

COMMENT INCLURE PLUS INTELLIGEMENT DANS LES ESSAIS CLINIQUES ?

Vincent Lévy

Département de Recherche Clinique
Hôpital Avicenne, Bobigny
CRESS U1153, équipe ECSTRRA, Hôpital Saint
Louis, Paris



Pierre Morel

Service Hématologie Clinique Therapie
Cellulaire, CHU Amiens Picardie
Hematim, EA4666, CURS, Amiens

Clinical trials generate the evidence based for decision making in all areas of medicine and they can be especially important for patients with serious or life threatening health condition that have limited treatments options

US General Accounting Office 1999

- Essais sur clinicaltrial.gov
 - N = 433.207
 - 83.134 en cours de recrutement
 - **CLL : 315; AML : 758**
- Pubmed (11/11/22)
 - Barriers/Inclusion/Clinical Trial : N=641, dont 175 revues
 - Cancer : 132
 - Hematology : 89

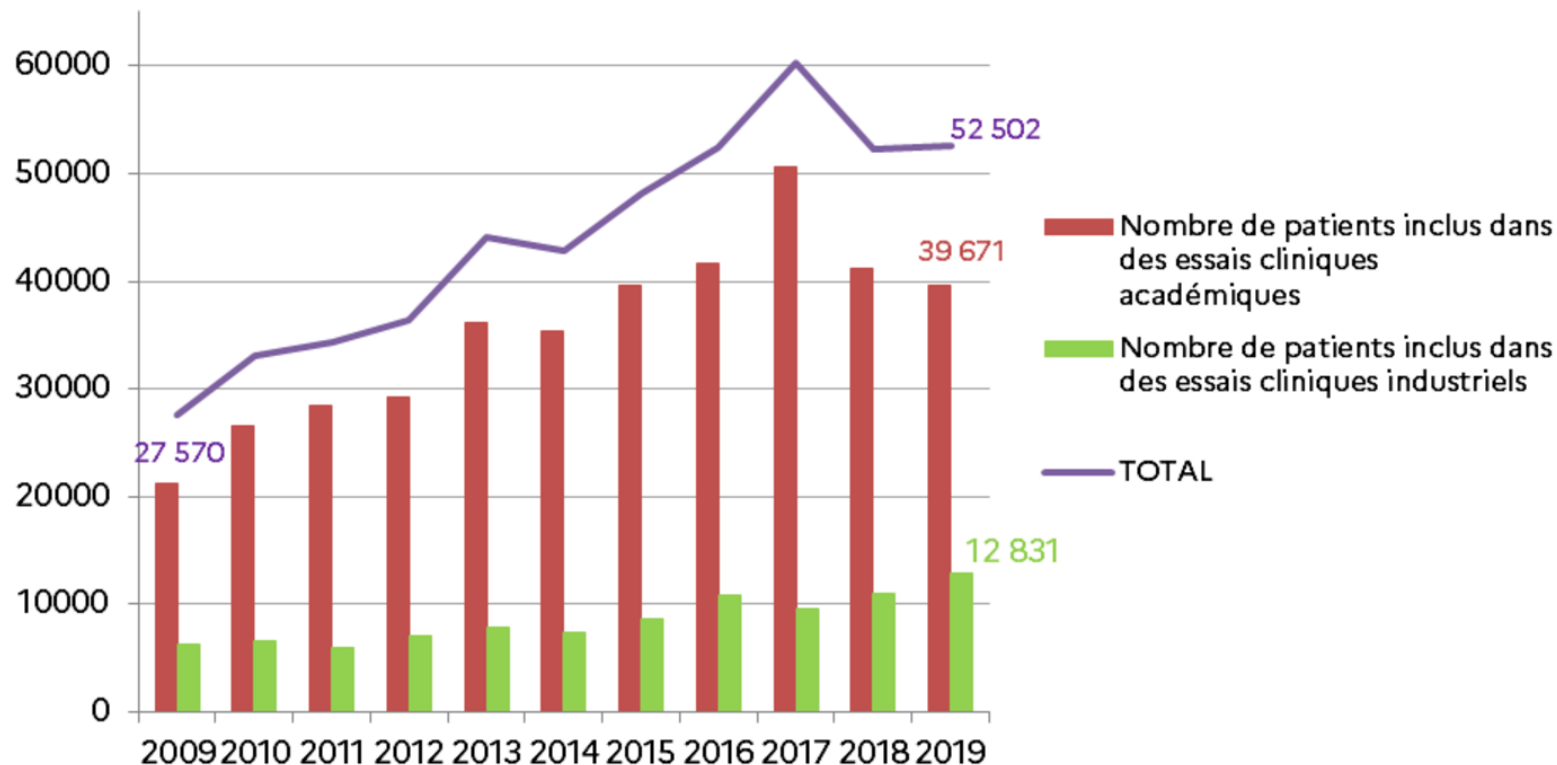
1: Tournoux C, Katsahian S, Chevret S, Levy V. Factors influencing inclusion of patients with malignancies in clinical trials. Cancer. 2006 Jan 15;106(2):258-70.

2: Rodríguez-Torres E, González-Pérez MM, Díaz-Pérez C. Barriers and facilitators to the participation of subjects in clinical trials: An overview of reviews. Contemp Clin Trials Commun. 2021 Aug 3;23:100829.

Introduction

Combien de patients en France ?

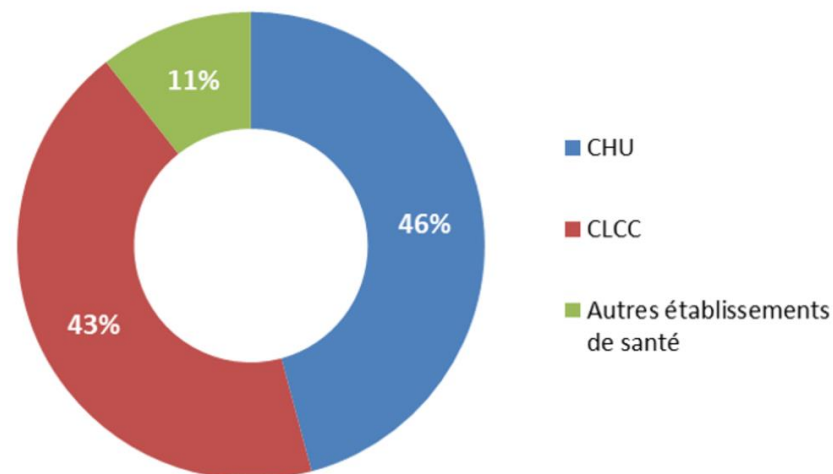
Évolution de l'inclusion des patients dans les essais cliniques en cancérologie entre 2009 et 2019



Introduction (2)

Etablissements et Hématologie

Distribution des inclusions dans les essais cliniques en cancérologie selon le type d'établissement en 2019



Nombre de patients inclus

Essais académiques : **4596**

Essais industriels : **1310**



: 60% des essais sont « thérapeutiques »

Les inclusions

- Revue générale (JAMA 2019) : participation en oncologie < 5% (mais spécificité de l'hématologie)
- Barrières liées aux pays, types d'hôpitaux, services, médecins, types de patients.
- Premier facteur : absence d'essai disponible
- Sur quel paramètre peut on jouer ?

Essentiellement les critères d'éligibilité

Dechartres A, Chevret S, Lambert J, Calvo F, Lévy V. Inclusion of patients with acute leukemia in clinical trials: a prospective multicenter survey of 1066 cases. *Ann Oncol.* 2011 Jan;22(1):224-233.

Bodicoat DH, Routen AC, Willis A, Ekezie W, Gillies C, Lawson C, Yates T, Zaccardi F, Davies MJ, Khunti K. Promoting inclusion in clinical trials-a rapid review of the literature and recommendations for action. *Trials.* 2021 Dec 4;22(1):880.

Denicoff AM, Ivy SP, Tamashiro TT, Zhao J, Worthington KH, Mooney MM, Little RF. Implementing Modernized Eligibility Criteria in US National Cancer Institute Clinical Trials. *J Natl Cancer Inst.* 2022 Nov 14;114(11):1437-1440

Les échantillons/la population

Population traitée

Suivi post AMM`

Accès précoce/ATU

Essai de phase III

Essai de
phase II

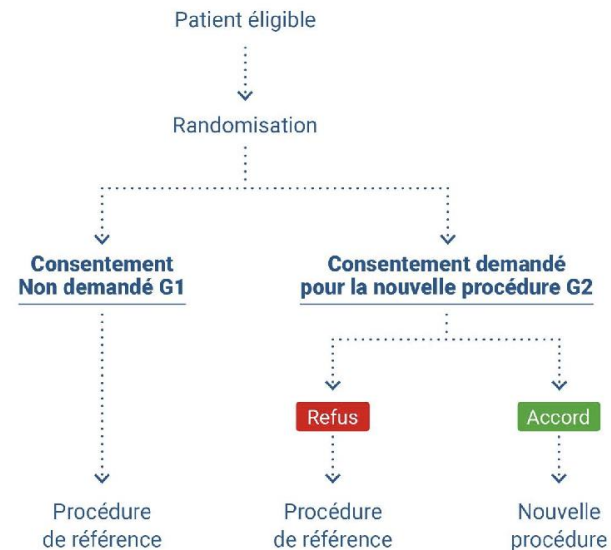
Les propositions (1)

- Disparition de ce certains critères de « routine » (ASCO 2017)
 - Présence de métastases cérébrales
 - Limite d'âge
 - HIV
 - Dysfonction d'organes (essentiellement foie et rein)
 - Antécédent de cancer
- Enquête 5499 pts, questionnaire sur **comorbidité** et proposition de participation à un essai (Jama oncol 2019)
- Résultats
 - 5 % ont ≥ 1 comorbidité, 1^{ière} = HTA
 - ≥ 1 comorbidité associé avec une diminution de la probabilité de discuter des essais, de proposer un essai et de participer à une essai
 - (44.1% vs 37.2%; OR, 0.86; 95% CI, 0.75-0.97; $P = .02$), (21.7% vs 15.7%; OR, 0.82; 95% CI, 0.70-0.96; $P = .02$) (11.3% vs 7.8%; OR, 0.76; 95% CI, 0.61-0.94; $P = .01$) respectivement

- KimES, BruinoogeSS, RobertsS, et al. Broadening eligibility criteria to make clinical trials more representative: American Society of Clinical Oncology and Friends of Cancer Research joint research statement. *J Clin Oncol*. 2017;35(33):3737- 3744.
- Unger JM, Hershman DL, Fleury ME, Vaidya R. Association of Patient Comorbid Conditions With Cancer Clinical Trial Participation. *JAMA Oncol*. 2019 Mar 1;5(3):326-333.

Les propositions « méthodologiques » (2)

- Randomisation déséquilibrée
 - Marketing...
 - Equipoise non respectée
 - Peu d'intérêt en termes d'effectif, voir délétère
 - ▼ de la puissance si rapport $\geq 3:1$
- Randomisation de Zelen
- Randomisation en cluster
 - Unité de randomisation ~~patient~~
 - Le centre/le service/l'hôpital
 - N ▲
 - Analyse + complexe



ET AU DELÀ DES ESSAIS CLINIQUES ?

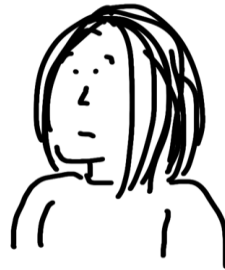
Tradition... EBM

Do you know about any RCTs that provide evidence that we should use RCTs?



RWE et niveau de preuve

Do you know about any **RWE** that provide evidence
that we should use **RWE**



When you start working on real-world
data



Faire coexister ETR et RWE ?

- FDA (2018)
 - *Real-World Data (RWD)* data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources”,
 - *Real-World Evidence (RWE)* as the “clinical evidence regarding the usage and potential benefits or risks of a medical product derived from analysis of RWD”
- Les essais hybrides
 - À la fois ETR et éléments pragmatiques
 - Randomisation puis certaines données sont collectés par CRF, d'autres par registres, entrepôts de données etc...

1. Zhu M, Sridhar S, Hollingsworth R, Chit A, Kimball T, Mummolo K, Greenberg M, Gurunathan S, Chen J. Hybrid clinical trials to generate real-world evidence: design considerations from a sponsor's perspective. *Contemp Clin Trials*. 2020 Jul;94:105856.

2. I. Ford, J. Norrie, Pragmatic trials, *N. Engl. J. Med.* 375 (2016) 454–463.

Essais hybrides

Comparison of randomized controlled trials, pragmatic trials and hybrid trials.

	Randomized controlled trials	Hybrid trials	Pragmatic trials
Randomization Blinding	Yes Generally yes	Yes Yes for outcome assessment (consider for other aspects of trial)	Generally not but occasionally randomized Generally no
Cost per participant Outcome ascertainment	High Collected through standardized case report forms (CRFs)	Lower Collected through all sources including CRFs, EHRs, registries or databases. Could consider a nested subset with collection of some outcome data through CRFs	Lowest Collected through routine medical sources; electronic health records (EHRs), registries, or other databases; outcome adjudication often required
Sample size	Depends on effect size, incidence of event & variability	Depends on effect size, incidence of event & variability. Usually requires larger sample size than RCT due to more heterogeneous population and higher variability	Depends on effect size, incidence of event & variability. Usually requires larger sample size than RCT due to more heterogeneous population and higher variability
Data quality (internal validity)	Strong, as data accrual is closely monitored. Endpoint is highly specific for treatment intervention. Collection of covariates controlled and standardized across trial sites	Generally lower than RCT but can be improved through outcome adjudication and partial standardized of EHRs across data sources and sites	Quality of data is heterogeneous across RW databases, consistency of capture of patient medical records is variable
Data relevance (external validity)	Limited generalizability beyond characteristics of patients in trial; relevance for routine medical practice is low;	Stronger external generalizability than traditional RCT	Captures routine medical practice & outcomes; generalizability to real-world patients high
Inclusion/exclusion criteria	Strict inclusion and exclusion criteria to enroll a more homogeneous study population.	Minimal inclusion and exclusion criteria so all participants with the disease or outcome of interest may be enrolled. May include a subset of participants for which a strict inclusion and exclusion criteria similar to those of RCT will be applied.	Minimal inclusion and exclusion criteria so all participants with the disease or outcome of interest may be enrolled.

Conclusion

- Vaste littérature sur les facteurs de non inclusion
- Sur quoi pouvons nous agir ?
 - Les critères
 - Mais surtout mettre à disposition des essais
 - Modifier la réglementation Française sur les essais pragmatiques
- Car il manque les comparaisons d'intérêt !

La suite avec le Pr Morel....

MENER UN ESSAI C'EST COMPARER

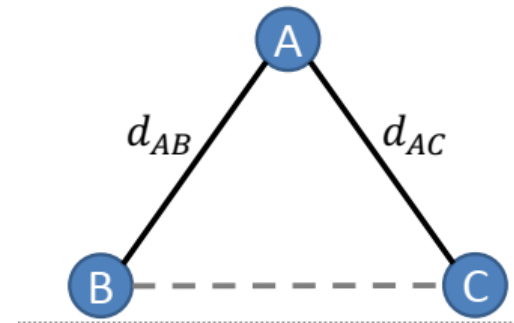


Comparer

- Pour rappel: Comparaison directe : essai randomisé
 - HR_{AB}
- Comparaisons indirectes
 - On souhaite comparer 2 traitements B et C
 - Ancré
 - B et C n'ont pas été étudiés lors d'un essai comparatif
 - mais par contre ont été comparés à un traitement identique A dans 2 essais AB et AC.
 - → Connaissant HR_{AB} et HR_{AC} inférer sur HR_{BC} :
 - $\Delta_{BC(P)} = \Delta_{AC(P)} - \Delta_{AB(P)}$
 - Non ancré
 - Absence d'essai randomisé et donc de comparateur commun: hypothèses plus fortes requises

Les comparaisons indirectes

- La comparaison indirecte « Standard »
 - $d_{BC} = d_{AC} - d_{AB}$
 - Biaisé s'il y a des déséquilibres dans les modificateurs d'effets (ME) entre AB et AC ; $d_{AB(AB)} \neq d_{AB(AC)}$
 - Suppose la stabilité des effets relatifs sur les 2 populations
- Les méthodes d'ajustement de population cherchent à ajuster
 - pour le déséquilibre
 - des facteurs pronostiques
 - Et des modificateurs d'effets (ME)*
 - L'hypothèse de stabilité des effets relatifs peut être mise en défaut
 - → Créer une comparaison équilibrée pour les facteurs confondants dans une population cible



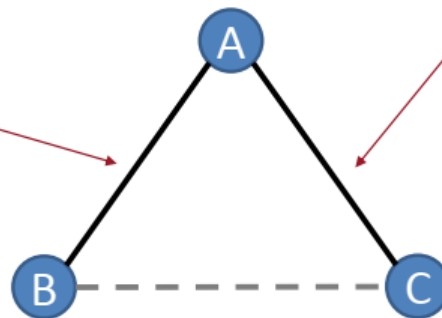
* **Modificateur d'effet (ME)**: modifie l'effet du traitement sur le critère de jugement sans forcément être pronostic
Peut dépendre de l'échelle utilisée

Comparaisons

- Scenario idéal: disposer de toutes les données individuelles des patients (DIP)
 - “Gold standard” – DIP meta-regression
- Scénario habituel: DIP limitées
 - Plusieurs methods récentes permettent l’utilisation de données mixtes:
 - Modéliser les interactions entre covariables des données agrégées

AB trial: IPD

Y_i	T_i	X_{1i}	X_{2i}	...



AC trial: aggregate data

$$\bar{Y}_A, \bar{Y}_C, \bar{X}_1, \bar{X}_2, \dots$$
$$\sigma_A, \sigma_C, f_X(\cdot)$$

Ajustement de la population

- Possibilité d'utiliser des données agrégées (A vs. C) avec des DIP (A vs. B)
- La comparaison se fait sur une échelle de prédiction linéaire → transformation: OR, HR
- MAIC et STC produisent des estimations de l'effet relatif du traitement qui sont spécifiques à la population AC
 - Il est peu probable que cela soit représentatif de la population cible de la décision (population traitée en routine)
 - Les estimations ajustées à cette population peuvent ne pas être pertinentes pour la décision dans la population cible...

Ajustement des populations: MAIC et STC

Matching-Adjusted Indirect Comparison	Simulated Treatment Comparison
Pondération des observations	Régression
Pondérer les individus de l'essai AB afin d'équilibrer la distribution des covariables et la rendre comparable avec celle de l'essai AC	1e étape; modélisation du devenir par régression sur l'essai AB
Estime le résultats sur A et B dans les conditions de l'essai AC grace à la pondération	Estime le résultat qu'aurait eu A et B dans l'essai AC en utilisant le modèle de régression obtenu sur AB avec les moyennes des variables explicatives observées sur AC
Vérifie la distribution des pondérations: taille effective de l'échantillon	Vérification habituelle du modèle AIC, résidus..

- Les populations AB et AC doivent avoir suffisamment de similitude (overlap)
- Comparer les distributions des covariables, les critères d'inclusion/exclusion
- Approches les plus populaires
- Disponibilité uniquement de la distribution marginale des covariables pour AC.

Hypothèses et propriétés :

La stabilité entre essais

	ancré	
Type de comparaison	Comparaison indirecte standard	Comparaison indirecte ajustée à la population, avec ancre
Hypothèse stabilité entre essais	Stabilité des effets relatifs	Stabilité conditionnelle des effets relatifs
	$d_{AB(AB)} = d_{AB(AC)}$	Predire $d_{AB(AC)}$ à partir de l'essai AB
Valide uniquement si...	Pas de déséquilibre de distribution des ME	Tous les ME sont connus et ajustés
Données	Nécessite uniquement des données agrégées	Nécessite des DIP pour au moins un essai

DIP: données individuelles patients

Recommandations NICE

- 1. Ancré +++++ (ou non ancré)
- 2. Justifier les comparaisons avec ancrage
 - **preuves que l'ajustement de population est susceptible de produire des estimations moins biaisées**
- 3. (Justifier les comparaisons sans ancrage)
- 4. Variables à prendre en compte pour l'ajustement
 - **Pour les comparaisons ancrées, type MAIC seul l'ajustement pour les ME est nécessaire pour minimiser le biais**
- 5. Échelle de comparaison
 - MAIC n'est pas "échelle-libre« : **échelle de prédiction linéaire**
- 6. Population cible
 - **La population cible de toute comparaison thérapeutique doit être explicitement énoncée**
 - **Si on dispose des DIP sur A vs B, MAIC va mimer l'effet qu'(aurait eu cette comparaison sur la pop de l'essai A vs C. Cette pop est elle un reflet exact de la pop cible??**
- 7. Qualité du rapport
 - **Des exigences strictes en matière de rapport sont recommandées, y compris l'évaluation des distributions des covariables, les preuves du statut de modificateur d'effet, la distribution des poids (le cas échéant) et les mesures appropriées de l'incertitude.**

exemple

Adv Ther (2017) 34:1650–1661
DOI 10.1007/s12325-017-0564-1



ORIGINAL RESEARCH

Comparative Efficacy of Ibrutinib Versus Obinutuzumab + Chlorambucil in First-Line Treatment of Chronic Lymphocytic Leukemia: A Matching-Adjusted Indirect Comparison

Suzy Van Sanden · Simona Baculea · Joris Diëls · Sarah Cote

A MAIC within a Bayesian framework [20, 21] was conducted using individual patient data (IPD) for the RESONATE-2 study of ibrutinib vs. chlorambucil [6], and the data published by Goede et al. [9] for comparison of obinutuzumab + chlorambucil vs. chlorambucil in the CLL11 study.

Step 1: Updating HRs for RESONATE-2

In a first step, the treatment effect of ibrutinib vs. chlorambucil is estimated using RESONATE-2 patient level data, simulating the HR that would have been obtained if RESONATE-2 had included a patient population identical to the CLL11 population. This is implemented by first excluding patients from RESONATE-2 that did not satisfy the CLL11 inclusion/exclusion

criteria: patients with a CIRS score of at most 6 and creatinine clearance of at least 70 mL/min, patients with creatinine clearance less than 30 mL/min, and patients with a diagnosis of small lymphocytic lymphoma (SLL) were excluded. To the remaining patients from

RESONATE-2, weights were assigned in such a way that the reweighted population resembles (or “matches”) the trial population from CLL11 as close as possible. The approach is very similar to propensity score weighting, whereby patients

For the base case analysis, all clinically relevant risk factors reported in [9] that may influence the relative treatment effect were included in the matching process: CIRS score, age, Binet stage, β_2 -microglobulin, del11q, ECOG performance status, creatinine clearance, gender, and having unmutated IGHV.



Table 1 Baseline characteristics of the CLL11 and RESONATE-2 trial populations before and after matching for base case and sensitivity analysis for the removal of IGHV status from matching

	Trial data		MAIC results		
	CLL11 [9]	RESONATE-2	RESONATE-2		
	ITT	ITT	After exclusion ^a	Base case matched (n = 13 variables)	SA matched (n = 12 variables)
N (Neff)	589	269	191	115 (35)	152 (48)
CIRS score (median)	8	5	6	8	8
CIRS score ≤6 (%)	26	64	56	26	26
Age (median)	73	73	73	73	73
Age (≥75 years) (%)	43	35	40	43	43
Binet stage A (%)	22	19	19	22	22
Binet stage B (%)	42	43	40	42	42
Binet stage C (%)	36	38	41	36	36
β2-Microglobulin ≥3.5 mg/L (%)	35	71	77	35	35
del11q (%)	17	22	24	17	17
ECOG (median)	1	1	1	1	1
Creatinine clearance (median, mL/min)	62	61.21	55.35	62.02	62.02
Male (%)	62	63	59	62	62
Unmutated IGHV (%)	61	59	58	61	57 ^b

CIRS Cumulative Illness Rating Scale, ECOG Eastern Cooperative Oncology Group, IGHV immunoglobulin heavy variable cluster, ITT intention to treat population, N number of patients analyzed, n number of variables matched, Neff effective sample size, SA sensitivity analysis

^a Excludes patients from RESONATE-2 with CIRS ≤6 and creatinine clearance ≥70 mL/min, patients with creatinine clearance <30 mL/min, and SLL patients

^b Baseline characteristic that was not matched and hence value represents the average for the characteristic unadjusted

Résultats comparaison « standard » et MAIC

Trial data	CLL11	N (Neff)	Treatment vs. chlorambucil	HR [95% CI] P value PFS by IRC assessment
		589	Obinutuzumab + chlorambucil	0.19 [0.14, 0.27] <0.0001
	RESONATE-2 (ITT)	269	Ibrutinib	0.161 [0.091, 0.283] <0.0001
MAIC results	RESONATE-2 (after exclusion) ^a	191	Ibrutinib	0.095 [0.044, 0.205] <0.0001
	RESONATE-2 (matched n = 13)	115 (35)	Ibrutinib	0.045 [0.008, 0.245] 0.0003

COMPARAISONS INDIRECTES

Un essai peut permettre plusieurs comparaisons?

Si le bras de référence est judicieusement choisi

Mais cette information reste de moindre qualité que l'essai randomisé

En tous cas, Danger des comparaisons directes