 preschools randomized to the intervention group—10 other preschools comprised the control group. Level 1, a universal media campaign highlighting parenting issues, was not offered not to "pollute" the control arm. We allowed self-selection of parents into the intervention, and thus, a real-world scenario was created. Efforts were put into making the intervention as readily accessible as possible rather than targeting vulnerable groups, according to the principles of proportional universalism. [11] Thus, child care services and interpretation were offered as well as different time options for participation, including weekends for working parents.

3.2.2. Exposure to intervention

Program exposure was registered. Over 12 months, 30% of the mothers and 16% of the fathers were self-selected into the intervention.
3.2.3. Outcome measures
Outcomes were collected on the individual level through questionnaires at baseline, at 6 and 12 months. Health outcomes were measured using the Eyberg Child Behavior Inventory (ECBI), [27] a commonly used parent report measure of externalizing behavior problems in children aged between 3 and 16 years. Parents were asked to indicate, on a seven-point Likert scale, how often a behavior occurs. A version with 22 items was used in this study, which was previously validated in a Swedish sample [27]. The ECBI total score ranges from 22 to 154 and is an aggregate of all intensity scores—higher scores indicate more problems.

From the 488 children initially included in the trial, data on outcomes measures was available for 355 children (73%). Outcomes were measured based on only one parent’s assessment of their child’s behavior. When assessments from both parents were available, only mothers’ assessments were selected because more mothers than fathers provided follow-up questionnaires.

4. Results
Statistical analyses showed that the health outcomes (the total ECBI intensity scores) followed the normal (Gaussian) distribution. The distribution of total ECBI intensity scores at baseline and 12-months follow-up for both the intervention and control groups are presented in Fig. 2A and B, respectively. Repeated-measures analysis of variance showed statistically significant differences between the intervention and control groups at 12-months follow-up. However, multilevel modeling—used to account for the clustering of our data based on preschools—did not reveal significant effects on child outcomes. We therefore looked for a different way to describe our findings.

The gray area between the two curves represents the health gains between baseline and follow-up. By using equation (A1:3) and (A1:10) and knowing the values of the parameters involved (Appendix A), the gray area between the two curves can be calculated for both the intervention and the control group. For given values \( \mu_1 = 57.48, \mu_2 = 53.55 \) and \( \sigma_1 = 13.81, \sigma_2 = 12.60, \) and calculated \( \lambda = 59.15 \) we obtain for health gain the result:

\[
f(\mu_1, \sigma_1, \mu_2, \sigma_2) = F_2 - F_1 = \Phi\left(\frac{59.15 - 53.55}{12.60}\right) - \Phi\left(\frac{59.15 - 57.48}{13.81}\right) = \Phi(0.44517) - \Phi(0.12159) = 0.12
\]

Let us now calculate the CI of \( f(\mu_1, \sigma_1, \mu_2, \sigma_2) \) (Appendix B). The 95% CI for the parameters are:

\[
\mu_1 = 57.48 (56.68 - 59.28); \mu_2 = 53.55 (51.49 - 55.61); \\
\sigma_1 = 13.81 (12.65 - 15.21); \sigma_2 = 12.60 (11.30 - 14.24)
\]

According to (A2:1), 95% CI: \( f(\mu_1, \sigma_1, \mu_2, \sigma_2) = (0.09 - 0.17) \) (11)

This means that 12% of the children in the intervention group, 95% CI [9%, 17%], have improved health outcomes. The calculations for the control group are based on the following parameters:

\[
\mu_1 = 53.69 (51.33 - 56.06); \mu_2 = 52.56 (49.90 - 55.21) \\
\sigma_1 = 13.09 (11.62 - 14.99); \sigma_2 = 12.54 (10.92 - 14.72)
\]

According to (A1:3), \( \lambda = 58.31 \) and the estimation of health gains according to (A1:10) is thus:

\[
f(\mu_1, \sigma_1, \mu_2, \sigma_2) = F_2 - F_1 = \Phi\left(\frac{58.31 - 52.56}{12.54}\right) - \Phi\left(\frac{58.31 - 53.69}{13.09}\right) = \Phi(0.13818) - \Phi(0.45992) = 0.03
\]

and the CI for change in the control group:

\[
95\% \text{ CI: } f(\mu_1, \sigma_1, \mu_2, \sigma_2) = (0.01 - 0.07)
\]
This means that approximately 3% of the children in the control group (95% CI [1%, 7%]) have improved health outcomes.

To calculate the absolute health gain between the intervention and the control group, the difference between both groups’ health gains can be estimated as:

$$\Delta = 0.12 - 0.03 = 0.09$$

Thus, the absolute health gain after the intervention was 9%. Because the CIs of health gain estimate for intervention and the control group do not overlap, this difference can be considered as a significant one.

For variables that are not normally distributed, a different set of mathematical formulae would have to be used as the distribution (ϕ) function, as well as the method to calculate the intersection point (λ), and the integrals used to calculate the area under the curve would be different. However, the principle remains the same.

4. Social inequalities and the impact of population interventions

We have used the same approach to estimate the impact of the parenting program on the target population’s health outcomes based on educational level. Participants were divided into groups with lower (high school or less) and higher educational levels. We calculated the proportion of the target population in each group that benefited from the intervention. Our calculations show that 15% CI [6% – 25%] of the participants with lower educational level have improved health outcomes, whereas only 6% CI [4% – 16%] of the participants with higher educational level benefited from the intervention (Fig. 3A and B). However, this difference did not reach statistical significance (see the overlapping CIs) in this small sample, but serves well for demonstration purposes.

5. Discussion

5.1. The “mission impossible” of public health intervention researchers

A common problem in trials attempting to examine the effectiveness of public health interventions is that the results are either nonsignificant, inconclusive, or effect sizes seem very small. Normal populations are normal by definition; and thus, there is relatively little potential for demonstrable improvement using traditional statistical inference methods developed for RCTs. Large data sets are often required for enough statistical power, which most often occurs in what could be called “natural experiments,” such as guideline introductions during an ongoing longitudinal study [15], or comparative studies of different health or reimbursement systems and their effects on health outcomes in various countries [28].

In this article, we provide a new way of assessing results in public health intervention trials through a population lens. We argue that the basic reason why someone might want to perform population interventions is to alter the distribution of a certain risk factor in the population. We provide a conceptual framework and mathematical formulae that can be used to calculate the total health gains in an intervention using the Gaussian distribution of the variable in the studied population before and after the intervention. We use the example of an RCT of a parenting program conducted in Sweden to demonstrate the method. Although the mathematical forms involved in calculating the overlapping area between the curves for other distribution functions than normal are not presented here, the same principle can be used.

Carr [29] has used a similar approach to calculate the impact of experimental single-case design interventions on children’s mental health. The method he describes allows for estimating the percentage of individuals who benefited from an intervention. However, Carr’s method
does not take into account the whole distribution of the target variable. When dealing with public health interventions, the aim is to be able to assess the impact for the whole target population taking into account all possible health gains, either improvement in mean or in SDs. Therefore, we believe that the approach we propose better serves the purposes of evaluating public health interventions. In addition, the method can help understand how interventions can affect social inequities in health.

5.2. Self-selection: pain or gain?

According to the prevention paradox [8], interventions that are valuable for the population as a whole may not necessarily benefit the individual. The prevention paradox may be acceptable if no harm is done and if the cost of intervention is low per person. Parenting programs were not originally intended for universal use but are increasingly applied as population-based interventions. Because of the costs and the effort involved, it would not be desirable to aim for universal uptake of parenting programs and the prevention paradox could become a problem.

On the other hand, when we allow self-selection, and thus address the problem of the prevention paradox, exposure to the intervention will be reduced. One might think that program effectiveness in terms of population effect would also diminish and selection bias would be introduced instead. However, if the intervention has the potential to differentially affect segments of the population based on their socioeconomic or risk profiles, selection becomes less of a problem. In fact, we have demonstrated in an earlier study that mothers self-selected into the Triple P intervention based on experiences of exacerbated child behavior problems (R. Salari, M. Wells, and A. Sarkadi, unpublished data, 2013). Although mothers with higher educational levels were more likely to participate, those with lower educational levels still seemed to benefit more, as demonstrated in this article. One could say that the self-selection of participants became an asset in the study in that those with higher problem levels participated and among them—the program differentially affected those with lower education. If public health interventions can be developed to have differential impact based on risk profiles [30], some of the major challenges of public health [31] could be effectively addressed.

5.3. Limitations

The method has some limitations that should be observed. This proposed approach does not allow controlling for potentially confounding factors, something that one often wants to do when analyzing a trial. What the approach does is estimate an effect on the population level “taking all things together.” However, subgroup analysis using the methods do allow the researchers to gain a fuller understanding of how the different subpopulations might have been affected is described in this article.

The method does not solve the problem of the lack of a comparison group that many public health interventions suffer from. Thus, limitations to what kind of causal inferences can be drawn will always be down to the design of the study.

Finally, estimating the proportion of the population showing health gain from the intervention will have to be interpreted based on the outcome measure at hand. The method does not establish a certain effective cutoff or threshold effect and does not have implications for clinical significance per se. This has to be decided from case to case. Also, by geometrical symmetry, gains at one end mean losses at the other end of the distribution curve. Whether or not these are of importance have to be judged depending on the outcome or exposure at hand. In any case, such redistribution effects are not caused by the method per se: we merely provide a tool to assess the possible shape and form of redistribution of the outcome measure occurring in the population.

6. Conclusions

It is possible to calculate the impact of public health interventions by using the distribution curve of a variable. By calculating CIs, the precision of the estimate of the observed changes can be can be assessed. The method can be used to evaluate the differential impact of population health measures and their potential to improve health inequities. The method does not, however, allow for confounder control. Further methodological development will be necessary to be able to take into account the differential exposure to risk factors and interventions in the population.

Acknowledgments

The authors wish to thank Professor Michael Liberman, Nordic Institute for Theoretical Physics, for assistance with the mathematical forms used in the paper. They also wish to express their gratitude to Dr. Alistair Fischer at NICE for review of the statistical and mathematical methods used.

Appendix

Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jclinepi.2013.12.012.

References


