Article title: From clinical trial efficacy to real-life effectiveness: why conventional metrics do not work
Journal name: Drugs – Real World Outcomes
Author names: Jean-Pierre Boissel, Frédéric Cogny, Nicholas Marko, François-Henri Boissel
Corresponding author: François-Henri Boissel, Novadiscovery, Lyon, France; francois.boissel@novadiscovery.com

Why we need to improve methods for translating efficacy results from clinical trials to real-life effectiveness in populations

Generally the endpoints in phase III clinical trials reflect clinical outcomes that are binary variables, such as death or occurrence of a stroke or cancer relapse. The efficacy estimate is calculated using the rate of outcomes observed in the control group (R_c) and the experimental (treated) group (R_t). These are analysed using summary metrics (or statistics) of treatment efficacy, such as the odds ratio (OR), relative risk (RR), relative benefit (RB), absolute benefit (AB) and number needed to treat (NNT). Tables 1 and 2 illustrates how the two rates are calculated and the summary metrics are derived.

Table 1: Estimation of treatment efficacy using results from a randomised controlled trial: example for a binary outcome. In a clinical trial, the occurrence of at least one outcome in a patient is called an endpoint. The outcome rates can be calculated in the treated group, R_t as: a/N_1 and in the control group, R_c, as b/N_2.

		Number of patients		
		With endpoint	Without endpoint	Randomised
Group	Treated	a	N_1 - a	N_1
	Control	b	N_2 - b	N_2

Table 2: Equation used to estimate current treatment efficacy metrics. Using the calculated outcome rates in the two groups, R_t and R_c treatment efficacy metrics can be estimated

Metrics	As a function of R_c and R_t	Connecting metrics
Odds ratio (OR)	$[R_t /(1 - R_t)]/[R_c /(1 – R_c)]$	$RR[(1 – R_c)/(1- R_t)]$
Relative risk (RR)	R_t / R_c	
Relative benefit (RB)	$1 - R_t / R_c$	$1 - RR$
Absolute benefit (AB)	$AB = R_c - R_t$	
Number needed to treat (NNT)	$1/(R_c - R_t)$	$1/AB$

Drug license approval decisions are based mainly on the available evidence about the safety and efficacy of the drug, with this evidence being summarised using *OR* and *RR*, and more recently, *NNT*. A couple of decades ago, efficacy metrics were only used to summarise randomised clinical trial findings and to help interpret them once a statistically significant difference between the outcome rates between the treatment and control groups had been established. Statistical inference is used to decide whether the observed difference is due to a real effect of the treatment being evaluated or if it is due to chance alone.
With the growing interest in evidence-based medicine, the public health impact has become an important element to be considered, particularly since health budgets are increasingly

restricted. As a result, cost-effectiveness research (CER) has been developed [1]. The importance of real-world outcomes and cost-effectiveness in HTA has grown, with a focus on more rigorous and reliable extrapolation of results from populations included in phase III clinical trials to larger real-world populations [2]. In summary, in a few decades we have shifted from a qualitative vision of treatment efficacy, i.e., 'is the treatment efficacious?' to a quantitative vision of effectiveness, i.e., 'how much benefit for a patient, a group of patients or a population from the treatment?'

Despite recent efforts to standardise the translation process of efficacy metrics from clinical trials to effectiveness metrics for clinical practice, there is no universally-accepted methodology. The translation process generally relies on input from 'expert' opinion [3]. The first step to address this issue is to decide which is the best metric to express effectiveness at a population level. Examples of the metrics that can be used to extrapolate from clinical trial summary efficacy data to real-world effectiveness and cost-effectiveness; represented as population metrics are given in Table 3. Each population metric starts with a specific trial efficacy metric that is used in an equation for the translation process. This uses specific population descriptors, and provides a specific population efficacy metric (Table 2).

Table 3: Trial efficacy metrics used in the simulation to extrapolate to population effectiveness descriptors, NPE_{pop} and NNT_{pop}

Trial efficacy metric	Population descriptor	Translation equation	Resulting population metric
OR_{trial}	N	$NxOR_{trial}$	NPE_{pop}
OR_{trial}	Ec_{pop}	$Ec_{pop}xOR_{trial}$	NPE_{pop}
RR_{trial}	N	$N.RR_{trial}$	NPE_{pop}
RR_{trial}	Ec_{pop}	$Ec_{pop}.RR_{trial}$	NPE_{pop}
RR_{trial}	$aveRc_{pop}$	*equation (5) below*	NNT_{pop}
NNT_{trial}	*none*	NNT_{trial}	NNT_{pop}
AB_{trial}	N	$N.AB_{trial}$	NPE_{pop}

OR: odds ratio; RR: relative risk; NNT: number needed to treat; AB: absolute benefit; N: population size; Ec_{pop}: total events if untreated; $aveRc_{pop}$: average rate of event if untreated; NPE_{pop}: number of prevented events in the population; NNT_{pop}; number needed to treat in the population.

There are three population metrics required for this translation process: N (population size), E_{cpop} (total events if untreated), $_{ave}Rc_{pop}$ (average rate of event if untreated). For example, the current translation process to predict the number needed to treat at the population level (NNT_{pop},) relies on the following translation equation, based on the average R_c in the population and the trial RR, where NNT_{trans} is assumed to predict NNT_{pop} and Rc_{pop} is equivalent to $_{ave}Rc_{pop}$:

Equation (1)

$$NNT_{trans} = \frac{1}{Rc_{pop}(1 - RR_{trial})}$$

Parameters required for translation of trial efficacy to real-world efficacy

Outcomes of interest

The first step in translating efficacy into real-world effectiveness requires the identification of an outcome that is meaningful for risk-benefit assessment in the real-world population. Contacts with the healthcare system and cost are important for health status, at the real-world population level. We can consider a clinical condition that is binary i.e., either present or absent and which drives an individual to seek healthcare as a meaningful outcome. Impairment of quality of life can also be considered as a binary outcome if we assume that individuals will seek healthcare when their quality of life is impaired below a certain threshold,

albeit a subjective threshold. Thus, in this paper, we have considered outcomes only as binary e.g., acute myocardial infarction, death, occurrence of a disease (for a preventive treatment) or impairment of quality of life as outcomes of interest that the treatment aims to prevent.

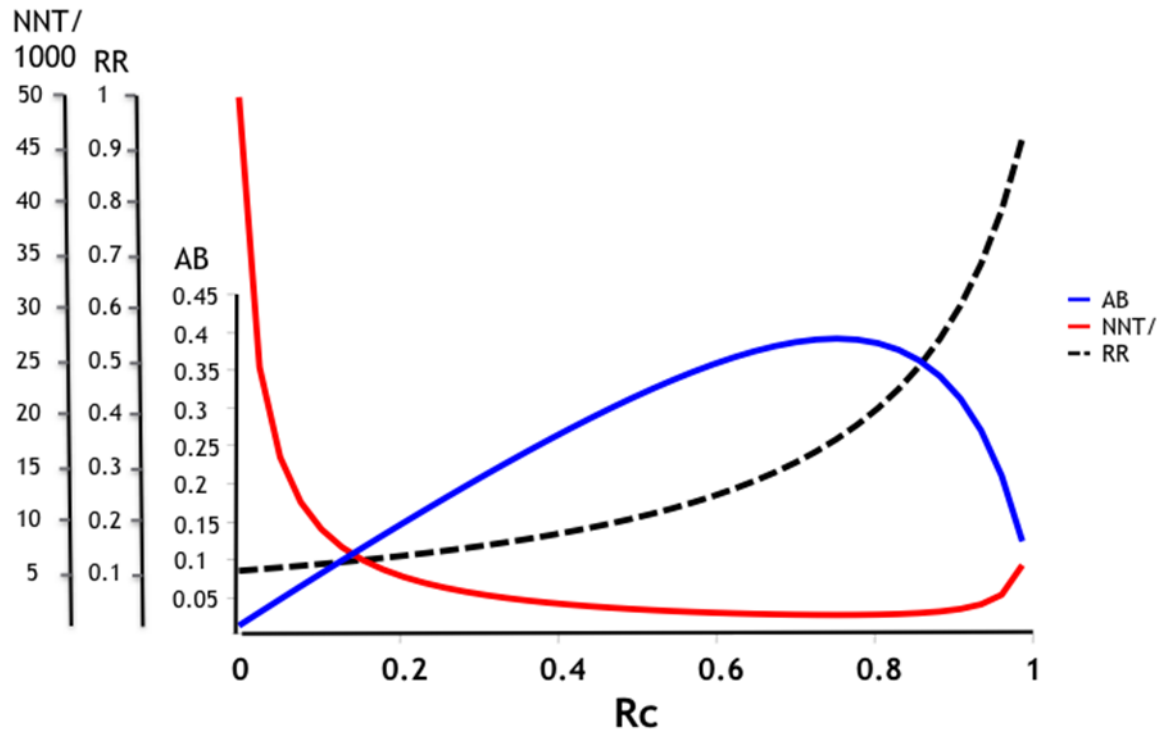
Current efficacy metrics calculated using clinical trial data

The event rates (or risk) of the event occurring in the treatment and control groups characterise the trial groups and are used to calculate efficacy metrics such as OR, RR, AB and NNT (Table 1). However, using frequencies observed in a population are not applicable at the individual patient-level because either the event occurs or it does not for a patient. In addition, since the size of the benefit is patient-dependent, as we will see below, the probability of the event occurring in a patient has to be estimated with an appropriate algorithm.

Sensitivity of efficacy metrics to the risk in the control group

Efficacy metrics are calculated using the risk (or event rate) in the control group as the baseline risk i.e., the event rate without treatment. However, since this risk varies between patients and between trials, the efficacy metrics will also vary between patients and between trials. However, it is often assumed that *ORs* and *RRs* are constant and that only *AB* and *NNT* vary. This assumption is true when the relation between R_t and R_c is linear and $R_t = 0$ when $R_c = 0$ [4]. In reality, the relation between R_t and R_c, which has been called the effect model (EM), is linear only in a few particular situations [5]. Fig 1 shows an example of a situation where *OR* is constant while the three other metrics vary with R_c [6].

Fig 1: Example of the variability of RR, AB and NNT as a function of risk in the control group, Rc when OR is constant (Wang model) [6]

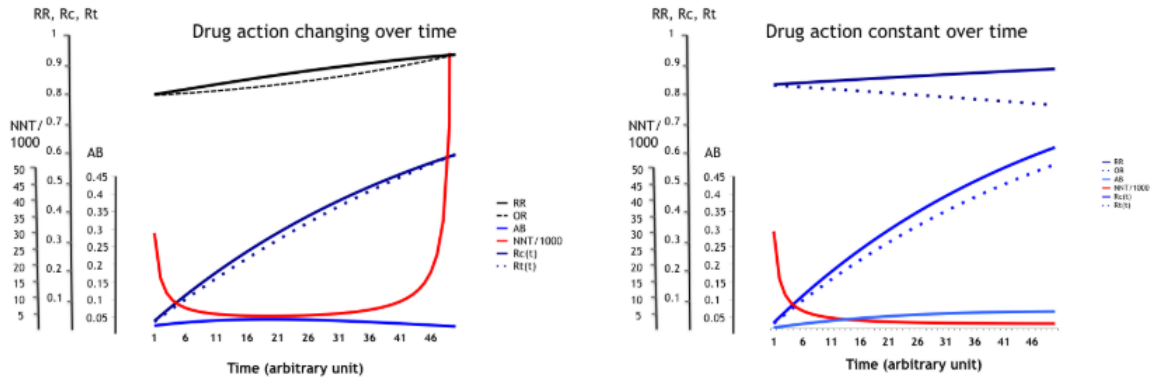


Effect of time on efficacy metrics for chronic treatment

Although it is not completely understood why the efficacy of chronic treatment can vary over time, the emergence of resistance is one factor responsible for decreasing efficacy over time in the same population for some anticancer drugs [7, 8]. We simulated the variation in efficacy metrics for a drug whose activity decreased over time (Fig 2a). The drug activity was defined as the ratio of treatment hazard over control hazard, where hazard was the instantaneous risk of the event. However, even when the hazard ratio was constant, i.e. drug activity was constant, the efficacy metrics varied over time (Fig 2b). *AB* and *NNT* were particularly sensitive to time. Since R_c varies over time, the efficacy metrics, both at the trial and population levels, are only relevant for a given period.

Fig 2: a) Simulation of changes in efficacy metrics values over time when treatment action decreases as treatment duration increases; b) Simulation of changes in efficacy metrics values over time when treatment action remains constant as treatment duration increases. In these examples, time is expressed in an arbitrary unit

a) b)



Population-level effectiveness metrics

It is important to identify the most appropriate metric to use before translating efficacy observed in a clinical trial population to effectiveness in a real-world population. In the following, to avoid confusion, the metrics computed for population data are marked by the subscript 'pop', e.g. OR_{pop}.

Using clinical trial efficacy metrics as population effectiveness metrics

The 'average' event rates can be computed as the means of the corresponding distributions in the overall untreated and treated population, respectively, assuming for the latter that none of the treated individuals have compliance issues. These variables can then be used to calculate the population level metrics in the same way as the clinical trial metrics, but they are not very informative because they do take into consideration the structure of the population.

Population-specific metrics

Number of prevented events

The number of prevented events (NPE_{pop}) is the difference, at any time point, between the number of individuals in the population who have experienced (observed) or who will experience (predicted) at least one outcome when no individual is treated and when all individuals are treated. The NPE_{pop} is what we need to know about the effectiveness of the new treatment in the population of interest over a specific period. As previously stated, the outcome can be either the occurrence of a disease (prevention), a clinical event (e.g. a myocardial infarction or cancer recurrence), or an unacceptable impairment of quality of life.

Number needed to treat and equivalent

NNT is a popular metric because it is perceived as a single, simple measurement. Although it is derived from R_t and R_c, NNT is not a pure efficacy/effectiveness metric and requires additional information for use. *NNTs* are calcuated for a group, but they can also be used by doctors in their prescription decision for an individual patient, which is possible since *NNTs* are the inverse of AB, which is a patient-based matric. At a population level, there are several similar effectiveness metrics that are calculated in a similar way to the *NNT* that is computed using trial data:

NNT_{pop}

Equation (2)

$$NNT_{pop} = \frac{1}{aveRc_{pop} - aveRt_{pop}}$$

$_{ave}NNT_{pop}$

Unlike the previous *NNT*, $_{ave}NNT_{pop}$ is the sum of the individual *NNT*s:

Equation (3)

$$aveNNT_{pop} = \left(\frac{1}{N}\right)\sum_{k=1}^{N}\frac{1}{AB}_k$$

$_{equiv}NNT_{pop}$

Dividing the number of treated patients by the number of prevented events gives rise to another *NNT*-like metric, $_{equiv}NNT_{pop}$, that expresses more directly the efficiency of the treatment:

Equation (4)

$$equivNNT_{pop} = \frac{N}{NPE}$$

Actually, NNT_{pop} and $_{equiv}NNT_{pop}$ are equivalent since:

$$NNT_{pop} = \frac{1}{aveRc_{pop} - aveRt_{pop}}$$

hence:

$$NNT_{pop} = {1}\Big/{\left(\frac{1}{N}\right)\sum_{k=1}^{N}(Rc_k - Rt_k)}$$

thus:

$$NNT_{pop} = {N}\Big/{\sum_{k=1}^{N}AB_k}$$

and:

$$NNT_{pop} = \frac{N}{NPE}$$
$$NNT_{pop} = equivNNT_{pop}$$

The Wang model

The Wang model is the simplest model of drug action on a clinical outcome that takes into consideration the main features of both the drug's pharmacological action on its biological target and the consequences on the course of a disease [6]. It assumes that the probability of the outcome under treatment (or the event rate, R_t) follows a logistic function of the drug's pharmacodynamic effect with two parameters: β_0, the intercept, and S, the coefficient of E, which can be interpreted as the scale of the drug effect size [9].

Then if we assume that the treatment affects E through a direct pharmacodynamic dose-response model, the Hill model, where E is the pharmacodynamic effect and E_0 is the baseline value of E; E_{max} is the maximum theoretical effect; D is the dose; ED_{50} is the dose at which 50% of the maximum effect is achieved, and γ is the sigmoidicity parameter (Table 3) [10]. The event rate in treated individuals, R_t can be calculated as shown in equation (5):

$$Rt = \frac{Rc.e^{S.\frac{E_{max}.D^{\gamma}}{ED_{50}^{\gamma}+D^{\gamma}}}}{1-Rc+Rc.e^{S.\frac{E_{max}.D^{\gamma}}{ED_{50}^{\gamma}+D^{\gamma}}}}$$

For a patient with an event rate without treatment of R_c, the rate under treatment R_t, can be calculated with equation 5. The equation was used to determine the population benefit of treatment with drug 1 or 2 in population A (trial 1) and in a non-random sample of population A (trial 2).

Calculations

We calculated the true NPE and the average NNT for the three populations. For a given drug, Rci, the central value in the range i for the population under consideration, was multiplied by the number of patients in this range and by ABi, the absolute benefit from the drug. ABi was calculated from the value of Rti, the event rate under treatment with the drug, given by the Wang model: ABi = Rci – Rti. The number of prevented events with the drug was the sum of the ABi across all ranges of I and the average ABi was NPE/N. The true values for NPE, NNT and for the other usual efficacy metrics (RR, OR, AB) were compared with the values calculated using the trial summary data, as is usually done in the current translation process.

Results from simulated translation

Table 4 summarises the values of estimated efficacy metrics based on the results from trials 1 and 2 and the values of metrics calculated in populations A, B and C for each of the two drugs. These data show that use of the trial efficacy metrics for inferring population benefit results in erroneous population metrics.

The bias was lowest when AB, computed with trial summary data, was used for the translation. The NNTpops were all overestimated, except for NNTpop for population A, translated from trial 1 summary data, the ideal trial since the whole population was included in the trial.

Table 4 Comparison of the values for efficacy metrics calculated using data from trials 1 and 2 for populations A, B and C

	Drug 1					Drug 2				
	Trial 1	Trial 2	Pop A	Pop B	Pop C	Trial 1	Trial 2	Pop A	Pop B	Pop C
OR	0.182	0.125	*0.182*	*0.124*	*0.165*	0.932	0.913	*0.932*	*0.908*	*0.927*
AB	0.261	0.166	*0.261*	*0.290*	*0.173*	0.016	0.014	*0.016*	*0.022*	*0.013*
RR	0.255	0.150	*0.255*	*0.174*	*0.201*	0.955	0.929	*0.955*	*0.938*	*0.942*
NNT	4	6	*4*	*3*	*6*	63	72	*63*	*46*	*79*

References

[1] Neumann, P.J. 2013 Communicating and promoting comparative-effectiveness research findings. *N Engl J Med* **369**, 209-211. (doi:10.1056/NEJMp1300312).

[2] Eichler, H.G., Abadie, E., Breckenridge, A., Flamion, B., Gustafsson, L.L., Leufkens, H., Rowland, M., Schneider, C.K. & Bloechl-Daum, B. 2011 Bridging the efficacy-effectiveness gap: a regulator's perspective on addressing variability of drug response. *Nat Rev Drug Discov* **10**, 495-506. (doi:10.1038/nrd3501).

[3] European Medicines Agency. Report of the CHMP working group on benefit-risk assessment models and methods. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2010/01/WC500069668.pdf. Accessed on: 30 July 2016.

[4] Boissel, J.P., Cucherat, M., Nony, P., Chabaud, S., Gueyffier, F., Wright, J.M., Lievre, M. & Leizorovicz, A. 2008 New insights on the relation between untreated and treated outcomes for a given therapy effect model is not necessarily linear. *J Clin Epidemiol* **61**, 301-307. (doi:10.1016/j.jclinepi.2007.07.007).
[5] Boissel, J.P., Collet, J.P., Lievre, M. & Girard, P. 1993 An effect model for the assessment of drug benefit: example of antiarrhythmic drugs in postmyocardial infarction patients. *J. Cardiovasc. Pharmacol.* **22**, 356-363.
[6] Wang, H., Boissel, J.P. & Nony, P. 2009 Revisiting the relationship between baseline risk and risk under treatment. *Emerg Themes Epidemiol* **6**, 1. (doi:10.1186/1742-7622-6-1).
[7] Boutitie, F., Gueyffier, F., Pocock, S.J. & Boissel, J.P. 1998 Assessing treatment-time interaction in clinical trials with time to event data: a meta-analysis of hypertension trials. *Stat Med* **17**, 2883-2903.
[8] Hildebrandt, M., Vervolgyi, E. & Bender, R. 2009 Calculation of NNTs in RCTs with time-to-event outcomes: a literature review. *BMC Med Res Methodol* **9**, 21. (doi:10.1186/1471-2288-9-21).
[9] Bagley, S.C., White, H. & Golomb, B.A. 2001 Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. *J Clin Epidemiol* **54**, 979-985.
[10] Gabrielsson, J. & Weiner, D. 2000 Pharmacokinetic and pharmacodynamic data analysis: Concepts and applications. (pp. 177-189, 3rd edition ed. Stockholm, Sweden, Swedish Pharmaceutical Press.