

# Quand le délai devient une propriété du traitement

## Framework d'optimisation thérapeutique dynamique

### Thèse

Pour les interventions dont la cinétique de production rivalise avec la cinétique de la maladie, **le traitement ayant la meilleure efficacité moyenne n'est pas nécessairement celui qui maximise l'utilité attendue pour un patient donné.**

L'utilité dépend de **l'état anticipé du patient au moment où ce traitement pourra réellement être administré.**

ZUMA-7, TRANSFORM et BELINDA offrent un cas d'école. Leurs divergences d'efficacité ne résultent pas d'une source unique (biologie du CAR, domaine CD28 vs 4-1BB, population). Elles reflètent aussi **comment chaque stratégie délivre le traitement au patient, et dans quel état clinique ce patient se trouve à ce moment.**

## 1. Les trois essais et l'énigme partagée

### Contexte commun

Les trois essais adressent une **question stratégique identique** en deuxième ligne de traitement : chez les patients atteints de LBCL (lymphome B diffus agressif) primaire réfractaire ou rechute précoce ( $\leq 12$  mois), **substituer un CAR-T anti-CD19 au parcours classique** (chimiothérapie de rattrapage suivie d'autogreffe chez les répondants).

L'enjeu : cette population a **démontré sa résistance chimio**. Le parcours classique requiert une réponse d'abord. Le CAR-T contourne cette barrière.

## Résultats cliniques observés

| Essai     | CAR-T    | N total | N CAR-T arm | CR (%)     | Mesure principale                              | Résultat  |
|-----------|----------|---------|-------------|------------|------------------------------------------------|-----------|
| ZUMA-7    | Axi-cel  | 359     | 180         | NR         | EFS 2-year 41% vs 16% (HR 0.40)                | ✓ Positif |
| TRANSFORM | Liso-cel | 184     | 92          | 74% vs 43% | vs EFS médian 29.5 mo vs 2.4 mo (HR 0.375)     | ✓ Positif |
| BELINDA   | Tisa-cel | 322     | 173         | 46% vs 43% | vs EFS médian 3.0 mo vs 3.0 mo (HR 1.07, p=NS) | ✗ Négatif |

## L'énigme

Liso-cel et tisa-cel partagent **le même domaine de costimulation (4-1BB)**. L'un réussit (TRANSFORM). L'autre échoue (BELINDA).

Si la biologie du CAR était le déterminant unique, **TRANSFORM serait difficile à expliquer**. Or TRANSFORM réussit.

Où chercher? **Pas une source unique. Les trois contribuent** : propriétés du produit, caractéristiques de la population, architecture du système qui délivre le traitement au patient.

## 2. Trois objets centraux — Architecture théorique

### Objet 1: État clinique dynamique du patient

$$S_i(t) = \text{état multidimensionnel du patient } i \text{ au temps } t$$

Cet état inclut : charge tumorale, vitesse de croissance, ECOG, LDH, inflammation systémique, fonction organique, fitness immunitaire.

**L'état n'est pas binaire.** C'est une trajectoire continue qui se dégrade (ou s'améliore) selon la maladie, le bridging, les comorbidités.

## Objet 2: Délai de délivrance comme distribution

$$\tau_{ij} \sim F_{ij}(\tau) = \text{distribution du délai entre décision et administration effective}$$

Pas un délai unique. Une distribution :

$$\mathbb{E}[\tau_{ij}], \text{Var}(\tau_{ij}), P_{90}(\tau_{ij})$$

Cette distribution capture :

- **Manufacturing** : designs du CAR, vein-to-vein time
- **Patient** : bridging permis, comorbidités induisant des retards
- **Système** : capacité centre, slots aphérèse, logistics, qualifications
- **Aléatoire** : infections, complications administratives, events imprévisibles

**La variance et la tail latency important autant que la médiane.**

## Objet 3: Fonction de réponse du traitement

$$R_j(S) = \text{réponse/utilité attendue du traitement } j \text{ dans l'état clinique } S$$

Chaque traitement possède une **fonction de performance continue** qui dépend de l'état du patient.

**Exemples :**

- CAR-T autologue : réponse maximale si ECOG  $\leq 1$ , burden modérée, LDH  $< 500$ . Réponse diminue progressivement si ECOG augmente ou état se dégrade.
- Allogénique : réponse conditionnée par tolérance du rejet, capacité immunosuppression.
- Bispécifiques : réponse moins dépendante de l'état initial mais soumise à autres contraintes (persistance pharmacocinétique, microenvironnement).

**Pas de frontière nette** (in/out). Une dégradation progressive.

Possibilité de définir un seuil d'utilité :

$$\Omega_j(u^*) = \{S: R_j(S) \geq u^*\}$$

Mais cette région **dépend du seuil choisi** et n'est pas une propriété biologique absolue du traitement.

### 3. Valeur attendue et décision

Combiner les trois objets :

$$V_{ij} = \mathbb{E}_{\tau, S} [R_j(S_i(t_0 + \tau))]$$

**Traduction** : Utilité attendue du traitement  $jj$  pour le patient  $ii$  quand on tient compte :

- De la distribution du délai  $\tau \sim F_{ij} \tau \sim F_{ij}$
- De la trajectoire incertaine  $S_i(t)$
- De la fonction de réponse  $R_j(S)$  du traitement dans cet état

La décision devient :

$$j^* = \arg \max_j V_{ij}$$

**Le meilleur traitement pour ce patient à cet instant est celui qui maximise utilité attendue, intégrant timing et état anticipé.**

### 4. Implications conceptuelles

La fenêtre thérapeutique dépend du couple (patient, intervention)

Pas de «  $T_W$  patient » universel. Plutôt : chaque traitement impose une relation différente entre délai et état.

Un patient peut être :

- Hors du domaine utile du CAR-T autologue (délai trop long, état trop dégradé)
- Mais encore dans le domaine du bispécifique (disponibilité immédiate)
- Ou du CAR-T allogénique (délai modéré, domaine de tolérance différent)

Le choix optimal exploite ces décalages.

Les étapes d'une intervention time-constrained

**Étape 1** : Patient trajectory  $S_i(t)$  commence sa dégradation dès la rechute/ « réfractarité »,

**Étape 2** : Chaque traitement  $jj$  introduit un délai  $\tau_{ij}$  avant d'être administrable,

**Étape 3 :** Au moment  $t_0 + \tau_{ij}$ , le patient se trouve dans un état  $S_i(t_0 + \tau_{ij})$ ,

**Étape 4 :** Ce traitement  $jj$  possède une utilité  $R_j(S_i(t_0 + \tau_{ij}))$ ,

**Étape 5 :** Intégrer la distribution de  $\tau$  donne l'utilité attendue  $V_{ij}$ .

## 5. Décomposition des déterminants du délai

Le délai observé n'est pas une simple addition. C'est une fonction complexe :

$$\tau_{\text{observe},ij} = g(\tau_{\text{stratégie}}, \tau_{\text{patient}}, \tau_{\text{systeme}}, \tau_{\text{stochastique}})$$

où :

- $\tau_{\text{stratégie}}$  : délai intrinsèque au produit/protocole (design, manufacturing, release)
- $\tau_{\text{patient}}$  : délai induit par agressivité maladie, fitness, bridging endogène
- $\tau_{\text{systeme}}$  : délai imposé par capacité centre, slots, logistics
- $\tau_{\text{stochastique}}$  : événements aléatoires (infections, blocages administratifs)

**Les composantes sont simultanées et interactives.** L'infection peut survenir during manufacturing sans ajouter linéairement. Le bridging peut être causé par manufacturing delay.

### Implication pour l'estimand

- **Pour un treatment-policy estimand** : le délai observé fait partie de l'effet de la stratégie assignée, quel que soit son déterminant. On ne l'« ajuste » pas en le neutralisant statistiquement.
- **Pour une analyse causale comparant architectures** : cette décomposition devient indispensable, car une fraction du délai reflète des caractéristiques du patient, non du design thérapeutique.
- **La bonne doctrine** : ne pas ajuster naïvement une variable post-randomisation sans avoir d'abord précisément défini l'estimand cible et le DAG causal.

## 6. Réinterprétation des trois essais

### ZUMA-7 : Résultat positif

#### Axi-cel (CD28) :

- Manufacturing time : tolérable
- Fonction de réponse  $R_{axi}(S)$  : large domaine d'état utile (tolère ECOG 2, burden modérée)
- Délai distribution : médiane acceptable pour majorité des patients
- **Conclusion** : majorité des patients recrutés ont possédé état  $S$  compatible avec  $R_{axi}$  au moment infusion. Efficacité observée reflète succès temporel + réponse biologique.

### TRANSFORM : Résultat positif

#### Liso-cel (4-1BB, architecture CD4/CD8 contrôlée) :

- Manufacturing time : court
- Bridging : 63% des patients
- Fonction réponse  $R_{liso}(S)$  : domaine comparable à axi-cel
- **Conclusion** : résultat positif valide que 4-1BB n'est pas intrinsèquement inférieur. Architecture manufacturing courte + sélection bridging favorable = majorité des patients encore dans domaine utile.

**Important** : ne réfute **pas** que CD28 vs 4-1BB soit biologiquement équivalent. Simplement : TRANSFORM démontre que si la biologie est différente, l'architecture temporelle du système peut compenser.

### BELINDA : Résultat négatif

#### Tisa-cel (4-1BB, sans contrôle CD4) :

- Manufacturing time : **très long** (52 jours)
- Bridging permissif : 83-97% des patients
- Population à aphérèse : très agressivement sélectionnée (ultime salvage)
- Fonction réponse  $R_{tisa}(S)$  : domaine comparable biologiquement à liso-cel
- **Problème critique** : délai très long + population ultra-malade = quand patient arrive à infusion (jour 52+), état  $S_i(t_0 + 52j)$  est souvent **en dehors du domaine** où  $R_{tisa}$  **conserve utilité**.

**Résultat EFS=3.0 mo vs 3.0 mo SOC** n'infirmes **pas** la biologie 4-1BB.

Il démontre qu'un long délai combiné à une population très agressivement rapide peut faire que **le patient quitte le domaine de performance utile** avant l'administration, indépendamment du design CD28 vs 4-1BB.

## 7. Vocabulaire central (quatre termes seulement)

**Potency** : propriété du produit. Capacité intrinsèque à reconnaître antigène, s'activer, proliférer.

**Efficacy** : effet observé sous conditions de l'essai. Ce qu'on mesure dans le protocole clinique.

**Effectiveness** : effet attendu dans contexte réel patient-système. Intègre délai, état réel du patient, constraints système réel.

**Utility** : effectiveness pondérée par toxicité, qualité de vie, préférences. **C'est ce que  $V_{ij}$  représente.**

## 8. Alternatives architecturales : Comparaisons

### Autologue

**Avantages** : compatibilité immunologique parfaite, persistance possible.

**Coûts** :

- $\tau_{stratégie}$  long (manufacturing)
- Patient's  $S(t)$  se dégrade pendant ce délai
- Si délai très long : patient peut quitter domaine où  $R_{autologue}$  utile

### Allogénique

**Avantages** : input standardisé, manufacturable,  $\tau_{stratégie}$  potentiellement plus court.

**Coûts** :

- Host-versus-graft rejection majeur
- Immunosuppression requise, infrastructure massive
- Domaine de performance  $\Omega$  différent (pas superset de autologue)

## Bispécifiques

**Avantages :**  $\tau_{stratégie}$  très court (généralement disponible immédiatement).

**Coûts :**

- Drug persistence (pharmacocinétique classique)  $\neq$  response durability (effet clinique prolongé)
- Microenvironnement non contourné (Treg, hypoxie)
- Escape antigénique identique au CAR-T

**Timing compétitif :** quand délai CAR-T devient pénalisant (maladie très rapide,  $\tau_{stratégie}$  long), leur **valeur relative augmente** dans calcul  $V_{ij}$ . Ils peuvent « rivaliser ».

## Bispécifiques anti-CD19/CD3 ou anti-CD20/CD3

**Avantage majeur :**  $\tau_{stratégie}$  très court (0-7 jours), souvent immédiatement disponibles.

**Mais :** profil de  $R_{bisp}(S)$  dramatiquement différent de CAR-T.

### La distinction drug persistence vs response durability

**Confusion centrale :** « Les bispécifiques persistent moins longtemps que CAR-T, donc moins utiles. »

**Techniquement :** vrai pour pharmacocinétique (T1/2 jours-semaines vs expansion cellulaire).

**Cliniquement :** faux pour réponse durable. Un anticorps peut disparaître en 4 semaines et induire réponse clinique qui dure mois-années.

**Mécanisme :** Le bispécifique ne persiste pas, mais induit la reconstitution de T cell immunity, élimination antigène persistant, remodelage tissulaire, puis ces effets persistent.

Donc :

$$R_{CAR}(S) = f(\text{auto-expansion cellulaire} + \text{persistance longue})$$

$$R_{bisp}(S) = f(\text{exposure-induced response} + \text{post-exposure durability})$$

Ces fonctions ne sont pas ordonnées, elle sont différentes.



**Données empiriques (2024-2025) :**

- **Glofitamab (anti-CD20/CD3)** : CR 93%, 2-year relapse-free ~75%
- **Epcoritamab (anti-CD20/CD3)** : CR 76%, 12-month RFS ~80%
- **Vs CAR-T** : CR 74-87%, 2-year RFS ~75-80%

Bispécifiques numériquement comparables sur réponse initiale et durabilité 2-year.

**Limitation critique : T cell fitness patient**

**CAR-T autologue** : infuse ses propres cellules génétiquement modifiées. On peut forcer même si les lymphocytes T du patient sont épuisés.

**Bispécifique** : recrute T cells endogènes du patient. Si patient a :

- CD3+ count <200
- CD8 <50
- Exhaustion markers élevés (PD-1+/TIM-3+ >50%)

Alors  $R_{bisp}(S)$  s'effondre drastiquement vs  $R_{auto}(S)$  qui tolère.

**Ce paramètre n'est pas mesuré en routine.** Donc au moment décision, aveugle.

**Microenvironnement non contourné**

CAR-T peut (partiellement) contourner obstacles (Treg, TGF- $\beta$ , hypoxie).

Bispécifique dépend du microenvironnement permissif pour recruitment T cells.

État S sévère avec Treg dominance, TGF- $\beta$  très élevé, microenvironnement hostile  $\rightarrow R_{bisp}(S)$  chute.

**Le rôle de  $\Theta$  pour bispécifiques**

$$\Theta_{bisp} = \frac{\tau_{R,bisp}}{\tau_{clinical\ evolution}} = \frac{1 - 7\text{ days}}{\text{doubling time}}$$

**Exemple 1 : Maladie ultra-rapide** (doubling time 7 jours, HGBCL Ki67 95%)

$$\Theta_{bisp} = \frac{4\text{ days}}{7\text{ days}} = 0.57 \rightarrow \text{compétitif}$$

$$\Theta_{auto} = \frac{28\text{ days}}{7\text{ days}} = 4.0 \rightarrow \text{patient quitte état utile}$$

**Sélection** :  $j^* = bisp$  (si T cell fitness acceptable)

**Exemple 2 : Maladie lente** (doubling time 60 jours)

$$\Theta_{bisp} = \frac{4 \text{ days}}{60 \text{ days}} = 0.07 \rightarrow \text{timing non décisif}$$

$$\Theta_{auto} = \frac{28 \text{ days}}{60 \text{ days}} = 0.47 \rightarrow \text{timing acceptable}$$

**Sélection :**  $j^* = auto$  (meilleur  $R_j$  sur durabilité + tolérance T cell)

Contexte clinique réaliste de sélection  $j^*$ 

**Scénario A : HGBCL, Ki67 95%, doubling ~7j, patient 55a, ECOG 1, Lym 600, T cells fatiguées**

$V_{auto}(S)$  : bas (état se dégrade J28,  $R_{auto}$  réduit)

$V_{bisp}(S)$  : élevé si T vitales, modéré sinon

$V_{allo}(S)$  : bas (trop lent pour cette cinétique)

**Décision :**  $j^* = bisp$  si **T cell fitness acceptable**, sinon CAR-T auto (risky mais seule option).

**Scénario B : DLBCL indolent-transformation, doubling ~60j, patient 65a, comorbidités, ECOG 1, Lym 400, CD8 très bas**

$V_{auto}(S)$  : élevé (état stable, peut forcer malgré T cells bad)

$V_{bisp}(S)$  : bas (dépend T cells, CD8 très bas)

$V_{allo}(S)$  : modéré (acceptable délai, mais risques supplémentaires)

**Décision :**  $j^* = auto$  **clairement**. Bispécifique compétitif seulement si patient refuse modification génétique.

## Les quatre vérités sur les bispécifiques

1. **Pas d'avantage absolu.** Meilleur dans contextes spécifiques ( $\Theta$  élevé, T cell fitness bon), pire ou égal ailleurs.
2.  **$\Theta$  très faible  $\rightarrow$  timing avantage nul en maladie lente.** Seul enjeu devient  $R_j$ , où CAR-T auto gagne (persistance).
3. **T cell fitness est LE déterminant.** CD3+, CD8, exhaustion markers (non mesurés routinièrement au moment décision).

4. **Durabilité long-terme TBD.** Bispécifiques ont 2-3 ans follow-up. CAR-T ont 5+.  
Gap diminue mais pas encore comblé.

## Synthèse

**Les bispécifiques ne « dominant » pas.** Ils rivalisent selon  $V_{ij}$  dynamique.

Framework permet de formaliser quand :

$$V_{bisp} > V_{auto}, V_{allo} \Rightarrow j^* = bisp$$

Vs quand CAR-T gagne.

**C'est intellectuellement beaucoup plus honnête** que « bispécifiques = futur ».

La vérité : **bispécifiques = option supplémentaire dans décision multi-bras**, et non pas supériorité universelle, et c'est déjà très utile.

## 9. Angles morts importants

### Therapeutic reliability : au-delà de la médiane

Un traitement en J25 médiane / P90=J60 peut être **moins utile** qu'un en J30 médiane / P90=J35.

$$\text{Therapeutic reliability} = f(\mathbb{E}, [\tau], \text{Var}(\tau)P_{90}(\tau))$$

Dans un système capacity-constrained exploité près de saturation, la **queue droite de la distribution de  $\tau$  explose**.

### Capacité de réserve comme capacité clinique

L'optimisation naïve max *utilization* ignore le coût clinique du tail latency.

$$\text{Total Value}(C) = \text{Economic Efficiency}(C) - \text{Clinical Loss from Tail}(C)$$

$$\frac{d\text{ClinicalLoss}}{dC} \text{ augmente dramatiquement près saturation}$$

Donc  $C^* < 100\%$  peut être optimal **à la fois économiquement et cliniquement**.

### Ratio dimensionless de compétition temporelle (heuristique)

$$\theta_{ij} = \frac{\mathbb{E}[\tau_{R,ij}]}{\tau_{\text{clinical evolution}}}$$

Une intervention appartient à « time-constrained therapeutics » si  $\theta \ll 1$

**Important** :  $\theta$  est une **heuristique dimensionnelle, pas un biomarqueur validé**. Le dénominateur reste à définir précisément (doubling time? temps jusqu'à ECOG 3? jusqu'à baisse utilité 50%?). Sans calibration, utiliser qualitativement (« compétition temporelle majeure ») plutôt que seuils numériques.

**Classe concernée** : CAR-T, TIL, vaccins néo-antigéniques, radiothérapie adaptative, transplantation, toute situation où  $\theta$  approche ou dépasse 1.

## 10. Implications opérationnelles

### Conséquence 1: Mesurer decision-to-treatment

Les KPIs actuels rapportent vein-to-vein ou apheresis-to-infusion. Ils ignorent decision-to-aphérèse.

**Le véritable enjeu** : patients peuvent perdre la fenêtre **avant même l'aphérèse**.

KPI candidat : decision point → effective administration.

### Conséquence 2: Mesurer la distribution, pas médiane seule

Rapporter  $\mathbb{E}[\tau]$ ,  $E[\tau]$ , variance, P90 tail latency.

P90 est cliniquement plus pertinent que médiane pour prédire % patients avec échec temporel.

### Conséquence 3: Intégrer $\tau_{ij}$ anticipé dans la sélection

La décision optimale doit explicitement considérer  $\tau_{R,ij}$  anticipé et son impact sur  $S_i(t_0 + \tau_{ij})$ .

**Exemplification** : pathways décision-support basés sur [ $\theta$  anticipé, alternatives temporelles disponibles], outcomes mesurés : decision-to-treatment, toxicités observées, survival-depuis-diagnostic.

## 11. Généralisation : Les interventions time-constrained

Toute intervention où  $\theta_{ij} \ll 1$  obéit aux mêmes principes :

1. État du patient au moment administration modifie valeur relative traitements
2. Capacité de réserve système = capacité clinique

3. Fenêtre thérapeutique dépend couple (patient, intervention)
4. Décision optimale = optimisation dynamique  $j^* = \arg \max_j V_{ij}$

## Métrique unifiée d'efficacité

Plutôt que de comparer « efficacité brute » inter-essais, utiliser **framework**  $V_{ij}$  **intégrant timing**.

Cela réconcilie :

- ZUMA-7 (efficacy visiblement supérieure à SOC)
- TRANSFORM (efficacy visiblement supérieure à SOC)
- BELINDA (efficacy = SOC)

**Non** en disant « 4-1BB inférieur ». Mais : « architecture temporelle + population ultra-rapide = échec temporel indépendamment du domaine costimulation. »

## 12. Conclusions théorique et pratique

### Conclusion théorique rigoureuse

Pour les interventions dont la cinétique de production rivalise avec celle de la maladie, **le traitement ayant la meilleure efficacité moyenne n'est pas nécessairement celui qui maximise l'utilité attendue pour un patient donné.**

L'utilité anticipée  $V_{ij}$  dépend de l'état  $S_i$  **au moment où le traitement  $j$  peut être administré**, pas seulement du traitement lui-même.

Donc :

***The relevant treatment is not necessarily the treatment with highest average efficacy. It is the treatment with the highest expected value given the patient's anticipated state when that treatment can actually be delivered.***

### Conclusion pour la pratique clinique

***The relevant treatment is not the best treatment. It is the best reachable treatment, the one that can reach the patient while the patient remains in its performance domain.***