



# From Clinical Trial Efficacy to Real-Life Effectiveness: Why Conventional Metrics do not Work

Jean-Pierre Boissel<sup>1</sup> · Frédéric Cogny<sup>1</sup> · Nicholas Marko<sup>2</sup> · François-Henri Boissel<sup>1</sup>

Published online: 29 July 2019  
© The Author(s) 2019

## Abstract

**Background** Randomised, double-blind, clinical trial methodology minimises bias in the measurement of treatment efficacy. However, most phase III trials in non-orphan diseases do not include individuals from the population to whom efficacy findings will be applied in the real world. Thus, a translation process must be used to infer effectiveness for these populations. Current conventional translation processes are not formalised and do not have a clear theoretical or practical base. There is a growing need for accurate translation, both for public health considerations and for supporting the shift towards personalised medicine.

**Objective** Our objective was to assess the results of translation of efficacy data to population efficacy from two simulated clinical trials for two drugs in three populations, using conventional methods.

**Methods** We simulated three populations, two drugs with different efficacies and two trials with different sampling protocols.

**Results** With few exceptions, current translation methods do not result in accurate population effectiveness predictions. The reason for this failure is the non-linearity of the translation method. One of the consequences of this inaccuracy is that pharmacoeconomic and postmarketing surveillance studies based on direct use of clinical trial efficacy metrics are flawed.

**Conclusion** There is a clear need to develop and validate functional and relevant translation approaches for the translation of clinical trial efficacy to the real-world setting.

## Key Points

The efficacy of treatments can be assessed in randomised, double-blind, clinical trials to minimise bias.

Most of these clinical trials do not include individuals from the population to whom efficacy findings will be applied in the real world, so the effectiveness of treatments must be ‘translated’ to these populations.

We show that current translation methods do not provide accurate predictions for effectiveness, highlighting the need to develop and validate functional and relevant translation approaches for the translation of clinical trial efficacy to the real-world setting.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s40801-019-0159-z>) contains supplementary material, which is available to authorized users.

✉ François-Henri Boissel  
[francois.boissel@novadiscovery.com](mailto:francois.boissel@novadiscovery.com)

<sup>1</sup> Novadiscovery, Lyon, France

<sup>2</sup> Department of Neurosurgery, MD Anderson Cancer Center, Houston, TX, USA

## 1 Introduction

Randomised, double-blind clinical trial methodology, if well-implemented, minimises bias in the measurement of treatment efficacy and allows any difference in outcomes to be attributed to the treatment effect. This provides an unbiased estimate of the size of the difference in outcome rates between patients in the treated and control groups. Clinical trials are designed to answer the question, ‘is the tested treatment better than the control?’ and to establish a causal link between receiving the tested treatment and the difference in outcome rates. The estimate of the size of this difference provides a quantitative estimate of how much better the treatment is for a group of patients or, even, a given patient [1, 2].

The patients in phase III clinical trials are included from the treatment target population using eligibility criteria, which often lead to a ‘selected’ trial population that is not always representative of the target population [3–10]. For example, various racial/ethnic populations, elderly people and women were shown to be underrepresented in 59 trials in heart failure [11]. In breast, colorectal, lung and prostate cancer clinical trials sponsored by the National Cancer Institute, participation varied significantly across racial/

ethnic and age groups, and in cardiovascular clinical trials funded by the National Heart, Lung, and Blood Institute, women were reported to be underrepresented [12, 13].

Attempts have been made to correct for this via specific trial designs, appropriate data analysis tools or using a pragmatic trial approach with more permissive eligibility criteria, but success has been limited [8, 14, 15]. There is heterogeneity between results from large, multicentre international trials assessing the same treatment, suggesting that the trial populations differ. For example, in a systematic review of remifentanyl compared with short-acting opioids for general anaesthesia, the observed overall frequency of postoperative nausea in 11 fentanyl control groups ( $N=3048$ ) ranged from 14 to 81% [16]. In the two largest trials ( $N=2437$ ;  $N=4787$ ), the frequencies were statistically significantly different (25% and 32%,  $p=0.0002$ ), implying heterogeneity in the patients' characteristics or the practice of care. Another example is the reported heterogeneity of the absolute benefit (AB) estimates from clinical trials assessing the same drug class [5, 17]. The results from 12 trials assessing the efficacy of  $\beta$ -blockers versus placebo or no  $\beta$ -blocker in reducing 1-year mortality rate in post-myocardial infarction patients were published between 1975 and 1990 (Table 1) [18–31]. The AB ranged from 0.0155 (an increase in mortality) to  $-0.0530$  (a reduction), and the corresponding number needed to treat (NNT) ranged from  $-421$  to 60. If we consider only the three trials with a  $p$  value  $<0.05$ , the range for the AB is  $-0.0167$  to  $-0.0530$  and 19–60 for the NNT. A meta-analysis of these trials showed heterogeneity for AB but not for relative risk (RR) [17].

This heterogeneity makes it difficult to generalise these trial results to the whole population. Thus, a translation

process must be used to extrapolate the efficacy for these populations. The goal of the translation process, which is sometimes termed the 'transportability' process, is to predict the impact of the tested treatment on the population of interest in a real-world setting, using the clinical trial results [32]. This translation process is integrated in a broader framework known as health technology assessment, which assesses the impact, safety and cost of a treatment on the health status of the target population.

Generally, the endpoints in phase III clinical trials reflect clinical outcomes that are binary variables, such as death or occurrence of a cancer relapse. Thus, the efficacy estimate is calculated using the rate of outcomes observed in the control group ( $R_c$ ) and in the experimental (treated) group ( $R_t$ ). These are analysed using summary metrics (or statistics) of treatment efficacy, such as the odds ratio (OR), RR, relative benefit, AB and NNT. See the Electronic Supplementary Material (ESM) for more information.

The purpose of this article was to compare estimated population-level benefit, based on summary clinical trial data, as is usually done, with that based on the true efficacy in the population of interest, translated from the efficacy observed in clinical trials.

The process of translating clinical trial findings to a given population involves using the trial efficacy metrics to compute population benefit metrics. In this article, we have limited our assessment to  $NPE_{pop}$  and  $NNT_{pop}$ , which we think are the most relevant population benefit metrics, as shown in the ESM. We assessed whether these population metrics derived from the clinical trial efficacy metrics could accurately predict real-world effectiveness over a given time.

**Table 1** Results from trials assessing  $\beta$ -blockers for the prevention of death (1-year mortality) in post-myocardial infarction patients (data standardised at 1 year of follow-up)

Beta-blocker	Treated group		Control group		$p$	Efficacy metrics			
	Event rate <sup>a</sup>	$R_t$	Event rate <sup>a</sup>	$R_c$		RR	OR	AB	NNT
Acebutolol [18]	17/298	0.06	34/309	0.11	<b>0.027</b>	0.52	0.49	$-0.0530$	19
Alprenolol [19]	5/114	0.04	8/116	0.07	0.41	0.64	0.62	$-0.0251$	40
Metoprolol [20]	65/1195	0.05	62/1200	0.05	0.76	1.05	1.06	0.0027	$-367$
Metoprolol [21]	25/154	0.16	31/147	0.21	0.28	0.77	0.73	$-0.0485$	21
Oxprenolol [22]	57/858	0.07	45/883	0.05	0.17	1.30	1.33	0.0155	$-65$
Pindolol [23]	36/263	0.14	33/266	0.12	0.66	1.10	1.12	0.0128	$-78$
Practolol [24]	45/1533	0.03	70/1521	0.05	<b>0.01</b>	0.64	0.63	$-0.0167$	60
Propranolol [25, 26]	70/1916	0.04	115/1921	0.06	<b>0.001</b>	0.61	0.60	$-0.0233$	43
Propranolol [27]	20/193	0.10	19/195	0.10	1.07	1.06	1.07	0.0062	$-162$
Propranolol [28]	25/278	0.09	37/282	0.13	0.12	0.69	0.65	$-0.0413$	24
Sotalol [29]	44/873	0.05	28/583	0.05	0.84	1.05	1.05	0.0024	$-421$
Timolol [30, 31]	72/945	0.08	106/939	0.11	<b>0.006</b>	0.67	0.65	$-0.0367$	27

Bold numbers indicate statistically significant  $p$  values. Italic numbers indicate negative NNT

AB absolute benefit, NNT number needed to treat, OR odds ratio,  $R_c$  risk in control group, RR relative risk,  $R_t$  risk in treated group

<sup>a</sup>Event rates are presented as population size/number of events

## 2 Materials and Methods

We used fictive individual patient data to simulate the translation process for clinical trial efficacy to real-life effectiveness because generally only aggregated data are publically available. Aggregated data provide estimates for the ‘average’ patient enrolled in the clinical trial, who is probably not representative of patients in the real world since higher-risk patients are rarely enrolled in phase III clinical trials. Hence, the results from these analyses should be assessed qualitatively and not quantitatively.

### 2.1 Simulation Framework

We simulated three populations, two drugs with different efficacies, and two trials with different sampling protocols.

#### 2.1.1 Populations

Each population, A, B and C, comprised 100,000 individuals who were all assumed to have the same disease, thus they were all at risk of the same clinical event, but the event rates in untreated individuals ( $R_c$ ) differed in each population (Fig. 1). The distributions of  $R_c$  differed, but the average  $R_c$  was the same for populations A and B (0.35) and was lower for population C (0.22). The effect of the two drugs on the clinical outcome was modelled with the Wang model (see Sect. 2.2). The population metrics for the beneficial effects of drugs 1 and 2 were then computed for the three populations.

#### 2.1.2 Drugs

In the simulation, both drugs 1 and 2 had the same mode of action but drug 1 was more potent than drug 2 (Table 2) [33].

#### 2.1.3 Clinical Trials

Two clinical trials were simulated in population A, one for each drug, to obtain two sets of summary trial metrics using different sampling processes. Trial 1 should have been run on a random sample of population A; however, since random variations and confidence intervals were not taken into consideration in our approach, the whole population A was used in trial 1, not a random sample. Hence, this can be considered as a random sample with the same average  $R_c$  as the overall population, without the random variations. A non-random sample from population A with an average  $R_c$  that was lower than that for the overall population was used in trial 2.

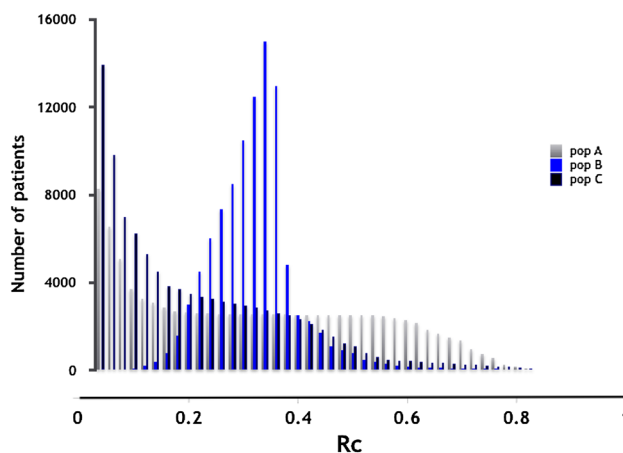


Fig. 1 Distribution of risk without treatment ( $R_c$ ) in three simulated populations, A, B and C, each comprising 100,000 individuals who were all assumed to have the same disease and, therefore, were all at risk of a clinical event but the event rates in the untreated individuals ( $R_c$ ) differed in each population

Table 2 Summary of characteristics of drugs 1 and 2 [33]

Parameter	Drug 1	Drug 2
Dose (mg)	3.5	5
Gamma	3	2.7
ED <sub>50</sub> (mg)	30	29
Stimulus	2	3
$E_{max}$ (mg)	35	35

Gamma and stimulus are two parameters of the Hill equation. For the purposes of this demonstration, we assigned the units for the drug as milligrams

$ED_{50}$  the amount of a drug that is therapeutic in 50% of the individuals or animals in which it is tested,  $E_{max}$  the amount of a drug that produces the maximum therapeutic effect

### 2.2 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 [34]. 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 ( $\beta_0$ , the intercept, and  $S$ , the coefficient of  $E$ ), which can be interpreted as the scale of the drug effect size [35]. See the ESM for more details.

### 2.3 Calculations

To assess any translation biases arising from the source of data used for the efficacy metrics calculation, we compared

the efficacy metrics of each of the two drugs (1) computed on the trial summary data (for the two trials with each of the two drugs), (2) computed on the three populations (for each of the two drugs) and (3) translated for the three populations from the trial summary data (for the two trials and the two drugs). More details about this process are provided in the ESM.

### 3 Results

#### 3.1 Simulated Clinical Trials

The results from the four simulated clinical trials, which were assumed to be statistically significant, showed that the efficacy metrics differed in the two populations for the same drug (Table 3). The efficacy metrics for the least potent drug, i.e. drug 2, were less favourable in both trials.

#### 3.2 Results from Simulated Translation

The results from the simulated translation of results from clinical trials with drugs 1 and 2 are summarised in Table 3. Although the  $NPE_{pop}$  should be constant for a given drug in a given population when translated using trial summary data, its value varied depending on the metric used to calculate it. For example, the  $NPE_{pop}$  for drug 2 in population C, calculated from trial summary data, was 125% of the true value when calculated with AB from trial 1 and 356% and 535% of the true value when calculated with the RR and OR, respectively, from the same trial. For  $NNT_{pop}$ , the ratio varied from one population to another within the single trial, which could have been anticipated. The estimates of real-world effectiveness metrics using the clinical trial efficacy metrics differed

from the values calculated for the trial populations, with the exception of population A in trial 1 with both drugs, since the whole population was included in this trial. The number of prevented events (NPEs) and NNTs were under-predicted for drug 1 and over-predicted for drug 2 when the RRs and ORs from the clinical trials were used for the translation (Table 4). RR varies with  $R_c$ , and the RRs varied between trials and with the population (Fig. 2; Table 4). This variation was greater for drug 2, which was less efficacious than drug 1.

Table 5 summarises the values of estimated effectiveness metrics in populations A, B and C based on the trial efficacy metrics estimated from trial 1 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. For example, the observed RRs in trial 1 with drugs 1 and 2 were 0.255 and 0.955, respectively. When these were used to translate to the three populations individually, we observed the same values for population A because the whole population was included in the trials, but the RRs for populations B and C were, respectively, 0.174 and 0.201 for drug 1 and 0.938 and 0.942 for drug 2. The bias was lowest when AB, computed with trial summary data, was used for the translation. The RR computed using trial 2 summary data differed from the RR for the true population A since trial 2 was run on a selected sample of population A (see the ESM).

### 4 Discussion

The observed differences between trial efficacy metrics and real-world effectiveness metrics is due to differences in  $R_c$  distributions in trial and real-world populations. We demonstrated that it is possible to translate an appropriate trial

**Table 3** Summary of efficacy metrics from two clinical trials for drugs 1 and 2

	Drug 1		Drug 2	
	Trial 1 ( $N=100,000$ )	Trial 2 ( $N=52,963$ )	Trial 1 ( $N=100,000$ )	Trial 2 ( $N=52,963$ )
Control group: $R_c$	0.351	0.195	0.351	0.195
Treated group: $R_t$	0.089	0.029	0.335	0.181
Efficacy metrics				
OR	0.358	0.181	0.978	0.945
RR	0.255	0.150	0.955	0.929
AB	0.261	0.166	0.016	0.014
NNT	4	6	63	72

In trial 1, the whole population was included as it was not possible to include a randomised sample because the approach does not take into consideration random variations and, thus, confidence intervals. In trial 2, a non-randomised sample was selected, with a lower average risk in the untreated patients than that for the whole population

AB absolute benefit,  $N$  trial population size, NNT number needed to treat, OR odds ratio,  $R_c$  risk in control group, RR relative risk,  $R_t$  risk in treated group

efficacy metric to a population effectiveness metric if the trial is undertaken on a random sample of the population of interest. However, in most diseases, except rare (orphan) diseases, it is extremely difficult, if not impossible, to recruit

patients into clinical trials who are truly representative of the population that will be treated. Although the clinical trial population is drawn from the treatment target population, it is selected using eligibility criteria that result in a

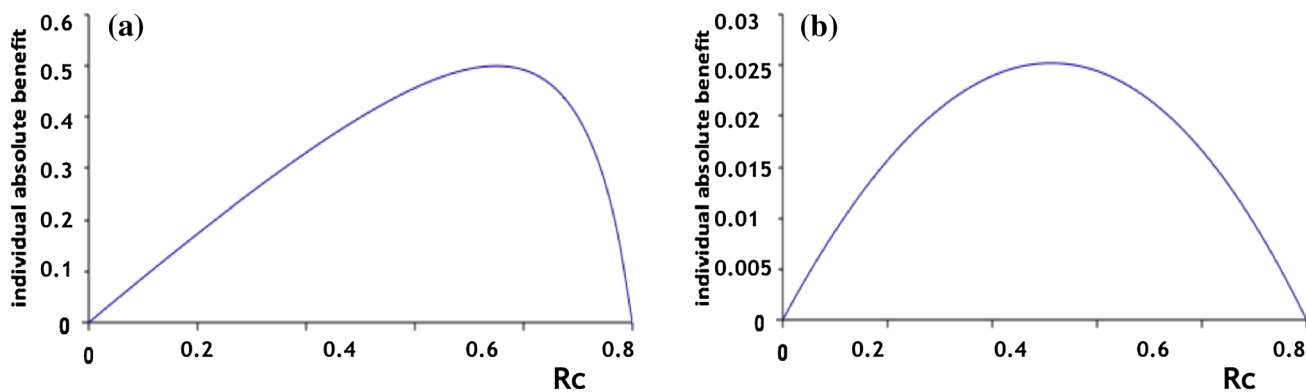
**Table 4** Number of prevented events and number needed to treat<sup>a</sup>, translated in populations A, B and C, using various trial efficacy metrics estimated using summary data from trials 1 and 2, for drugs 1 and 2

Efficacy metric	Drug 1						Drug 2					
	Trial 1 in population			Trial 2 in population			Trial 1 in population			Trial 2 in population		
	A	B	C	A	B	C	A	B	C	A	B	C
<b>NPE<sub>pop</sub></b>												
OR	313	282	105	335	302	507	429	313	535	548	400	684
RR	285	257	148	325	293	492	285	208	356	448	327	560
AB	100	90	151	63	57	96	100	73	125	87	64	109
<b>NNT<sub>pop</sub></b>												
RR	100	111	107	88	97	94	100	137	130	64	87	83

Example of interpretation: the true NPE for population 2 and drug 2, directly computed on the  $R_c$  distribution in population 2 with the Wang model (see Sect. 2) is 1265. The NPE in this population with drug 2 computed with the AB from trial 2 summary data is 1382. The ratio 1382/1265 in percent is 109%. It gives the value of the bias in estimating population C NPE with the AB computed on trial data. As it is often done in translation process, RR computed from trial summary data is used for translating the efficacy of new drug. In the considered case, this gives a translated NPE = 7084, thus a bias ratio = 560%

AB absolute benefit, N population size, NNT number needed to treat,  $NNT_{pop}$  number needed to treat in the population,  $NPE_{pop}$  number of prevented events in the population, OR odds ratio, RR relative risk

<sup>a</sup>The NPE and NNT are expressed as a percentage of the true values



**Fig. 2** Variation of absolute benefit with risk without treatment for two drugs (1 and 2) in the same population A. **a** The absolute benefit (AB) as a function of the risk without treatment ( $R_c$ ) in population A for drug 1; **b** AB as a function of  $R_c$  in population A for drug 2

**Table 5** Comparison of the values for odds ratio, absolute benefit, relative risk and number needed to treat effectiveness metrics calculated for populations A, B and C and drugs 1 and using efficacy metrics from trial 1 for each drug

	Drug 1				Drug 2			
	Trial 1	Population			Trial 1	Population		
		A	B	C		A	B	C
OR	0.182	0.182	0.124	0.165	0.932	0.932	0.908	0.927
AB	0.261	0.261	0.290	0.173	0.016	0.016	0.022	0.013
RR	0.255	0.255	0.174	0.201	0.955	0.955	0.938	0.942
NNT	4	4	3	6	63	63	46	79

AB absolute benefit, NNT number needed to treat, OR odds ratio, RR relative risk

subpopulation of the treatment target population that does not have the same characteristics as the whole population; in particular, the  $R_c$  differs.

The translation issue has been explored both through a factual approach by comparing clinical trial results with observational data and through a theoretical statistical approach [1, 14, 36]. The factual approach produced results that were difficult to interpret because of the variability of postmarketing (phase IV) studies and their limited capability to manage bias. The theoretical statistical approach was not intuitive for the medical community and will require more work to provide a practical solution.

As mentioned, simulation is a simple way to explore translation, although it does not resolve the issue, particularly when the models and simulated data have not been validated. However, since there is no alternative approach for exploring this issue, our results should be interpreted cautiously.

The NNT metric is generally used by teachers, regulators, authors and pharmaceutical companies to benchmark treatments or assess the relative benefits of a treatment. The treatment with the lowest NNT is generally taken to be the most efficacious. It has been suggested that NNT ‘has that clinical immediacy’ (of clinical applicability), which is one reason why it is such a popular measure [37]. However, this is not true when the NNT is computed on clinical trial data for translation purposes or for comparing drug efficacies, as frequently occurs. We showed that the same drug in different trials can lead to different NNT values when the  $R_c$  for the trial populations differ and are different from those for the population of interest; therefore, the translated NNT should be interpreted cautiously. Several authors have warned against the sensitivity of NNT to factors that change baseline risk, e.g. patients’ characteristics, secular trends in incidence and case fatality and delay to event [38, 39]. The value of NNT is not the same if the treatment effect is immediate or if the effect is to delay an outcome rather than prevent it [40].

As mentioned, evidence exists that patients included in clinical trials, although taken from the overall target population, are not representative of all patients to whom the new treatment will be prescribed. The main differences between the trial population and the real-world population are the risk of the outcome ( $R_c$ ) and the presence of concomitant diseases [5]. Although it is often assumed that the populations will be sufficiently similar to support the hypothesis that the new treatment will also be efficacious in the real-world population, we cannot extrapolate the size of the treatment effect from the clinical trial to the real-world population. Even when recruitment criteria focus on high-risk patients, it has been observed that trial patients are at lower risk than real-world ‘high-risk’ patients and the exclusion criteria often prevent patients with concomitant diseases and those who cannot respect the clinical trial schedules from being

recruited to the trial, although these patients will potentially receive the new drug [18]. The results from trials assessing  $\beta$ -blockers illustrate the fact that the same treatment can show differing efficacy when used in different patient populations who, in theory, all have the same disease or risk (Table 1).

With a low-risk population, there would be differences, but if the treatment were moderately efficacious, these differences would be modest. Since, for most drugs, treatment efficacy assessed in trials is modest, the population-level effectiveness metrics obtained by translating trial efficacy metrics values may be viewed as satisfactory. However, this would be without taking into account the issue of responders/non-responders, which is more important when the observed efficacy is low.

Population efficacy metrics computed on clinical trial data were first used to assess the validity of the trial results. However, the statistical assessment with traditional null hypothesis testing is based on the assumption that the analysed trial is a random sample of an infinite number of similar trials and therefore the observed trial efficacy is representative of the true treatment efficacy.

Regulators and payers continue to focus on statistical significance and  $p$  values and have not adequately addressed the issue of translation of treatment efficacy from a trial setting to treatment effectiveness in a real-life setting. The guidelines published in 2007 by the European Medicines Agency explicitly mentioned the translation issue in the objectives section, but the issue was not properly formulated and addressed in the rest of the report, although the NNT was discussed in a way that came close to the fundamental issue [41]. This fundamental issue concerns the fact that we are dealing with non-linear effects, whereas the metrics used in translation assume a linear effect.

Pragmatic clinical trials were designed to address the translation issue by providing evidence for adoption of treatment in real-world clinical practice [42, 43]. Since then, only a few truly pragmatic trials have been published, essentially because the rules that define a pragmatic trial are difficult to put into practice. For example, the patients should be similar to those who will receive the intervention in real life, but they must accept being randomised to the new treatment or the comparator, which is usually the current standard of care. In addition, the investigators, who should be real-world prescribers and not trialists, can decide how to administer the treatment. Alternatively, model-based methods can be used to translate observed trial results to a specified target population, but this approach can only take into consideration a small number of covariates [14].

We propose that the effect model (EM) could be used to translate trial metrics to population metrics. The EM approach models the relationship between AB and  $R_c$ , which is a characteristic of the treatment at a given time point. This

has been demonstrated using simulated populations and been reported in real life [44, 45].

## 5 Conclusion

This analysis clearly shows that more appropriate and accurate tools are needed to be able to translate clinical trial efficacy to population-level effectiveness. We showed that two population efficacy metrics,  $NPE_{pop}$  and  $NNT_{pop}$ , could be used to compare two or more treatments (e.g. drugs 1 and 2 in populations A, B or C), irrespective of whether the trials had been run on random samples of the corresponding populations or whether unbiased translation has been achieved. This approach requires prior knowledge of, at least, the target population distribution of  $R_c$  and the treatment EM.

**Acknowledgements** The authors thank Dr. Robert Dahan for his helpful comments on the content of this manuscript and Melanie Senior for her critical appraisal of our work. They would also like to thank Dr. Margaret Haugh (MediCom Consult) for editorial assistance, funded by Novadiscovery.

## Compliance with Ethical Standards

**Funding** Novadiscovery funded medical writing and editorial services for the preparation of this manuscript.

**Conflict of interest** Jean-Pierre Boissel, Frédéric Cogny and François-Henri Boissel are employees and shareholders of Novadiscovery, which currently has a patent pending on the Effect Model Business Applications. Nicholas Marko has no conflicts of interest that are directly relevant to the content of this article.

**Ethical approval** The research described in this article did not involve the participation of individuals and therefore did not require ethical approval.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Pearl J. Generalizing experimental findings. *J Causal Inference*. 2015;3:259–66.
- Shadish WR, Cook TD, Campbell DT. *Experimental and quasi-experimental designs for generalized causal inference*. Belmont: Wadsworth Cengage Learning; 2002.
- Huls H, Abdulahad S, Mackus M, Van de Loo JA, Roehrs T, Roth T, Verster CJ. Inclusion and Exclusion Criteria of Clinical Trials for Insomnia. *J Clin Med*. 2018. <https://doi.org/10.3390/jcm7080206>.
- Boissel JP, Leizorovicz A, Picolet H, Ducruet T. Efficacy of acebutolol after acute myocardial infarction (the APSI trial). The APSI Investigators. *Am J Cardiol*. 1990;66:24c–31c.
- Boissel JP, Peyrieux JC. Sub-grouping of post myocardial infarction patients according to their one-year death risk. *Eur Heart J*. 1987;8:1272–80.
- Cebul RD, Snow RJ, Pine R, Hertzner NR, Norris DG. Indications, outcomes, and provider volumes for carotid endarterectomy. *JAMA*. 1998;279:1282–7.
- Wennberg DE, Lucas FL, Birkmeyer JD, Bredenberg CE, Fisher ES. Variation in carotid endarterectomy mortality in the Medicare population: trial hospitals, volume, and patient characteristics. *JAMA*. 1998;279:1278–81.
- Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. *JAMA*. 2007;298:1209–12. <https://doi.org/10.1001/jama.298.10.1209>.
- Cars O, Molstad S, Melander A. Variation in antibiotic use in the European Union. *Lancet*. 2001;357:1851–3. [https://doi.org/10.1016/s0140-6736\(00\)04972-2](https://doi.org/10.1016/s0140-6736(00)04972-2).
- Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet*. 1993;342:1317–22.
- Heiat A, Gross CP, Krumholz HM. Representation of the elderly, women, and minorities in heart failure clinical trials. *Arch Intern Med*. 2002;162:1682–8.
- Murthy VH, Krumholz HM, Gross CP. Participation in cancer clinical trials: race-, sex-, and age-based disparities. *JAMA*. 2004;291:2720–6. <https://doi.org/10.1001/jama.291.22.2720>.
- Harris DJ, Douglas PS. Enrollment of women in cardiovascular clinical trials funded by the National Heart, Lung, and Blood Institute. *N Engl J Med*. 2000;343:475–80. <https://doi.org/10.1056/nejm200008173430706>.
- Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: the ACTG 320 trial. *Am J Epidemiol*. 2010;172:107–15. <https://doi.org/10.1093/aje/kwq084>.
- ISIS-2 Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. *Lancet*. 1988;2:349–60.
- Komatsu R, Turan AM, Orhan-Sungur M, McGuire J, Radke OC, Apfel CC. Remifentanyl for general anaesthesia: a systematic review. *Anaesthesia*. 2007;62:1266–80. <https://doi.org/10.1111/j.1365-2044.2007.05221.x>.
- The Beta-Blocker Pooling Project Research Group. The Beta-Blocker Pooling Project (BBPP): subgroup findings from randomized trials in post infarction patients. *Eur Heart J*. 1988;9:8–16.
- Boissel JP, Leizorovicz A, Picolet H, Peyrieux JC. Secondary prevention after high-risk acute myocardial infarction with low-dose acebutolol. *Am J Cardiol*. 1990;66:251–60.
- Vedin A, Wilhelmsson C, Werko L. Chronic alprenolol treatment of patients with acute myocardial infarction after discharge from hospital. *Acta Med Scand Suppl*. 1975;575:3–56.
- Lopressor Intervention Trial Research Group. The Lopressor Intervention Trial: multicentre study of metoprolol in survivors of acute myocardial infarction. *Eur Heart J*. 1987;8:1056–64.
- Olsson G, Rehnqvist N, Sjogren A, Erhardt L, Lundman T. Long-term treatment with metoprolol after myocardial infarction: effect on 3 year mortality and morbidity. *J Am Coll Cardiol*. 1985;5:1428–37.
- European Infarction Study Group. European Infarction Study (E.I.S.). A secondary prevention study with slow release

- oxprenolol after myocardial infarction: morbidity and mortality. *Eur Heart J*. 1984;5:189–202.
23. Australian and Swedish Pindolol Study Group. The effect of pindolol on the two years mortality after complicated myocardial infarction. *Eur Heart J*. 1983;4:367–75.
  24. Multicentre International Study Group. Reduction in mortality after myocardial infarction with long-term beta-adrenoceptor blockade. Multicentre international study: supplementary report. *Br Med J*. 1977;2:419–21.
  25. Beta-Blocker Heart Attack Trial Research Group. A randomized trial of propranolol in patients with acute myocardial infarction. I. Mortality results. *JAMA*. 1982;247:1707–14.
  26. Hawkins CM, Richardson DW, Vokonas PS. Effect of propranolol in reducing mortality in older myocardial infarction patients. The Beta-Blocker Heart Attack Trial experience. *Circulation*. 1983;67:194–7.
  27. Baber NS, Evans DW, Howitt G, Thomas M, Wilson T, Lewis JA, Dawes PM, Handler K, Tuson R. Multicentre post-infarction trial of propranolol in 49 hospitals in the United Kingdom, Italy, and Yugoslavia. *Br Heart J*. 1980;44:96–100.
  28. Hansteen V, Moinichen E, Lorentsen E, Andersen A, Strom O, Soiland K, Dyrbekk D, Refsum AM, Tromsdal A, Knudsen K, et al. One year's treatment with propranolol after myocardial infarction: preliminary report of Norwegian multicentre trial. *Br Med J*. 1982;284:155–60.
  29. Julian DG, Prescott RJ, Jackson FS, Szekely P. Controlled trial of sotalol for one year after myocardial infarction. *Lancet*. 1982;1:1142–7.
  30. The Norwegian Multicentre Study Group. Timolol-induced reduction in mortality and reinfarction in patients surviving acute myocardial infarction. *N Engl J Med*. 1981;304:801–7. <https://doi.org/10.1056/nejm198104023041401>.
  31. Gundersen T. Influence of heart size on mortality and reinfarction in patients treated with timolol after myocardial infarction. *Br Heart J*. 1983;50:135–9.
  32. Pearl J, Bareinboim E. External validity: from do-calculus to transportability across populations. *Stat Sci*. 2014;29:579–95.
  33. Gabrielsson J, Weiner D. Pharmacokinetic and pharmacodynamic data analysis: concepts and applications. 3rd ed. Stockholm: Swedish Pharmaceutical Press; 2000. p. 177–89.
  34. Wang H, Boissel JP, Nony P. Revisiting the relationship between baseline risk and risk under treatment. *Emerg Themes Epidemiol*. 2009;6:1. <https://doi.org/10.1186/1742-7622-6-1>.
  35. Bagley SC, White H, Golomb BA. Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. *J Clin Epidemiol*. 2001;54:979–85.
  36. Hannan EL. Randomized clinical trials and observational studies: guidelines for assessing respective strengths and limitations. *JACC Cardiovasc Interv*. 2008;1:211–7. <https://doi.org/10.1016/j.jcin.2008.01.008>.
  37. McQuay HJ, Moore RA. Using numerical results from systematic reviews in clinical practice. *Ann Intern Med*. 1997;126:712–20.
  38. Kristiansen IS, Gyrd-Hansen D. Cost-effectiveness analysis based on the number-needed-to-treat: common sense or non-sense? *Health Econ*. 2004;13:9–19. <https://doi.org/10.1002/hec.797>.
  39. Smeeth L, Haines A, Ebrahim S. Numbers needed to treat derived from meta-analyses—sometimes informative, usually misleading. *BMJ*. 1999;318:1548–51.
  40. Aaron SD, Fergusson DA. Exaggeration of treatment benefits using the “event-based” number needed to treat. *CMAJ*. 2008;179:669–71. <https://doi.org/10.1503/cmaj.080018>.
  41. European Medicines Agency. Report of the CHMP working group on benefit-risk assessment models and methods. [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Regulatory\\_and\\_procedural\\_guideline/2010/01/WC500069668.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2010/01/WC500069668.pdf). Accessed on 30 July 2016.
  42. Ford I, Norrie J. Pragmatic trials. *N Engl J Med*. 2016;375:454–63. <https://doi.org/10.1056/NEJMr1510059>.
  43. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Chronic Dis*. 1967;20:637–48.
  44. Boissel J-P, Kahoul R, Marin D, Boissel F-H. Effect model law: an approach for the implementation of personalized medicine. *J Pers Med*. 2013;3:177.
  45. Kahoul R, Gueyffier F, Amsallem E, Haugh M, Marchant I, Boissel FH, Boissel JP. Comparison of an effect-model-law-based method versus traditional clinical practice guidelines for optimal treatment decision-making: application to statin treatment in the French population. *J R Soc Interface*. 2014;11:20140867. <https://doi.org/10.1098/rsif.2014.0867>.