ORIGINAL RESEARCH

Analyzing noninferiority trials: it is time for advantage deficit assessment – an observational study of published noninferiority trials

Beryl Primrose Gladstone
Werner Vach

Clinical Epidemiology Group,
Center for Medical Biometry and
Medical Informatics, Medical Center,
University of Freiburg, Freiburg,
Germany

Abstract: The concept of noninferiority (NI) trials is based on a belief that the new therapy may potentially offer a benefit for the patient or society in spite of it having a slightly lower efficacy. We introduce advantage deficit assessment (ADA), a simple framework similar to the benefit-risk assessment in superiority trials. ADA balances the advantage gained against the deficit in efficacy on a two-dimensional plane. It requires that NI trials provide quantitative information on both the advantage as well as on efficacy on scales, which can be compared in a meaningful manner. From this perspective, we study the feasibility of ADA among a set of NI trials published in four major medical journals. Among 113 published NI trials, about half assessed and reported at least one claimed advantage. For most other studies, an assessment seems to be feasible if considered in the planning of the study. Many studies claiming noninferiority report a positive gain in advantage. These trials have the potential to demonstrate a significant net benefit in an ADA, substantially changing the final judgment of the study results. ADA seems promising as it overcomes the current limitation of NI trials to demonstrate "only" noninferiority and brings them back to the mainstream of superiority trials by aiming to demonstrate a positive net benefit. ADA seems to be feasible in the majority of NI trials.

Keywords: noninferiority, new treatment, potential advantage, advantage deficit assessment, benefit-risk assessment, loss of efficacy

Introduction

The concept of noninferiority is based on allowing a trade-off between loss of efficacy and gain in advantage. The advantage may be in the form of safety, ease of administration, tolerability, or costs.[1–3] A small loss of efficacy is regarded as acceptable with the belief that the new therapy may offer an advantage for the patient or society. Traditionally, the advantage of the tested treatment has been one of the factors used to guide and justify the choice of the study design as well as the noninferiority margin. Regulatory guidelines from the US Food and Drug Administration and European Medicines Agency also require the advantage of the new treatment to justify the choice of a noninferiority (NI) trial design.[4,5] In its recommendation for the choice of noninferiority margin, the European Medicines Agency states that a lenient choice of margin could be justified by the presence of advantage of the new treatment.[4] However, evaluation of the advantage is not required.

Earlier on, Garattini et al pointed out the need to demonstrate the potential advantages of the tested treatment in addition to its noninferiority in terms of efficacy.[6] Similarly, in their 2003 article on experiences with design issues in NI trials, D'Agastino et al[1] mentioned assessment of a composite objective of noninferior efficacy and

Correspondence: Beryl Primrose Gladstone
Clinical Epidemiology Group, Center for Medical Biometry and Medical Informatics, Medical Center, University of Freiburg, Stefan-Meier-Strasse 26, D-79104 Freiburg, Germany
Tel +49 76 1203 6697
Fax +49 76 1203 6711
Email beryl@imbi.uni-freiburg.de

superior safety as a possibility. Recently, several scientists have recommended similar strategies.[7–10]

Evaluation of safety along with efficacy in superiority trials has been around for a long time, and solid methodologies have been developed for simultaneous comparisons and multiple endpoint testing.[11–18] Adapting these to NI trials, several approaches for multiple comparison procedures allowing simultaneous testing of noninferiority and superiority of the multiple endpoints in NI trials are available. These include a mixture of graphical approaches, alpha allocation, and hierarchical gatekeeping, eg, Bretz et al[12,13] Bauer et al,[11] Bristol,[18] Nishikawa et al[17] and Guilbaud.[16]

Balancing efficacy against safety using hypothesis testing, however, neglects the quantitative nature of this concept. Hence, quantitative benefit-risk assessment has been highlighted by the regulatory authorities for more than a decade now[19–23] in the context of superiority trials. The US Food and Drug Administration and European Medicines Agency are improving their guidelines on benefit-risk assessment, with recent attention on methodology.[24,25] A review of quantitative risk-benefit methodologies for assessing drug safety and efficacy by Guo et al summarized 12 methods.[26] Many of these methods are based on a two-dimensional view of benefit versus risk,[27,28] ie, they consider a benefit-risk plane similar to cost-effectiveness planes in health economics.[29] Such a visualization of the incremental risk and benefit can assist in discussing the acceptable trade-off between risks and benefits with a decision-maker, a regulator/third-party payer, physician, or patient.

In a superiority trial, the benefit usually relates to efficacy and the risk to safety issues. On the contrary, in case of an NI trial, benefit is the gain in (potential) advantage of the new treatment while risk is the loss in efficacy. In an NI setup, a two-dimensional approach with the deficit, ie, the loss in efficacy of the new treatment as compared with the standard treatment on one axis and the gain in advantage on the other axis, provides a simple framework for discussing the trade-off, similar to the cost-effectiveness framework widely used in health economics.[30] Balancing of the gain in advantage versus the deficit in efficacy can then be approached, when both are expressed on comprehensible scales. We refer to this concept as advantage deficit assessment (ADA).

In the first part of this paper, we introduce the concept of ADA in analyzing NI trials. In the second part, we investigate its feasibility by empirically studying recently published NI trials. We also investigate the potential changes in conclusions among these trials when performing an ADA.

Materials and methods
Advantage deficit assessment
We assume, in the following, that we can quantify the gain in advantage by a value ΔA, which may be, for example, a difference in the rate of adverse events or mean values of a quality of life score, where the difference is taken in a way that positive values of ΔA reflect a gain in advantage by the new treatment. The gain in advantage has to be balanced against a deficit in efficacy which we assume to be quantifiable by a value ΔE, eg, the difference in median survival between the old and the new treatment. Positive values of ΔE reflect a lower efficacy of the new treatment. A simultaneous consideration of both ΔA and ΔE can be facilitated by plotting the corresponding point on a two-dimensional plane with ΔA on the x axis and ΔE on the y axis (Figure 1). The loss in efficacy increases along the x axis from left to right, representing a lesser and lesser efficacious new treatment. The gain in advantage increases along the y axis from bottom to the top with the coordinates cutting each other at 0. The positive values above the horizontal axis represent an increasingly advantageous new treatment compared with the standard treatment. Labeling the four quadrants as north east, north west, south east, and south west, the north west quadrant indicates a more efficacious and more advantageous new treatment, while the south east quadrant indicates a less efficacious and less advantageous new treatment.

While the north west and south east quadrants represent straightforward situations, the north east quadrant is a gray area where a trade-off needs to be done. It is possible to divide this quadrant into an acceptable and unacceptable area by drawing a straight line corresponding to a specific value of the advantage deficit ratio (ADR), defined as

$$\mathrm{ADR} = \frac{(\Delta A)}{(\Delta E)}.$$

The decision-maker, eg, the patient, has the freedom to decide where to draw this line, ie, to fix the threshold λ for the ratio as the minimum amount of gain in advantage required per unit deficit in efficacy. If the estimates are above the line, the observed ADR is above the threshold; if it is below the line, the observed ADR is below the threshold. Various approaches are available to support the discussion about influence of choice of threshold on the final decision. They are based either on directly considering the ratio or the so-called net benefit (NB).

$$\mathrm{NB}\,(\lambda) = (\Delta A) - \lambda \cdot (\Delta E)$$

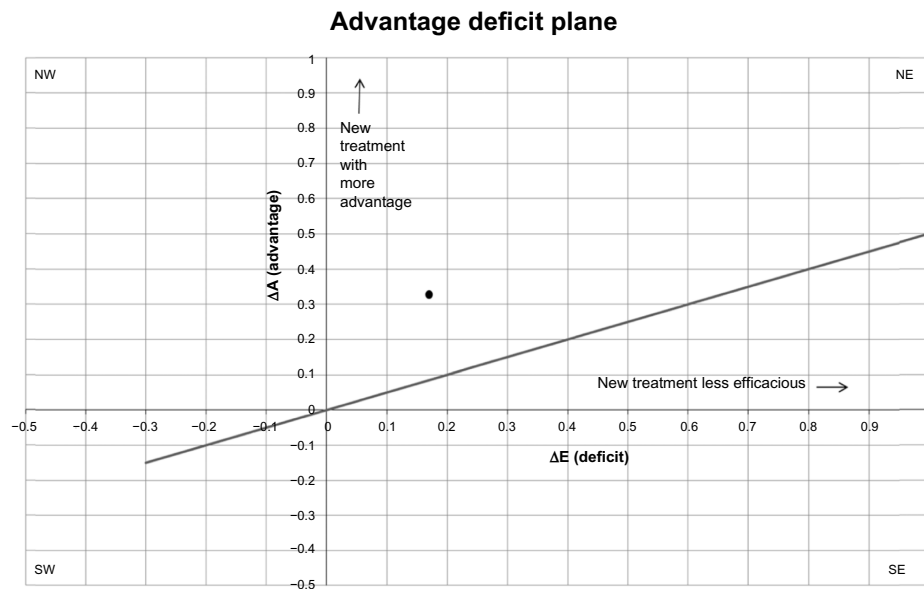**Advantage deficit plane**

Figure 1 Advantage deficit plane useful for advantage deficit assessment of a new treatment. The dot represents the results of a trial and the line corresponds to a prespecified value λ of the ADR. NW refers to the northwest quadrant and SE to the southeast quadrant.
Abbreviations: ADR, advantage deficit ratio; NE, north east; NW, north west; SE, south east; SW, south west.

ie, the "additional" benefit beyond a desired value λ for the ADR. Such approaches often try to take the stochastic uncertainty also into account.[31–33]

Advantage or efficacy measures can be based on events such as cure, failure, death, recurrence, or side effects, or based on quantities such as biological measurements (eg, blood pressure, blood glucose level, serum levels), scores (eg, quality of life scores, disease severity scores), and costs.

However, an ADA is possible only when the gain in advantage and the deficit in efficacy can be related in a meaningful way. In the following sections, we highlight three important issues to be considered here.

Type of statistical measure for gain and deficit

The numerical interpretation of ADRs would be simple when the advantage or deficit is expressed as a difference in proportions/rates of events or as a difference in means of quantitative measures, as shown in Table 1. Relative measures, such as hazard ratios or relative risk, do not refer to an absolute number of events, and this makes them hard to relate. For example, it is hard to compare a risk increase by a factor of 1.25 with respect to a failure of a treatment with a risk decrease by a factor of 0.8 for an adverse event if we do not know the absolute rates of the treatment failure and the adverse event, respectively.

Perspective

ADAs can be performed from different perspectives, eg, societal, patient, or clinician. A societal perspective would

be appropriate if a general choice for a whole population is to be made. Important outcomes from such a perspective are costs and use of resources. A clinician perspective would be appropriate if clinicians have to make the treatment decision or if they have to support patients in their choice of treatment. A patient perspective is particularly appropriate if both treatments will be available in the future, and based on their personal preferences, patients have to make their own choice. However, the patient perspective should ideally be part of both the societal and clinical perspectives.

Table 1 Interpretation of advantage deficit ratio depending on type of outcome and statistical measures used to assess advantage and efficacy

	Advantage (ΔA)	
	Difference in proportions/rates of advantage events	Difference in means of quantitative advantage measures
Deficit (ΔE)		
Difference in proportions/ rates of efficacy events	Number of events gained in advantage per one event lost in efficacy	Gain in mean advantage per one efficacy event lost in 100 patient (year)s
Difference in means of quantitative efficacy measures	Number of events gained in advantage in 100 patients or 100 patient years per unit mean lost in efficacy	Gain in mean advantage per unit mean lost in efficacy

Note: Proportions are assumed to be expressed as percentages, rates as number of events per 100 patient years.

ADAs require that the events or quantitative outcomes are related to and comprehensible from the desired perspective. In the case of a patient perspective, it seems important to distinguish between events directly related to the experience of the patient (eg, emesis, fatigue) and events only comprehensible to a clinician (eg, toxicity defined by laboratory measurements). Similar distinctions can also be made for quantitative outcomes, with patient-reported outcomes related to quality of life and satisfaction as prominent examples for outcomes related to the patient perspective.

Outcome units

When quantitative outcomes are used in computing an ADR, the unit of the outcome is directly involved in its interpretation (as in Table 1). Even if a variable is in general relevant and comprehensible to a patient, concrete units may lack a direct interpretation allowing a balancing against another variable. Raw scores or z-scores on quality of life or visual analog scale measurements of pain, are examples of this kind. From the societal perspective, use of resources expressed in number of investigations or number of hours are less useful than expressing them in a specific currency.

Illustrative examples of ADAs in NI trials

In the following examples, we illustrate some of the points mentioned above. Table 2 presents data related to ADA from five published NI trials.

In the first example, an oral drug, mycophenolate mofetil, was compared with intravenously administered cyclophosphamide, the then standard treatment for lupus nephritis, in an NI trial.[34] The important anticipated advantage of mycophenolate mofetil is avoidance of the potentially severe toxic effects of cyclophosphamide, including severe infections. The results showed that the mycophenolate mofetil arm had a significantly higher complete remission rate in addition to a significantly lower rate of severe infections. Here, the result of an ADA is rather clear; the new drug is better with regard to advantage as well as efficacy.

In the second example, intravenous terutroban was compared with aspirin, with the anticipated advantage of a lower rate of bleeding and a plan to claim noninferiority at a maximum hazard ratio of 1.05 with respect to a cardiovascular composite endpoint.[35] The results, however, showed a higher bleeding rate along with lower efficacy. Although these tendencies are not significant, there is no reason for

Table 2 Examples of currently published noninferiority trials with regard to feasibility of advantage deficit assessment

Sno	Research area	Advantage					Deficit					NI margin	Focus
		Outcome	New	Standard	Advantage	95% CI	Outcome	New	Standard	Deficit	95% CI		
1[34]	Treatment of lupus nephritis	Severe infection rate	1%	8%	7%	0.5% to 14%	Complete remission rate	22.5%	5.8%	−16.7%	−5.6% to −27.9%	10%	Better in both
2[35]	Prevention of recurrence of cerebral/CV ischemic events	Bleeding rate	15%	14%	−1%	−2% to 0.002%	CV composite endpoint	11%[b]	11%[b]	0.009[a]	−0.027 to 0.049[a]	0.021[a]	Worse in both
3[36]	Treatment of acute pulmonary embolism – OP care versus IP care	Number of hospital days	0.5	3.9	3.4	2.7 to 4.1	Rate of recurrent venous thromboembolism	0.6%	0%	0.6%	−2.1% to 2.7%	4%	Useful to do ADAs
4[37]	In vitro fertilization	Multiple pregnancy rate	0.5%	13.1%	12.6%	7.8% to 17.4%	Fertility rate	43.4%	44.7%	1.3%	−7.2% to 9.8%	13%	Individual decision possible
5[38]	Radiotherapy for breast cancer	Radiotherapy-associated toxicity	0.5%	2.1%	1.6%	0.66% to 2.5%	Recurrence rate	1.2%	0.95%	0.25%	−1.04% to 1.54%	2.5%	Multiple advantage measures with gain and loss, complicated to use ADAs
		Complications	17.6%	15.5%	−2.1%	−5% to 1%							

Notes: [a]Values are log hazard ratios; [b]proportion of patients who experienced first event (CV composite endpoint).
Abbreviations: ADA, advantage deficit assessment; CI, confidence interval; CV, cardiovascular; IP, inpatient; OP, outpatient; Sno, Serial number; NI, noninferiority.

a more formal ADA involving ADRs as there is no hint to any benefit at all.

Our third example[36] shows the result of a trial comparing outpatient care versus inpatient care for venous thromboembolism (VTE), with an anticipated advantage of a shorter hospital stay, implying more comfort from the patient perspective and less expenditure from the societal perspective. Given that outpatient care was associated with a slightly higher rate of VTE recurrence and a distinct decrease in hospital days, it may be helpful to explicitly balance deficit and gain in this case. Number of days of hospital stay, recurrence rate of VTE, and their units are comprehensible to the patient. Taking the estimates as true values results in an observed ADR of 0.6/3.4=0.18, which means that a patient would have a 0.18% higher probability of having a VTE recurrence on average if he/she is to stay one day less at the hospital. This allows the patient/society to make a choice by answering the question, "Am I/Are we ready to accept 1.8 more events in 1,000 patients for staying on an average one day less at the hospital?" In a worst case scenario (based on taking the bounds of the confidence interval [CI] as true values), the ADR increases to one more event in 100 patients per one day less at the hospital.

Similarly, in the fourth example,[37] comparison of a mild version of fertility treatment versus the standard treatment procedure with an anticipated advantage of a lower multiple pregnancy rate shows that a deficit of 1.3% lesser probability

of becoming pregnant is to be weighed against a 12.6% lower probability of multiple pregnancy. Here, acceptable ADRs are probably highly variable from couple to couple, and the ADA should support an individual choice. Figure 2 illustrates how an ADA might look like in this situation: a couple with an individual threshold of 5 would accept the new treatment, as the point estimate (corresponding to an observed ADR of 12.6/1.3=9.7) is above the line. A couple with an individual threshold of 20 would not accept the new treatment, as the point estimate is below the line. However, it may happen that the advantage is comprised of several components that act in opposite directions.

In our fifth example, intraoperative radiotherapy versus whole breast radiation for breast cancer was compared, with the potential advantage of less radiotherapy-associated toxicity.[38] Although the trial found a significant gain in advantage and a loss of efficacy within the noninferiority margin, the rate of complications associated with intraoperative radiation was found to be higher. Such findings complicate ADA assessment and require appropriate strategies to handle multiple outcomes.

Assessment of potential advantage of the new treatment in current NI trials

To study the feasibility of ADA in NI trials, we investigated trial reports with respect to statements about the expected/potential advantage, availability of data on the gain in
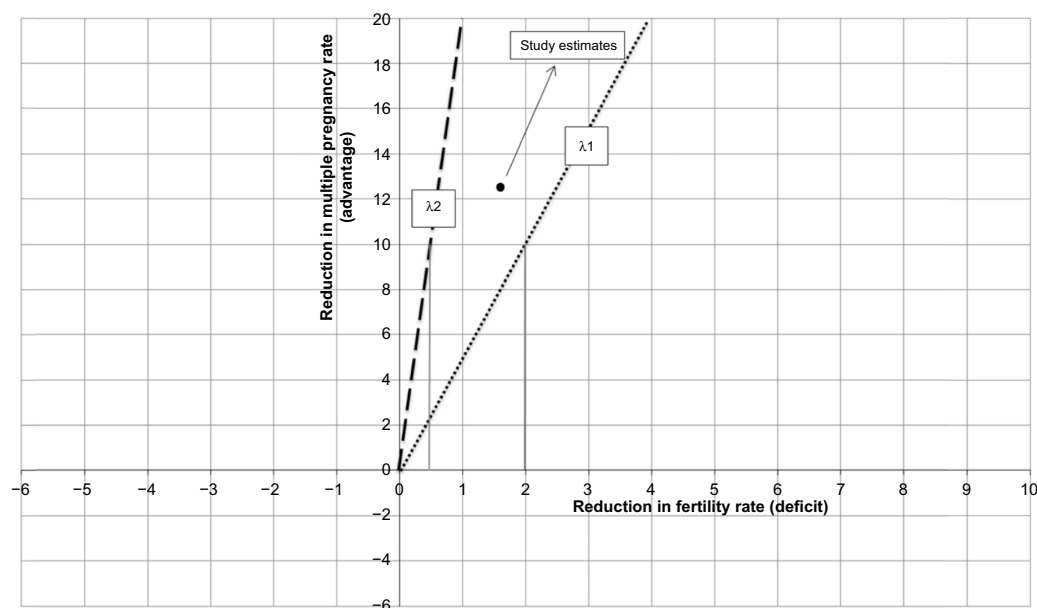

Figure 2 Advantage deficit plane based on a randomized controlled trial studying fertility treatments by Heijnen et al[37] (mild treatment versus conventional standard treatment). Threshold ADR, λ1 of 5: couple 1 requires a 10% reduction in multiple pregnancy rate to accept a 2% reduction in fertility rate. Threshold ADR, λ2 of 20: couple 2 requires a 10% reduction in multiple pregnancy rate to accept a 0.5% reduction in fertility rate.
Abbreviation: ADR, advantage deficit ratio.

advantage, and type of outcomes used for assessing the advantage and the loss in efficacy. We looked for these parameters in a set of NI trials assembled from four major medical journals known for their long history of intensive and high-quality publications in medicine, ie, "Journal of American Medical Association", "New England Journal of Medicine", "The Lancet", and the "British Medical Journal", assuming that they represent a high standard of study methodology, conduct, and reporting. We included all NI trials published from 2005 to 2011 and studying noninferiority of efficacy of a new drug/treatment/therapeutic procedure/ diagnostic procedure as the primary objective. NI trials aimed at determining the optimal dose of a drug but without comparison with a standard drug were excluded. Vaccine trials were excluded because they typically have many primary endpoints studying various strain-specific/subtype-specific antibodies and often consider protective rates close to 100%. Identification of trials from the major journals is presented elsewhere in detail.[39]

Claim of a potential advantage

The presence of a claim of potential advantage of the tested treatment was assessed based on the introduction and methods sections of the published article. Such a claim may be mentioned explicitly as a benefit or advantage or vaguely as a proposition. A potential advantage is also regarded as claimed when the disadvantage of the standard treatment is mentioned and the new treatment, either in a clear or subtle way, is said to overcome it. The claimed potential advantage could be stated specifically eg, fewer days in hospital, less cost, or in general terms, eg, safety or tolerability.

Assessment of the claimed advantage

A claimed advantage is said to have been assessed when the advantage/one or more strongly related advantage variables have been studied in the trial and results specific to the treatment groups have been published in the article; a formal comparison of treatment groups may or may not have been performed. A single trial may have one or more assessed advantages. When a trial is claiming a general advantage such as better safety/tolerability or less complications, it may use a set of advantage-related variables (such as data for various adverse events) and/or summarize them in a composite summarizing variable (such as total number of patients with adverse events). The composite variable, wherever available, was included in the further analysis; if not, the whole set of advantage-related variables was included. For each claimed but not assessed advantage, we judged whether it could be

assessed, in principle, by measuring some variables at the individual level.

For each trial, the primary outcome used in testing noninferiority was chosen as the efficacy variable. For all advantage and efficacy variables, we extracted the estimate of the group difference and any available measure of its precision (95% CI, standard error, standard error/standard deviation/interquartile range in each arm). A significant gain was when the point estimate of the gain in advantage was greater than 0 and the 95% CI of the measure did not include 0 (point estimate favors the new treatment and is significant). A nonsignificant gain was when the point estimate was greater than 0 but the 95% CI included 0 (point estimate favors the new treatment but is not significant). It was considered as no gain when the point estimate was less than or equal to 0 (point estimate does not favor the new treatment).

A variable was classified as quantitative when it was measured on a continuous scale or presented as a count. It was classified as event-related when the variable was binary with the proportion of events reported for each treatment arm or when it was time to event with incidence rates or event probabilities at a certain time point, extractable from the publication. Otherwise, it remained unclassified. Additionally, each efficacy and advantage variable was classified either as patient-comprehensible or not. The variable was called patient-comprehensible if it related to the patient's subjective wellbeing and was based on patient reporting (eg, quality of life scores, pain scores) and/or relates to events or circumstances that can be directly experienced by the patient (eg, vomiting, number of days in hospital). If a variable was not patient-comprehensible, we classified it either as clinician-comprehensible (if it was related to biological or biochemical measurements, eg, serum creatinine level or thickness of the carotid intima) or as societal-comprehensible (if it was related to use of resources or costs). Further, the units of quantitative and patient-comprehensible variables were classified as patient-comprehensible (if they referred to counts, eg, of certain events) or patient-incomprehensible (if they were raw scores or z-scores).

Results

Of the 113 published NI trials identified from the four major journals, 87 (78%) claimed a potential advantage for the tested treatment. The advantage was mentioned explicitly in 58 trials, only as a vague mention among eight, and indirectly referred to by mentioning the disadvantage of the standard treatment in 21 trials. Overall, 65 trials with one or more specific advantages and 22 with general advantages resulted in 170 advantage claims.

At least one claimed advantage was assessed in 55 (63%) of the trials, which added up to 80 assessed advantage claims. Only 21 (24%) of the trials assessed all of their claimed advantages. A composite advantage variable was found among 16 of the 22 trials mentioning general advantage.

The various types of advantage claimed are presented in Table 3. Safety was the most frequent advantage factor, accounting for 42% of all the claimed advantages, followed by resource/cost (15%) and ease of administration (14%). All the seven advantage claims related to patients' subjective well-being and the one related to compliance were assessed. Among the safety-related advantages, 79% were assessed, but the rates were lower for the other groups. Advantages related to ease/mode of administration and duration of treatment were very rarely assessed.

Of the 90 claimed but not assessed advantages, it would have been possible, in our opinion, to assess the advantage in 71 (79%) of them. The 19 nonassessable advantages included ten at the level of the health care system, such as accessibility or development of antibiotic resistance. The others were four long-term risks like infertility and cardiovascular events which require a longer follow-up than feasible in a clinical trial and five in vivo or molecular level mechanisms, such as rapid endothelialization, less inflammation, and less stent thrombosis, which are often not directly measurable.

The 55 trials with advantage assessment contributed 55 efficacy variables and 128 advantage variables. Ninety-nine percent and 76% of the advantage and efficacy variables, respectively, were presented originally as differences in rates/proportions/events while the rest were presented as relative measures such as hazard ratios, odds ratios, and relative risks. However, event probabilities or rates could be extracted for all those with relative measures. All variables could be subsequently classified as quantitative or event-related and 14% of the advantage and 9% of the efficacy measures were quantities; with regard to comprehensibility, 78% and 87%, respectively, were patient-comprehensible. Quantitative efficacy measures were always patient-incomprehensible, whereas the quantitative advantage measures were often patient-comprehensible. Eleven (65%) of the 17 comprehensible advantage measures had comprehensible units as well. Considering all 128 advantage and efficacy variable pairs (Table 4), 75 (59%) involved patient-comprehensible measures as well as units.

Of the 128 assessed advantage variables, data on precision were available for 119 variables representing 49 trials (Table 5). The gain in advantage was significant among 55/119 (46%) advantage variables, while 34 (29%) variables did not favor the new treatment. When the combined results for advantage and efficacy were studied, we found three large interesting groups: 27 pairs with a significant gain in the advantage measure and a noninferior deficit, 21 with a significant gain in the advantage measure and a nonsignificant gain in efficacy, and 18 with no gain in the advantage measure and a noninferior deficit in efficacy. In the first two groups, an ADA has the potential to assess a significant net benefit, whereas in the last group, an ADA will probably question any recommendation for the new treatment.

Table 3 Potential advantages mentioned for the new treatment in the current published noninferiority trial reports

Broader classification of advantage	Advantages involved	Number claimed	% claimed	Number assessed	% assessed
Safety	Toxicity, tolerability, adverse event, side effect, complication or safety	71	42%	56	79%
Resource	Cost or resource	25	15%	5	20%
Patient's subjective wellbeing	Patient burden of treatment, satisfaction or quality of life	7	4%	7	100%
Accessibility	Availability, accessibility	4	2%	0	0%
Ease of administration	Easy to use, no monitoring needed, fixed dose	24	14%	2	8%
Mode of administration	Oral/sublingual versus SC/IM/IV	9	5%	0	0%
Duration of treatment	Shorter or less often	15	9%	2	13%
Antibiotic use	Avoidance/shorter duration	6	4%	2	33%
Compliance		1	1%	1	100%
Medical process averted	No general anesthesia No pacemaker implantation	2	1%	1	50%
Others	Less inflammation Rapid endothelialization Less thrombosis	6	3%	4	60%
Total		170	100%	80	47%

Abbreviations: IM, intramuscular; IV, intravenous; SC, subcutaneous.

Table 4 Comprehensibility assessment of efficacy and advantage measures

	Efficacy measures			Total
	Events		Quantitative measures	
	Patient-comprehensible	Patient-incomprehensible	Patient-incomprehensible	
Advantage measures				
Events				
Patient-comprehensible	68	5	10	83
Patient-incomprehensible	21	1	5	27
Quantitative measures				
Patient-comprehensible with comprehensible units	7		4	11
Patient-comprehensible with incomprehensible units	6			6
Patient-incomprehensible	1			1
Total	103	6	19	128

Discussion

The aim of demonstrating noninferiority in NI trials often puts the focus on the tolerated loss in efficacy rather than the advantage traded for. Similar to the importance of safety assessment in superiority trials, evident in the ongoing efforts by the US Food and Drug Administration and European Medicines Agency,[24,25] an explicit assessment of advantage can contribute substantially to the interpretation and acceptance of NI trials. If an NI trial can demonstrate an advantage of sufficient magnitude, It may be possible to demonstrate a positive net benefit instead of just demonstrating non-inferiority. On the other hand, if there is no evidence of advantage, a conclusion of noninferiority may no longer be satisfactory. Balancing the gain in advantage against the deficit in efficacy is naturally the starting point. Transferring the concept of benefit-risk assessment to NI trials, we have presented in this paper some general thoughts as to how such an ADA can be performed.

In the first part of the paper, we have tried to clarify some conceptual issues related to interpretation of an ADR. It is preferable that advantages and deficits are expressed as differences in rates/proportions or means, as this facilitates the numerical interpretation of ADRs; the outcomes and their

units should be relevant to and comprehensible from the perspective of being used for ADA.

Based on these general thoughts, we investigated published NI trials with respect to the feasibility of ADA. Our investigation reveals that half of the NI trials published today include some form of assessment of advantage. This is in line with Bernabe et al,[40] who reported that 22/41 (54%) published (Phase IV) NI trials mentioned an additional benefit of the study drug. Similarly, Schiller et al found that 48% of 167 NI trials published in 2009 justified the choice of design and 34% reported the advantage of the new treatment in detail.[41] Moreover, for the majority of the investigated trials which had not assessed its claimed advantage, an assessment within the trial seemed possible and hence an ADA would have been feasible if planned from the beginning. Nonfeasibility seems to be mainly related to advantages perceived only at the level of the health care system and advantages involving long-term prognosis.

Using ADA as a standard tool for analyzing NI trials not only seems to be feasible but also promising in the sense that it might change substantially the final judgment of the study results. We propose ADA assuming that all other conceptual issues related to NI trials except the choice of noninferiority margin have been addressed effectively. Of the 113 joint

Table 5 Direction and significance of the effects of the advantage and efficacy measures

	Efficacy measures				Total
	Deficit and not noninferior	Deficit but noninferior	Nonsignificant gain and noninferior	Significant gain	
Advantage measures					
No gain	2	18	12	2	34
Nonsignificant gain	0	14	13	3	30
Significant gain	4	27	21	3	55
Total	6	59	46	8	119

18

Open Access Journal of Clinical Trials 2015:7

assessments of efficacy and advantage from 44 NI trials that declared noninferiority, 51 (45%) assessments from 23 trials had a significant advantage gain. In all these 23 trials, there is hence a potential to finally come to a significant net benefit in an ADA instead of a simple noninferiority statement specifically, given that the gain in efficacy was positive in 15 of them. On the other hand, we could identify 18 trials that declared noninferiority, but did not have a significant gain in any advantage variable.

The patient perspective is an important issue for benefit assessments performed by regulatory or health technology assessment agencies today.[25,42,43] It is worth noting that the majority of potential ADAs identified by us among the current NI trials could be based on patient-comprehensible outcomes. Moreover, it was always possible to express observed treatment effects as a difference in means, proportions, or rates. Both these features together allow patients potentially benefiting from these new therapies to decide for themselves based on their preferences, circumstances, and perception. The major source for lack of patient comprehensibility was serum biochemistry levels, eg, neutropenia, leukopenia, and lymphopenia, and pathologic changes such as a venous thrombotic event, stenosis, pericardial effusion, fluid retention, and neurotoxicity.

ADA should be planned prospectively in future NI trials. One or more advantage-related variables need to be measured at the individual level to be able to assess the anticipated advantage. Even when the advantage seems to be a simple consequence of the new treatment (eg, ease of administration or shorter treatment duration), it is a good idea to make an assessment in order to ensure that there are no other unexpected circumstances (eg, noncompliance or adverse effects) acting against this advantage. ADA planning should include the choice of a specific pair of advantage and efficacy measures which is to be combined into an ADR as the primary endpoint as well as a prespecified threshold λ. This would also allow power calculations to be performed. However, secondary analyses could be carried out to include various other choices of λ or ADRs later on.

Today, the use of benefit-risk assessments is already widespread in the routine analysis of clinical trials, if the studies follow a superiority setup. However, the concept of ADA in NI trials would mean a major change in the current concepts regarding planning and analysis of such trials. In this paper, we have not discussed the statistical techniques used to perform ADAs, because well established methods from cost-effectiveness analyses or benefit-risk assessment can be used. There is, however, one subtle difference, namely that in an ADA, an unexpected change of quadrant may happen

more frequently, as we assume a priori ΔE to be close to zero. In particular, if a treatment turned out to have a negative gain in advantage but a gain in efficacy, one should not use the ADR thresholds, chosen prior to the study when the opposite situation was anticipated, uncritically. Another issue to be considered is that improved safety is a very common advantage. It is often assessed by adverse events, and adverse events are often rare. For this reason, one may expect power problems in assessing a gain in advantage. This expectation is not completely unsubstantiated, as 13 of the 56 binary advantage variables in our investigation had a prevalence below 5% or above 95%. Therefore, the choice of advantage outcomes with sufficient prevalence may be an issue in planning ADAs.

Under some circumstances, an ADA can be based on comparing advantage and efficacy at the individual level instead of the study level. If both are, for example, measured at a continuous scale, ratios or net benefits can be computed for each individual. Durkalski and Berger[15] made a suggestion applicable also for categorical outcomes, ie, to translate the combined values into at least a partially ordered score.

There are some limitations to our study. We used NI trials published in four major journals, which may not be representative of all NI trials. The fraction of trials claiming and assessing the advantage may be too optimistic; however, our feasibility assessment of ADA is not dependent on the quality of reporting or trial conduct, but only on the type of anticipated advantage. Our results were based on 113 trials, and only 55 trials contributed to an analysis of efficacy and advantage measures. All our analyses were of an exploratory nature and aimed to identify frequent or infrequent patterns. Most of our conclusions are qualitative rather than quantitative, in the sense that we can see a potential for improvement. The degree to which this potential can be realized will become clearer in the future. Another limitation is that our illustrations of ADR were limited by the fact that we could not include inferential statistics, eg, CIs for the ADR, because these depend on the correlation between ΔA and ΔE. It is not usual to report these correlations in publications.

Conclusion

Our study indicates that it is time to use ADA as a main analytical approach for NI trials. We could not identify major obstacles to the use of ADA. ADA overcomes the current limitation of NI trials to demonstrate "only" noninferiority, bringing NI trials back to the mainstream of superiority trials by aiming to demonstrate a positive net benefit.

Acknowledgment

Author contributions

Both authors have made substantial contributions to the conception and design, or acquisition of data, or analysis and interpretation of data; have been involved in drafting the manuscript or revising it critically for important intellectual content; have given final approval of the version to be published; and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Disclosure

The authors report no competing interests in this work.

References

1. D'Agostino RB, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues – the encounters of academic consultants in statistics. *Stat Med.* 2003;22:169–186.
2. Fleming TR, Odem-Davis K, Rothmann MD, Shen YL. Some essential considerations in the design and conduct of non-inferiority trials. *Clin Trials.* 2011;8:432–439.
3. Fueglistaler P, Adamina M, Guller U. Non-inferiority trials in surgical oncology. *Ann Surg Oncol.* 2007;14:1532–1539.
4. Committee for Medicinal Products for Human Use. Guideline on the choice of the noninferiority margin. London, UK: European Medicines Agency; 2005. Available from: http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003636.pdf. Accessed January 1, 2012.
5. US Food and Drug Administration. Guidance for industry non-inferiority clinical trials – draft guidance. 2010. Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM202140.pdf. Accessed November 30, 2014.
6. Garattini S, LiBassi L, Bertele V. Placebo or active control? Either, as long as it is in the patient's interest. *WHO Drug Information.* 2003;17:253–256.
7. Wangge G, Klungel OH, Roes KCB, de Boer A, Hoes AW, Knol MJ. Should non-inferiority drug trials be banned altogether? *Drug Discov Today.* 2013;18:601–604.
8. Guyatt GH, Mulla SM, Scott IA, Jackevicius CA, You JJ. Patient engagement and shared decision-making. *J Gen Intern Med.* 2014;29:562.
9. Hoffman RM, McNaughton-Collins M. The superiority of patient engagement and shared decision-making in noninferiority trials. *J Gen Intern Med.* 2014;29:16–17.
10. Mulla SM, Scott IA, Jackevicius CA, You JJ, Guyatt GH. How to use a noninferiority trial: users' guides to the medical literature. *JAMA.* 2012;308:2605–2611.
11. Bauer P, Brannath W, Posch M. Multiple testing for identifying effective and safe treatments. *Biom J.* 2001;43:605–616.
12. Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. *Stat Med.* 2009;28:586–604.
13. Bretz F, Posch M, Glimm E, Klinglmueller F, Maurer W, Rohmeyer K. Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. *Biom J.* 2011;53:894–913.
14. Burman C-F, Sonesson C, Guilbaud O. A recycling framework for the construction of Bonferroni-based multiple tests. *Stat Med.* 2009;28:739–761.
15. Durkalski VL, Berger VW. Re-formulating non-inferiority trials as superiority trials: the case of binary outcomes. *Biom J.* 2009;51:185–192.
16. Guilbaud O. Note on simultaneous inferences about non-inferiority and superiority for a primary and a secondary endpoint. *Biom J.* 2011;53:927–937.
17. Nishikawa M, Tango T, Ohtaki M. Statistical tests based on new composite hypotheses in clinical trials reflecting the relative clinical importance of multiple endpoints quantitatively. *Biom J.* 2009;51:749–762.
18. Bristol DR. Superior safety in noninferiority trials. *Biom J.* 2005;47:75–81.
19. Eichler H-G, Bloechl-Daum B, Abadie E, Barnett D, König F, Pearson S. Relative efficacy of drugs: an emerging issue between regulatory agencies and third-party payers. *Nat Rev Drug Discov.* 2010;9:277–291.
20. Bennett CL, Nebeker JR, Lyons E, et al. The research on adverse drug events and reports (radar) project. *JAMA.* 2005;293:2131–2140.
21. Tsintis DP, Mache EL. CIOMS and ICH initiatives in pharmacovigilance and risk management. *Drug Saf.* 2004;27:509–517.
22. Edwards DIR, Wiholm B-E, Martinez C. Concepts in risk-benefit assessment. *Drug Saf.* 1996;15:1–7.
23. Zafiropoulos N, Phillips L. Evaluating benefit risk: an agency perspective. *Regulatory Rapporteur.* 2012;9:5–8.
24. Committee for Medicinal Products for Human Use. Benefit-Risk Methodology Project. London, UK: The European Medicines Agency; 2009. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Report/2011/07/WC500109477.pdf. Accessed November 30, 2014.
25. US Food and Drug Administration. Structured approach to benefit-risk assessment in drug regulatory decision-making: Draft PDUFA V implementation plan – February 2013, Fiscal Years 2013–2017. Silver Spring, MD, USA: US Food and Drug Administration; 2013. Available from: http://www.fda.gov/downloads/ForIndustry/UserFees/PrescriptionDrugUserFee/UCM329758.pdf. Accessed November 30, 2014.
26. Guo JJ, Pandey S, Doyle J, Bian B, Lis Y, Raisch DW. A review of quantitative risk–benefit methodologies for assessing drug safety and efficacy – report of the ISPOR Risk-Benefit Management Working Group. *Value Health.* 2010;13:657–666.
27. Lynd LD, O'Brien BJ. Advances in risk-benefit evaluation using probabilistic simulation methods: an application to the prophylaxis of deep vein thrombosis. *J Clin Epidemiol.* 2004;57:795–803.
28. Shaffer ML, Watterberg KL. Joint distribution approaches to simultaneously quantifying benefit and risk. *BMC Med Res Methodol.* 2006;6:48.
29. Drummond MF. *Methods for the Economic Evaluation of Health Care Programmes.* Oxford, UK: Oxford University Press; 2005.
30. Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. *Cost-Effectiveness in Health and Medicine.* 1st ed. New York, NY, USA: Oxford University Press; 1996.
31. Briggs A, Fenn P. Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. *Health Econ.* 1998;7:723–740.
32. Siegel C, Laska E, Meisner M. Statistical methods for cost-effectiveness analyses. *Control Clin Trials.* 1996;17:387–406.
33. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med Decis Making.* 1998;18(2 Suppl):S68–S80.
34. Ginzler EM, Dooley MA, Aranow C, et al. Mycophenolate mofetil or intravenous cyclophosphamide for lupus nephritis. *N Engl J Med.* 2005;353:2219–2228.

35. Bousser M-G, Amarenco P, Chamorro A, et al. Terutroban versus aspirin in patients with cerebral ischaemic events (PERFORM): a randomised, double-blind, parallel-group trial. *Lancet*. 2011;377:2013–2022.

36. Aujesky D, Roy P-M, Verschuren F, et al. Outpatient versus inpatient treatment for patients with acute pulmonary embolism: an international, open-label, randomised, non-inferiority trial. *Lancet*. 2011;378:41–48.

37. Heijnen EM, Eijkemans MJ, De Klerk C, et al. A mild treatment strategy for in-vitro fertilisation: a randomised non-inferiority trial. *Lancet*. 2007;369:743–749.

38. Vaidya JS, Joseph DJ, Tobias JS, et al. Targeted intraoperative radiotherapy versus whole breast radiotherapy for breast cancer (TARGIT-A trial): an international, prospective, randomised, non-inferiority phase 3 trial. *Lancet*. 2010;376:91–102.

39. Gladstone BP, Vach W. Choice of non-inferiority (NI) margins does not protect against degradation of treatment effects on an average – an observational study of registered and published NI trials. *PLoS One*. 2014;9:e103616.

40. Bernabe RD, Wangge G, Knol MJ, et al. Phase IV non-inferiority trials and additional claims of benefit. *BMC Med Res Methodol*. 2013; 13:70.

41. Schiller P, Burchardi N, Niestroj M, Kieser M. Quality of reporting of clinical non-inferiority and equivalence randomised trials – update and extension. *Trials*. 2012;13:214.

42. Mullard A. Patient-focused drug development programme takes first steps. *Nat Rev Drug Discov*. 2013;12:651–652.

43. European Medicines Agency. Information on benefit-risk of medicines: patients', consumers' and healthcare professionals' expectations. London, UK: European Medicines Agency; 2009. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Other/2009/12/WC500018433.pdf. Accessed November 30, 2014.