

Essais de recherche de dose: la cancérologie face au reste du monde?

Xavier Paoletti

Institut Curie,

Sce de Biostatistique / Inserm U900



institutCurie

Ensemble, prenons le cancer de vitesse.

Des cytotoxiques vers les agents ciblés: un saut méthodologique?

1. Essais précoces : un carcan étouffant?
2. Essais précoces pour les traitements ciblés: une hotte à questions
3. Méthodes statistiques et biomarqueurs
4. Et de l'autre côté du miroir de la cancérologie...
5. Essais randomisés: quand la cancérologie ne sera plus l'exception

Contexte: Essais de recherche de dose

- Soit d_1, \dots, d_K , K niveaux de dose croissants
- Quelle est la dose optimale?
 - ➔ Toujours un compromis entre toxicité et l'activité
- Phase de développement:
 - en cancérologie: Phase I (dose finding)
 - pour les autres disciplines: phase II (dose ranging)

Contexte: Essais de recherche de dose

- En cancérologie:
 - Identifier la dose la plus haute possible (Dose Maximum Tolérée)
 - avec les biothérapie, tentatives pour changer de paradigme
Mais quelle définition?
- hors cancérologie:
 - dose la plus efficace parmi un ensemble de doses
- Dans les 2 cas: critiqué
 - " An inadequate dose-regimen selection is believed to be one of the key drivers of the high attrition rate in Phase III trials" (Kola 2004)
 - " The limited precision of phase I trials " (Christian 1994)

Contexte: Essais de recherche de dose

Principales méthodes de recherche de dose

Cancérologie:

- méthodes algorithmiques (3+3)
- méthodes utilisant un modèle (CRM)
- petits effectifs ($n < 50$)
- essais séquentiels, pas de tests

Hors cancérologie

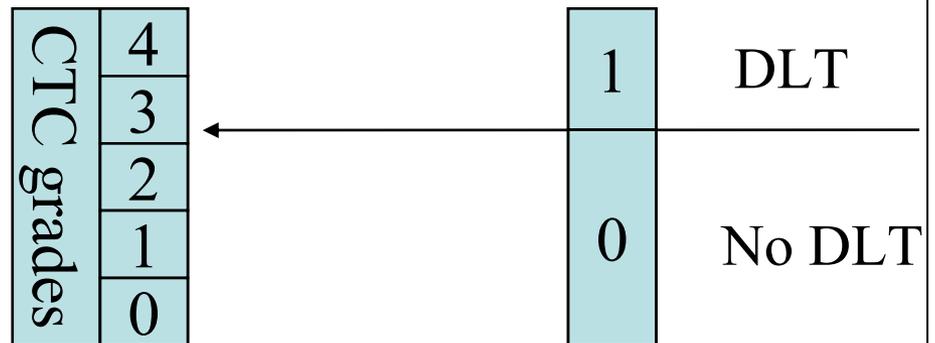
- tests de comparaison (gate keepers)
- utilisation d'un contrôle
- n entre (100, 500)
- bras parallèles, randomisation

Essais précoces en cancérologie: un carcan étouffant?

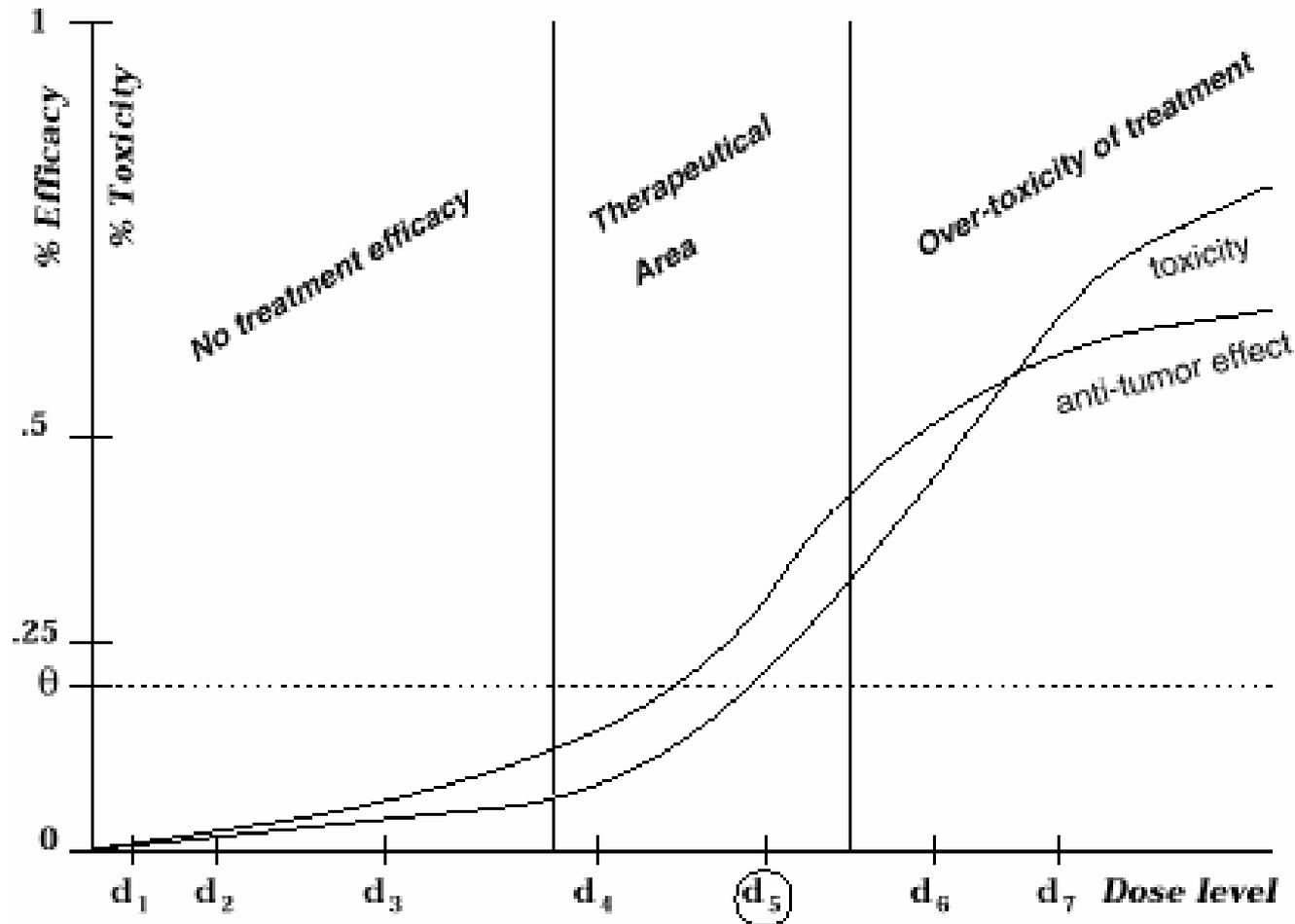
Hypothèse fondamentale: plus la dose est élevée, plus les chances de succès sont grandes

Phase I

- Objectifs
 - Identifier la Dose Maximale Tolérée
- Critères d'évaluation:
 - **La DLT au 1^{er} cycle...**
 - la PK
 - L'activité?



Essais précoces en cancérologie: un carcan étouffant?



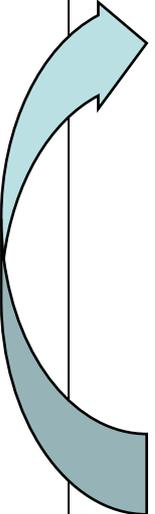
Essais précoces en cancérologie : un carcan étouffant?

- Contraintes
 - **Escalade de dose**: un palier de dose est administré si le précédent est jugé sûr
 - Ne pas sous-traiter les malades (pas de bénéfice escompté)
 - Ne pas sur-traiter les malades (risque de tox.)

Essais précoces : un carcan étouffant?

- Plans d'expérience
 - La méthode « standard »
 - 3 pts traités à une dose
 - si 0 DLT parmi 3 → escalader
 - si 1 DLT parmi 3 → inclure 3 autres patients
 - si 2 DLT parmi 3 ou 6 → désescalader
 - Les méthodes avec modèle (*Continual Reassessment Method O'Quigley et al 1990*)

CRM: principe

- Administrer à chaque patient la dose qui correspond à la meilleure estimation de la DMT.
 - Guidé par un modèle:
 - Ajuster un modèle de la relation dose-toxicité sur toutes les données recueillies;
 - Estimer les probabilités de DLT à chaque dose;
 - ⇒ Prochaine dose = dose dont l'estimation est la plus proche de la cible (20%) (meilleure estimation courante de la DMT).
 - Traiter le prochain patient à cette dose.
- 

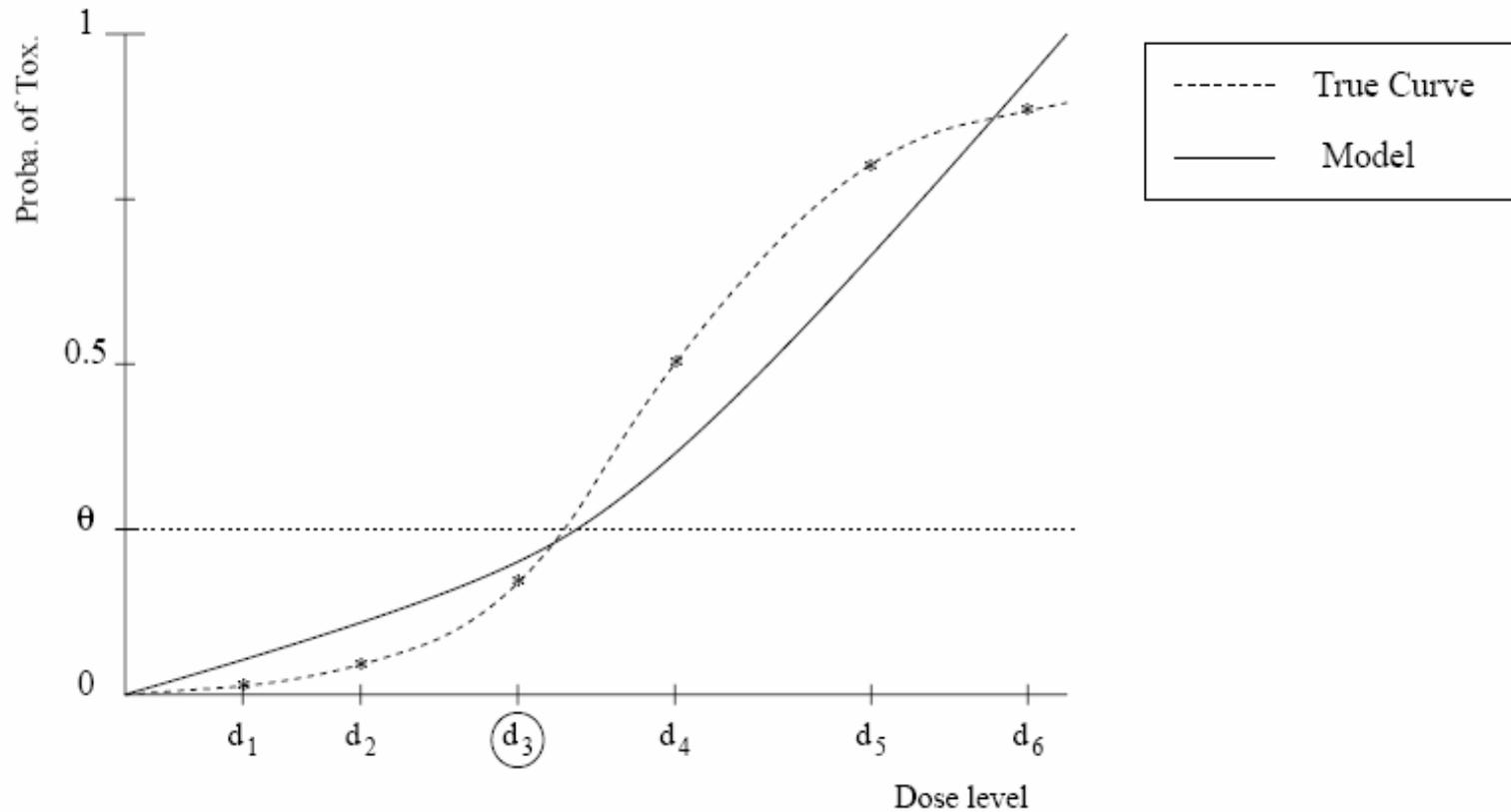
Aspects techniques

- Modèle **sous-paramétré**: $Pr(Y=1 | X=x)=\varphi(x,a)$
 - 1 parametre: a
 - Fonction monotone
 - Plusieurs candidats: puissance (x^a), logistic, probit etc.
Modèle x^a conduit à de bonnes performances (Paoletti and Kramar SiM 2009)
 - Adéquation locale (**modèle de travail**)
- Estimation:
Après chaque observation, mise-à-jour de l'estimation de a (et de $\varphi(d_k, \hat{a})$)
 - Inférence bayesian → spécifier la distribution a priori
 $f_{j+1}(a)=L(a) f_0(a)$
 - Inférence par la vraisemblance (O'Quigley and Shen 1996)

Propriétés statistiques

- Convergence asymptotique même si mauvais spécification du modèle (Shen and O'Quigley 1996):
Quand $n \rightarrow \infty$
 - La dose recommandée converge vers la vraie DMT
 - L'estimation de la probabilité à la dose recommandée converge vers la vraie probabilité.
- Intervalles de confiance uniquement à la dose recommandée.
 - Cornish-Fisher approximation (O'Quigley 1992)

Adequation locale: 1 paramètre est assez



DLT: Un critère peu informatif

Quelle est la performance des méthodes de recherche de dose?

- Proposition d'une méthode de référence (dite optimale) dans le cadre de simulation
 - utilise une information « complète », connaissant la relations dose-tox.
 - comme si chaque patient pouvait être traité à chaque dose
 - estime le risque de toxicité et permet d'identifier la DMT
- Donne les meilleures performances
 - pour une taille d'échantillon
 - pour une relation dose-tox
 - dans un contexte non paramétrique

DLT: Un critère peu informatif

- Et ce n'est pas brillant...

d_k	d_1	d_2	d_3	d_4	d_5	d_6
R_k	0.05	0.11	0.22	0.35	0.45	0.60
$p_k^{\text{CRM}}(16)$	0.05	0.26	0.42	0.21	0.06	0.0
$q_k^{\text{Opt}}(16)$	0.04	0.27	0.48	0.17	0.04	0.0

Essais précoces pour les thérapies ciblées: une hotte à questions

- Quelle est la dose optimale?
Dose optimale \neq Dose Maximale Tolérée?

Toxicité n'est plus le seul critère de jugement!

- Comment la définir
 - quel critère de jugement utiliser ?
 - quand et « où » le mesurer ?
 - sur quelle population?
- Quelle combinaison de traitements?

Utilisation de critères d'activité et toxicité

- Depuis 1998, de nombreuses propositions
 - Thall et Russel (1998) suivi de Thall and Cook (2004)
 - O'Quigley et al (2001)
 - Braun et al. (2002)
 - Zhang et al. (2006)
 - Houédé et al. (2009)
 - ...
- Diffèrent essentiellement par le choix du modèle
 - Pour trouver une dose qui satisfasse
 - activité (réduction de la tumeur > 30% O/N)
 - et toxicité (DLT O/N)

Utilisation de critères d'activité et toxicité

3 propositions:

- **Critère trinaire** (Thall and Russel, Zhang et al, O'Quigley et al)

toxicité

activité	0	1
	1	1

- *Succès* si activité et pas de toxicité
- *Toxicité* si toxicité
- *Non réponse* si ni toxicité ni activité

➔ 4-parameter **continuation ratio model**

- **Critère binaire bivarié** (Thall and Cook)

➔ odds ratio model avec 6 paramètres

- **Critère trinaire bivarié** (Houédé et al.)

➔ Modèle marginal d'Aranda Ordaz généralisé + copula

Réponse clinique pire que la DLT?

Horstmann (NEJM 04), Roberts (JAMA 05), Italiano (Annals 08)

Sur la période 1991-2002, et 2003-2006

- **Taux de réponse clinique** (partielle ou complète)
Varie si agent seul ou en combinaison
 - 5% pour les agents cytotoxiques en monothérapie
 - 3-4% pour les traitements ciblés en monothérapie
 - >10% si en combinaison avec traitement enregistré
- Longues stabilisations (contrôle de la maladie)
 - En augmentation: 35% à 45% des patients
- Morts toxiques en baisse: <0.05%

Essais précoces pour les traitements ciblés: une hotte à questions

- Critères de jugement alternatifs

- La pharmaco-cinétique (PK)/PD (aire sous la courbe, pic de concentration, % d'inhibition etc.)
- Des critères cliniques
 - comme la vitesse de la progression tumorale
 - La durée avant progression
- Des biomarqueurs

→ Phase I ressemble de plus en plus à une phase II

→ 1ère réaction: augmenter la taille de l'échantillon à la dose recommandée (+5 à +10 pts)

Quels critères de jugement en pratique

Parulekar et Eisenhauer JNCI 2004

Recommended Phase II dose	Primary basis for recommendation	No of trials	No of agents
Not stated		6	6
Not recommended		2	1
Recommended	Toxicity	35	19
	PK	11	7
	Other trial results (toxicity)	2	2
	Clinical activity	1	1
	PBMC finding	1	1
	Effect in tumour	1	1
	Convenient dosing schedule	1	1

Quels critères de jugement en pratique

Parulekar et Eisenhauer JNCI 2004

- Après la toxicité, **PK est le critère le plus utilisé**
- Niveau plasmatique assez courant après sélection sur la tox
 - présuppose une association entre le niveau d'exposition et l'inhibition de la cible
- Pourtant, mesure de l'inhibition de la cible rarement invoquée:
 - renforce une décision plutôt que guide la sélection

Surprenant considérant le poids des investigations réalisées pour des études dites « preuves de concept »

Utilisation de la PK/PD

Iressa versus Tarceva

Iressa et Tarceva 2 inhibiteurs de tyrosine kinase

- **Iressa**

- **DMT \geq 700 mg/j,**
 - Mais à 150 mg/j
 - concentration plasmatique donne une activité pharmaco
 - preuve d'un effet biologique
 - Les doses 250 et 500 mg/j sont comparées chez les patients avec NSCLC avancé en phase II
 - Pas de différence d'activité
- 250 mg/j recommandée

Enregistrement puis restriction et peu d'utilisation en routine

Utilisation de la PK/PD

Iressa versus Tarceva

Iressa et Tarceva 2 inhibiteurs de tyrosine kinase

- **Tarceva**

- **DMT ≥ 150 mg/j** (au dessus diarrhées et tox cutanées),
- A 150 mg/j
 - PK: Variabilité intra et inter-pt
relation indépendante entre la dose et la PK
relation dose-activité
 - concentration plasmatique donne une activité pharmaco
 - Longues stabilisations (tête et cou et poumon)
- Phase II de screening
- ➔ 150 mg/jour recommandée

Enregistrement, utilisation et efficacité

PHARMACOKINETICS and TOXICITY

	Gefitinib	Erlotinib
Optimal biologic dose	>150mg	100mg
MTD	700mg	150mg
AUC for MTD	36.08	38.42
Affinity EGFR IC ₅₀	5nM	2nM
Dose for phase II	250mg	150mg
Rash frequency	37%	76%
Diarrhoea frequency	27%	55%

Utilisation de biomarqueurs: un avenir prometteur

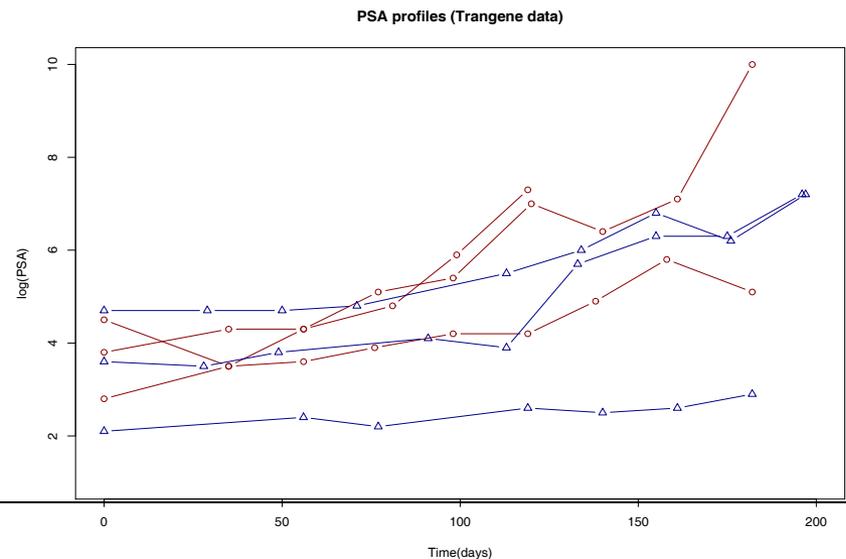
Proposition de plans d'expérience

- Chercher la dose
 - qui donne **la meilleure réponse** (inhibition, taille tumorale)
 - sous contrainte d'une toxicité acceptable (<30%)
- Cibler la DMT **et** la dose « optimale » (meilleure réponse observée)
- **Comparer les doses dans un schéma d'escalade**
 - Randomiser les paliers de dose 2 par 2
 - **Utiliser l'évolution du marqueur** au cours du temps
 - ➔ prendre en compte l'ensemble du suivi
 - ➔ utiliser des **variables continues**

Utilisation de biomarqueurs: un avenir prometteur

Etude d'un vaccin ds le cancer de la prostate (Transgene)

- Comparaison de plusieurs schémas d'administration
- Utilisation de l'évolution du PSA au cours du temps
- Mesure de la toxicité à chaque cycle
- Randomisation des patients



Utilisation de biomarqueurs: un avenir prometteur

Etude par simulations (AL Vataire) montre que:

- Puissance (taux de bonne identification) de la méthode augmente avec données répétées et continues
- Risque de toxicité grave est contrôlé
- Permet de trouver la meilleure dose et non de définir si activité suffisante
- Ne nécessite pas que le biomarqueur soit un « surrogate » mais qu'il permette de classer les doses
- La randomisation
 - élimine les biais de recrutement
 - fournit des données sur une « zone »
 - ➔ relation dose-effet
- Peut être utilisé dès que l'activité est mesurée par rapport à une valeur de départ

Utilisation de biomarqueurs

- Pas de critère d'activité « universel »
 - ➔ importance de la pré-clinique et de la compréhension des mécanismes d'action
 - ➔ rôle de la phase 0?
- Quel est le rôle de la phase I?
 - Impossible d'identifier le meilleur marqueur et de l'utiliser pour le plan d'expérience

Et pendant ce temps, de l'autre côté de la *Spécialité*?

- La recherche de dose optimale repose
 - Sur le développement de **gatekeepers**
Maurer, Hothorn, Lehmacher (1995)
 - **procédure de tests multiples** de type fermée (Bonferoni and Simes; Benjamini and Hochberg ; Dmitrienko etc.)
 - **procédure multi-étape**
 - Contrôle le risque de 1^{ère} espèce
 - Sur la **modélisation de la relation dose-effet**
 - estimer la Dose Minimale Active, la Dose Maximale Tolérée etc.
 - **Niveau de preuve plus élevé si relation dose-effet**

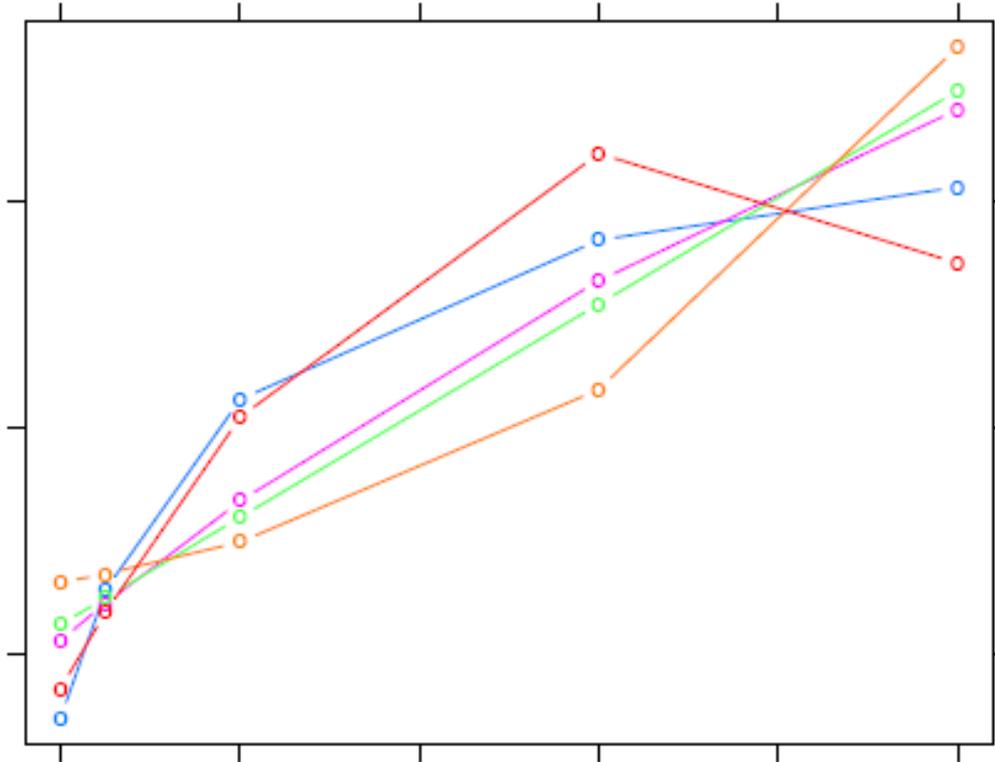
Approche par modélisation

Candidats types:

- modèle E_{\max} : $y = E_0 + \frac{E_{\max} - E_0}{1 + EC_{50} / C_e}$ où E_0 est la réponse pour une concentration nulle, E_{\max} est la réponse maximale et EC_{50} est la concentration conduisant à une « demi » réponse.
- Modèle quadratique
- Log-linéaire
- Exponentiel
- ...

Approche par modélisation: utilisation des contrastes

Exemple: Soit 4 doses et 1 placebo testés avec 250 patients randomisés en groupes parallèles. Les différents modèles sont fortement différents...



Approche par modélisation: utilisation des contrastes

Bretz, Pinheiro et Branson (2005)

- Combiner une approche de comparaison multiple et de modélisation pour
 - Bénéficier du contrôle d'erreur des techniques de comparaison multiples
 - De la flexibilité des modèles
- Soit un modèle $Y_{ij} = f(d_i, \theta) + \varepsilon_{ij}$, avec $\varepsilon_{ij} : N(0, \sigma^2)$ ^{ind}
- Chaque modèle m est associé avec $\mu_m = f_m(d, \theta)$, le vecteur des moyennes à chacune des doses
- Parmi un ensemble de modèles candidats, on teste ceux qui sont statistiquement différents de 0.

$$H_0^m: c_m' \mu_m = 0 \text{ vs } H_1: c_m' \mu_m > 0$$

Quelle place dévolue au modèle

- Reflet de la réalité ?
 - Trouver le meilleur modèle
- Outil de travail (modèle sous-paramétré type CRM) ?
- A quoi doit-il servir ?
 - Estimer des probabilités
 - Identifier une tendance
 - Faire des prédictions
- Modestement, à classer des doses entre elles à partir d'un critère sensible

Et si la cancérologie rejoignait la médecine interne?

- Prendre en compte l'activité pour trouver la dose optimale
→ ne plus se limiter à DLT O/N.
- Ne plus se satisfaire de 6 patients
- Statistiquement impossible de trouver la dose optimale avec 20 pts
→ augmenter la taille des échantillons?
→ randomiser entre plusieurs doses candidates?
- Importance de la randomisation car
 - pas de seuil d'activité validé (biomarqueurs, concentration plasma. etc.)
 - hétérogénéité dans la population
 - Recherche de longues stabilisations
- Se rapprocher des plans d'expérience ds les autres pathologies en conservant l'idée d'un *modèle de travail*.
 - Faire des phase I de faisabilité et de safety, sélection de plusieurs doses
 - Faire des phase II de recherche de dose