

Evaluation des médicaments : des essais de supériorité aux nouveaux schémas d'essais

Dominique Costagliola

Institut Pierre Louis d'Epidémiologie et de Santé Publique,
UMR-S 1136,

Sorbonne Universités, INSERM et UPMC Univ Paris 06



Institut Pierre Louis d'Epidémiologie et de Santé Publique
Pierre Louis Institute of Epidemiology and Public Health

Pierre Louis (French physician, 1787-1872) contributed to the development of epidemiology

Plan

- Les phases des essais de médicament
- Essais de non-infériorité
- Essais adaptatifs

Les différentes phases des essais visant à évaluer les médicaments avant la mise sur le marché

- **Essai de phase I : le produit est-il sûr ?**
 - Le produit expérimental, le candidat médicament, est testé chez un petit groupe de personnes (le plus souvent des **volontaires sains**), pour la première fois, dans le but d'**étudier comment il se comporte dans l'organisme humain**
 - L'essai de **phase I** a pour but de répondre aux questions suivantes :
 - Quelle est la **dose maximale tolérée** (études de tolérance et de toxicité) ?
 - Comment est-il assimilé dans l'organisme : administration, diffusion, métabolisme, excrétion (étude de **biodisponibilité** ou de **pharmacocinétique** clinique) ?
 - Quels en sont les **effets secondaires** dans ces conditions d'utilisation ?
- **Essai de phase II : le produit a-t-il un effet bénéfique ?**
- **Essai de phase III : le produit a-t-il un effet thérapeutique sur un plus grand effectif ?**

Les différentes phases des essais visant à évaluer les médicaments avant la mise sur le marché

- **Essai de phase I : le produit est-il sûr ?**
- **Essai de phase II : le produit a-t-il un effet bénéfique ?**
 - L'essai de **phase II** est mené sur un petit **groupe homogène de volontaires atteints de la maladie ciblée**
 - L'objectif est d'étudier l'**efficacité** du produit, d'approfondir les connaissances sur sa **tolérance** et de rechercher la plus petite dose efficace pour **déterminer la dose optimale** pour l'essai de phase suivante (phase III)
- **Essai de phase III : le produit a-t-il un effet thérapeutique sur un plus grand effectif ?**

Les différentes phases des essais visant à évaluer les médicaments avant la mise sur le marché

- **Essai de phase I : le produit est-il sûr ?**
- **Essai de phase II : le produit a-t-il un effet bénéfique ?**
- **Essai de phase III : le produit a-t-il un effet thérapeutique sur un plus grand effectif ?**
 - L'essai de **phase III** se fait sur une **grande population de volontaires atteints de la maladie ciblée**
 - L'objectif est de confirmer la **tolérance** (en surveillant l'apparition d'éventuels effets secondaires qui seraient passés inaperçus lors des phases précédentes) et l'**efficacité** du candidat médicament (**en comparant** son efficacité à celle d'un **placebo** ou d'un **traitement existant**)
 - Le but est de déterminer dans quelles conditions d'utilisation le produit est le plus efficace

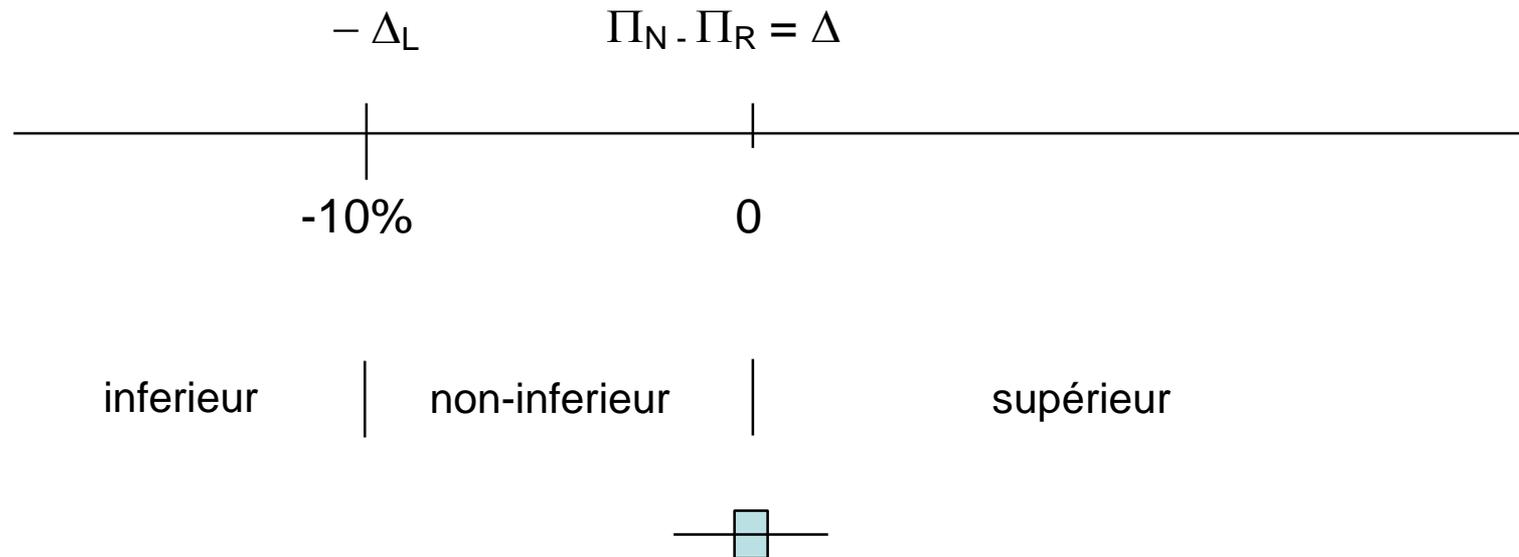
Essais de non-infériorité

Quel objectif pour un essai clinique de phase 3 ?

- Par comparaison à un traitement de référence R
 - Montrer la supériorité du nouveau traitement N
 - N est-il meilleur que R ?
 - Montrer la non-infériorité du nouveau traitement
 - N ne fait pas moins bien que R ?
 - Montrer l'équivalence du nouveau traitement
 - N fait aussi bien (ni mieux ni moins bien) que R ?
 - Bioéquivalence
 - Générique

Définition de la non-infériorité

- N ne fait pas moins bien que R



Choix de la borne de non-infériorité - 1

- Décision clinique et non statistique
- La plus grande différence qui reste cliniquement acceptable
- \leq différence utilisée dans les essais de supériorité du domaine
- Garantir que le nouveau traitement fait mieux que le placebo

Un exemple dans le domaine du diabète: HbA1c et risque de décès- 1

- Dans le diabète, on utilise l'HbA1c comme critère de jugement
 - Borne de non-infériorité en général 0,6%
 - Les essais de supériorité visent à montrer une différence du taux d'HbA1c de 1%

Un exemple dans le domaine du diabète: HbA1c et risque de décès- 2

- La réduction du taux d'HbA1c de 1% est associée à une baisse de
 - 21% des décès de cause liées au diabète (IC à 95 % :15-27%)
 - 14% des infarctus (IC à 95 % : 8-21%)
 - 37% des complications microvasculaires (IC à 95 % :33-41%)
 - Aucun seuil pour aucun des critères de jugements

Un exemple dans le domaine du diabète: HbA1c et risque de décès- 3

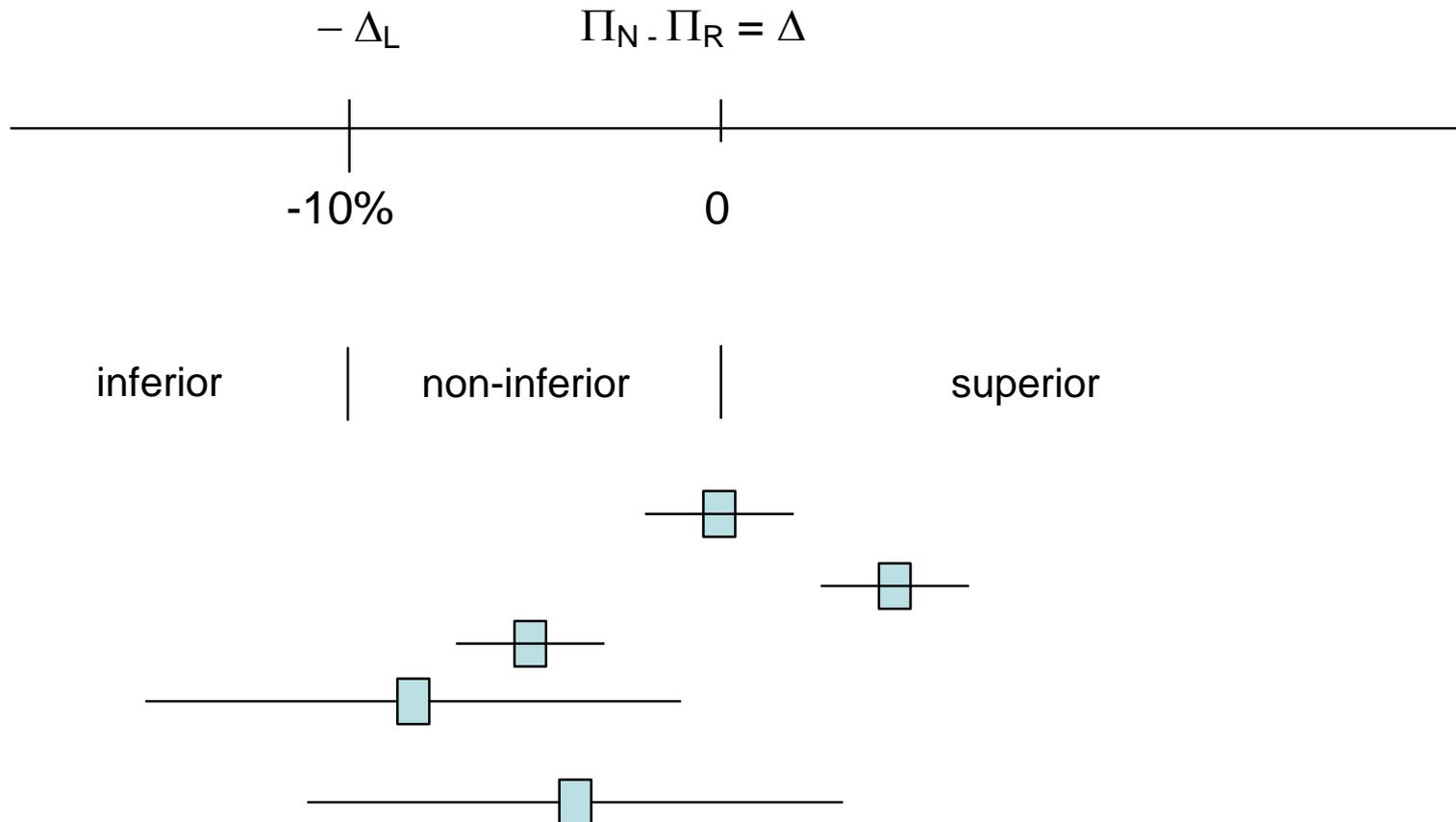
- Est-il possible de définir une borne de non-infériorité cliniquement acceptable dans ce contexte ?

Choix de la borne de non-infériorité - 2

- Comme choisir une borne de non-infériorité implique qu'on accepte de perdre sur l'efficacité
 - Il devrait y avoir un gain à utiliser un nouveau traitement
 - Facilité d'emploi
 - Tolérance
 - Coûts
 - ...

Quelques exemples d'interprétation

- La conclusion est basée sur la comparaison entre la borne inférieure de l'intervalle de confiance de la différence et la borne de non \square_L
- N ne fait pas moins bien que R



Conclusion - 1

- Essais de non infériorité
 - Si on accepte une perte de chance, quel est le gain espéré ?
 - Le choix de borne de non-infériorité est majeur
 - C'est un choix clinique pas une décision statistique
 - Qui doit garantir la supériorité par rapport au placebo
 - L'analyse en intention de traiter n'est plus l'analyse principale car elle biaise vers la non-infériorité
 - L'analyse en ITT et l'analyse per-protocole doivent donner le même résultat
 - La différence du nombre de sujets inclus dans chaque analyse donne une indication de la qualité de l'étude
 - Si la non-infériorité est démontrée on peut sans soucis majeurs examiner l'existence d'une supériorité
 - Pertinence clinique

Essais adaptatifs

Porcher R et al. Thérapie 2011
Blatt DL et Mehta C. NEJM 2016

Définition

Méthode adaptative pour un essai clinique

- Utilise les données accumulées dans l'essai pour éventuellement en modifier certains aspects
- Sans compromettre la *validité* et l'*intégrité* de l'essai

Validité

- Inférence statistique correcte (test et estimation)
- Garantir la cohérence entre les différentes phases de l'essai
- Minimiser les biais opérationnels

Intégrité

- Acceptabilité des résultats par la communauté scientifique
- Prévoir autant que possible les adaptations à l'avance
- Maintien de la confidentialité des données

Principaux types de schémas

Type d'essai adaptatif	Adaptations
Essai séquentiel par groupe	Arrêt précoce de l'étude
Réévaluation du nombre de sujets (en aveugle – variance, autres paramètres de nuisance ...)	Augmentation de la taille d'échantillon
Phase 1 escalade de doses CRM (<i>Continual Reassessment Method</i>)	Choix de la prochaine dose
Phase 1-2 intégrés	Choix de la prochaine dose
Phase 2 recherche de dose adaptative	Changement du ratio de randomisation
Réévaluation du nombre de sujets (unblinded – à partir de l'efficacité observée)	Augmentation de la taille d'échantillon
Enrichissement de la population	Modification des critères d'inclusion, de la population d'analyse → sous-groupe
Essais combinés phases 2-3 (ex -seamless)	Sélection de la dose

Perception des méthodes

- Perception par les agences de régulation
 - Objectifs
 - Essai confirmatoire (contrôle stricte du risque de 1^{ère} espèce)
 - Essai exploratoire
 - Méthodes
 - Méthodes bien connues / maîtrisées (*well understood*)
 - Autres méthodes (*less well understood*)
- Par les industriels
 - Méthode reconnue ou pas
 - Rapport bénéfice / risque pour l'essai
- Par les chercheurs (monde académique)
 - Méthode connue (du chercheur) ou pas

Choisir un schéma adaptatif

- Pourquoi?*
- Obtenir la même information qu'un essai « classique », avec une meilleure efficacité
- Augmenter la probabilité d'atteindre l'objectif de l'essai
- Améliorer la connaissance d'un traitement

- Avantages / inconvénients

+	-
Raccourcit les délais de développement	Nécessité d'un statisticien expert
Attractivité	Logistique (sauf phase 1)

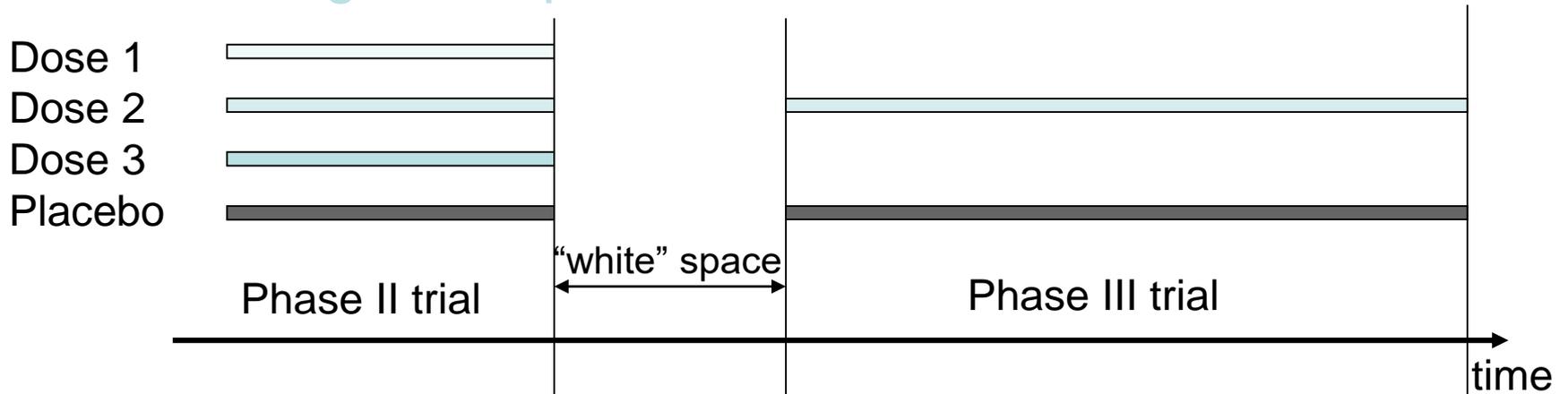
- Autres avantages / inconvénients spécifiques selon les objectifs et le type d'essai

Phases 2(b)-3 combinées

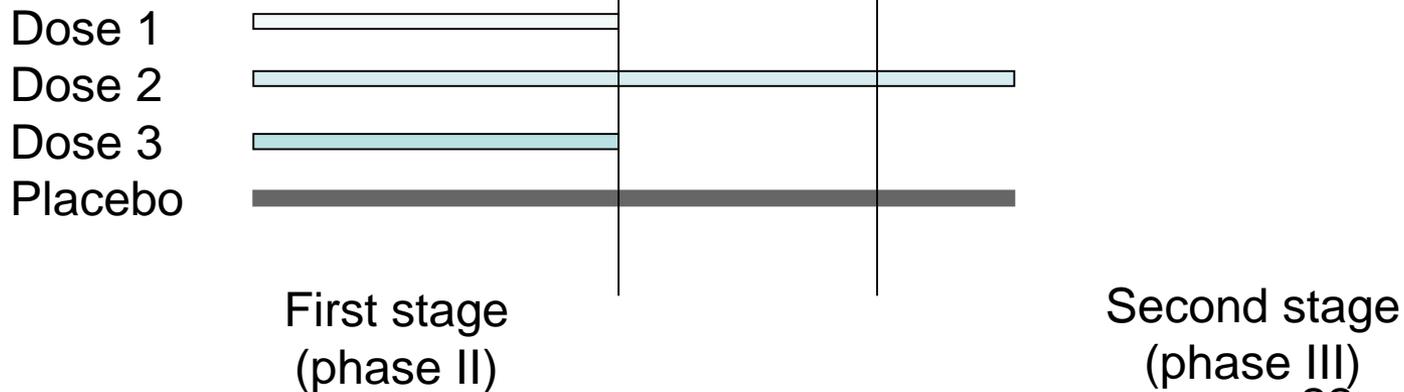
- Un seul essai, deux « phases » traditionnelles
 - Etape 1: phase 2 (p. ex. *dose ranging*)
 - Etape 2: phase 3
- Essai voulu comme confirmatoire
- Distinguer
 - Essais *opérationnellement* 2-3
 - Essais *inférentiellement* 2-3 → *seamless phase 2-3*
- Dans ce dernier cas: analyse finale sur l'ensemble des sujets inclus
- A été longtemps le « graal » de l'essai adaptatif

Schéma

Traditional drug development



Seamless phase II/III



Phase 3: enrichissement

- Essai qui débute avec une population « large »
- Et se poursuit potentiellement dans une sous-population « ciblée » si efficacité dans ce sous groupe
- Méthodologie reconnue quand
 - Les sous-groupes sont définis à l'avance
 - L'essai est planifié comme cela dès le départ
- Méthodes pour garantir le contrôle de l'erreur α
- Peut-être le nouveau « graal », en tout cas domaine de recherche actif

Enrichissement « post-hoc »

- Analyse non prévue au départ
- Voire essai non prévu comme adaptatif au départ
- Analyse dite post-hoc
- Rationnel « propre »: découverte d'un nouveau marqueur, non prévu au départ
- Autres rationnels, hum ...

Contraintes

- Planification
 - Nombreux scénarios → simulations numériques intensives
 - Pré-planification: prévoir les adaptations pour chaque scénario
- Logistique
 - Produits
 - Forme galénique « flexible », forme commerciale
 - Approvisionnement, stocks, ...
 - Données
 - Disponibilité rapide de données valides (gel de base)
 - Flux sécurisés +++
- Risques
 - Essai / réglementaire
 - Moyens
 - Patients
 - Plutôt moindres, mais possible manque de recul (ph1-2, ph 2-3)

Utilisation

- Difficile à quantifier
 - En augmentation
 - Assez souvent en cancérologie (phase 1)
 - Applications plus récente aux essais tardifs, mais ça arrive
- Monde industriel
 - Certains groupes ont une bonne expérience
 - Position des agences cruciale
 - Evaluer les risques « réglementaires » et justifier le choix
- Monde académique
 - Encore assez rare pour les phases tardives (plateforme MRC, par exemple)
 - Rapport bénéfice / contraintes (+coût) à évaluer

Conclusion -2

- Méthodologie séduisante
 - Mesurer les contraintes avant de se lancer
 - Importance de la pré-planification
- Taxonomie / classification
 - Normatif = réducteur
 - Tous les schémas n'entrent pas dans le cadre présenté
 - Langage commun → facilite la diffusion et l'acceptation
- Recherche
 - Domaine récent, en évolution
 - Adaptatif: possibilités étendues vs complexité des schémas
 - Méthodes centrées sur les tests : progrès à faire sur l'estimation

Remerciements

- Raphaël Porcher
- Chrsitine Rouzioux

Example - 1

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Dolutegravir plus Abacavir–Lamivudine for the Treatment of HIV-1 Infection

Sharon L. Walmsley, M.D., Antonio Antela, M.D., Ph.D., Nathan Clumeck, M.D.,
Dan Duiculescu, M.D., Andrea Eberhard, M.D., Felix Gutiérrez, M.D.,
Laurent Hocqueloux, M.D., Franco Maggiolo, M.D., Uriel Sandkovsky, M.D.,
Catherine Granier, D.E.S.S., Keith Pappa, Pharm.D., Brian Wynne, M.D.,
Sherene Min, M.D., and Garrett Nichols, M.D., for the SINGLE Investigators*

Example - 2

STATISTICAL ANALYSIS

According to the protocol, we could conclude that treatment with dolutegravir and abacavir-lamivudine was noninferior to treatment with efavirenz-tenofovir DF-emtricitabine if the lower boundary of a two-sided 95% confidence interval for the difference in the primary end point was less than 10 percentage points lower in the DTG-ABC-3TC group than in the EFV-TDF-FTC group. This margin is consistent with that in other trials in this population.¹⁹ Assuming a 75% response rate in the EFV-TDF-FTC group, we calculated that 394 participants who could be evaluated would need to be included in each group for the study to have 90% power to determine the noninferiority of dolutegravir and abacavir-lamivudine, at a one-sided significance level of 2.5%. Efficacy and safety analyses were performed in the intention-to-treat population and safety population, respectively; both populations included all participants who underwent randomization and received at least one dose of study drug. The two populations were identical in this study.

- No mention of the non inferiority before this section !
- No mention of non inferiority in the abstract

that are resistant to NNRTIs.⁷⁻⁹ We designed Study ING114467 (SINGLE) to assess the safety and efficacy of dolutegravir at a dose of 50 mg plus a fixed-dose combination of abacavir-lamivudine, as compared with fixed-dose efavirenz-tenofovir DF-emtricitabine, which is the only single-tablet regimen currently preferred in the U.S. HIV treatment guidelines^{1,2} and one of two currently recommended single-tablet regimens in the European treatment guidelines.³

Example - 3

STATISTICAL ANALYSIS

According to the protocol, we could conclude that treatment with dolutegravir and abacavir-lamivudine was noninferior to treatment with efavirenz-tenofovir DF-emtricitabine if the lower boundary of a two-sided 95% confidence interval for the difference in the primary end point was less than 10 percentage points lower in the DTG-ABC-3TC group than in the EFV-TDF-FTC group. This margin is consistent with that in other trials in this population.¹⁹ Assuming a 75% response rate in the EFV-TDF-FTC group, we calculated that 394 participants who could be evaluated would need to be included in each group for the study to have 90% power to determine the noninferiority of dolutegravir and abacavir-lamivudine, at a one-sided significance level of 2.5%. Efficacy and safety analyses were performed in the intention-to-treat population and safety population, respectively; both populations included all participants who underwent randomization and received at least one dose of study drug. The two populations were identical in this study.

Non-inferiority limit
10% (4 to 12%)

Expected success rate
75% (EFV)

One-sided Type I error
2.5%

Power 90 %

Example - 4

EFFICACY

A rapid and sustained virologic response was observed, with 88% of the participants in the DTG–ABC–3TC group, as compared with 81% of those in the EFV–TDF–FTC group, having the primary end point of a plasma HIV-1 RNA level of less than 50 copies per milliliter at week 48 (Fig. 2A). The adjusted treatment difference between the two groups was 7 percentage points (95% confi-

DTG–ABC–3TC group and 81% in the EFV–TDF–FTC group had an HIV-1 RNA level of less than 50 copies per milliliter (Fig. 3). The adjusted treatment difference between the two groups was 9 percentage points (95% CI, 4 to 13), again supporting the superiority of the dolutegravir and abacavir–lamivudine regimen. If both the per-protocol and intention-to-treat analyses showed noninferiority, then testing for superiority was to be conducted as described above. No P values were derived for the test of superiority for the per-protocol population; the confidence interval was sufficient to conclude superiority because it excluded 0. Data from the per-protocol population are presented only to support the results in intention-to-treat population.

dence interval [CI], 2 to 12), with dolutegravir and abacavir–lamivudine meeting the noninferiority criterion. In addition, the dolutegravir and abacavir–lamivudine regimen was shown to be statistically superior to the efavirenz–tenofovir DF–emtricitabine regimen ($P=0.003$). Overall differences in response (intention-to-treat analysis) were due primarily to discontinuations because of adverse events (10 of 414 participants [2%] in the DTG–ABC–3TC group and 42 of 419 [10%] in the EFV–TDF–FTC group) (Table 1).

Similar results were observed in the per-protocol population, from which 2% of the participants were excluded (11 of 414 participants [3%] in the DTG–ABC–3TC group and 7 of 419 participants [2%] in the EFV–TDF–FTC group), owing to a number of reasons, including the use of prohibited medication (in <1% of participants). In this analysis, 90% of the participants in the

ITT 88% vs 81%, 7% (2-12)

PP (98% of ITT) 90% vs 81%, 9% (4-13)

Example - 5

RESULTS

A total of 833 participants received at least one dose of study drug. At week 48, the proportion of participants with an HIV-1 RNA level of less than 50 copies per milliliter was significantly higher in the DTG-ABC-3TC group than in the EFV-TDF-FTC group (88% vs. 81%, $P=0.003$), thus meeting the criterion for superiority. The DTG-ABC-3TC group had a shorter median time to viral suppression than did the EFV-TDF-FTC group (28 vs. 84 days, $P<0.001$), as well as greater increases in CD4+ T-cell count (267 vs. 208 per cubic millimeter, $P<0.001$). The proportion of participants who discontinued therapy owing to adverse events was lower in the DTG-ABC-3TC group than in the EFV-TDF-FTC group (2% vs. 10%); rash and neuropsychiatric events (including abnormal dreams, anxiety, dizziness, and somnolence) were significantly more common in the EFV-TDF-FTC group, whereas insomnia was reported more frequently in the DTG-ABC-3TC group. No participants in the DTG-ABC-3TC group had detectable antiviral resistance; one tenofovir DF-associated mutation and four efavirenz-associated mutations were detected in participants with virologic failure in the EFV-TDF-FTC group.

CONCLUSIONS

Dolutegravir plus abacavir-lamivudine had a better safety profile and was more effective through 48 weeks than the regimen with efavirenz-tenofovir DF-emtricitabine. (Funded by ViiV Healthcare; SINGLE ClinicalTrials.gov number, NCT01263015.)

Example - 6

- What would the gain be?
 - Better safety profile
 - Lower frequency of AE?
- No longer an issue if there is a clinically significant superiority
 - Is a 7% difference clinically important when the non-inferiority limit was set to 10%?
 - If we can neglect a 10% difference, a 7% difference is not of interest
- Response rate was 80% not 75%
 - Study was overpowered, increasing the likelihood of finding a difference and concluding on superiority

Phase 1

- Méthode : (modified) CRM vs 3+3 (standard)
- Avantages
 - Meilleure estimation de la dose maximale tolérée
 - Attribution plus rapide de la bonne dose
 - Moins de patients exposés à des doses toxiques (ou inefficaces)
- Limites, difficultés
 - Sensibilité plus grande aux erreurs de mesure de la toxicité (ex: cancéro)
 - Sensibilité à toxicité observée à la 1^{ère} dose
- Recommandation (TR Giens 2010) +++

Phase 1-2

- Avantages
 - Evaluation tolérance et efficacité
 - Moins de risque d'éliminer des doses intéressantes
- Limites
 - Recul insuffisant si toxicités tardives
- Recommandation (TR Giens 2010)
 - Risqué pour une molécule peu connue
 - Plutôt pour une monothérapie

Phase 2

- Essai d'efficacité précoce et exploratoire
 - Recherche de dose avec randomisation adaptative
 - Essais de priorisation: traditionnellement traités dans des essais séparés
- Avantages
 - Moins de patients **OU** exploration possible d'un plus grand nombre de doses (mm nb de patients)
 - Plus de patients traités à la bonne dose
- Limites
 - Logistique importante
 - Complexité
 - Non applicable quand le critère de réponse est tardif
- Recommandation (TR Giens 2010): ++
 - Pas de risque majeur identifié
 - **Attention néanmoins à la variabilité des ratios d'allocation! (2015)**

Phase 3: SSR

- Ré-évaluation du nombre de sujets nécessaire
- Deux « paradigmes »
 - En aveugle (des résultats d'efficacité)
 - Au vu des résultats d'efficacité
- Objectifs différents
 - Premier cas: ré-évaluer des paramètres de nuisance
 - Deuxième cas: plus compliqué