

*COLLÈGE NATIONAL
DES GYNÉCOLOGUES ET OBSTÉTRICIENS FRANÇAIS
Président : Professeur B. Hédon*

Cinquième partie
**Pathologies mammaires
et cancer du sein**



*38^{es} JOURNÉES NATIONALES
Paris, 2014*

Pathologies mammaires et cancer du sein : modèles statistiques

R. ROUZIER^{1, 2}, C. BONNEAU^{1, 2}, C. HUCHON³, C. COUTANT⁴
(Saint-Cloud, Versailles, Poissy, Dijon)

Résumé

L'utilisation de modèles mathématiques peut améliorer la prédiction individuelle tant de l'histoire naturelle d'une maladie que des bénéfices d'une prise en charge diagnostique ou thérapeutique. Elle permet ainsi d'aboutir à une prise en charge personnalisée et de proposer des résultats sous forme de concepts organisés en arbres binaires de décision. Elle rend possible l'intégration de nouvelles données haut débit, en particulier les données de génomique et de protéomique. Ces outils vont probablement devenir indispensables, mais la réalisation de grands essais étant de plus en plus lourde est coûteuse, la démonstration du service rendu sera complexe. Les modèles mathématiques et l'utilisation de techniques d'intelligence artificielle peuvent potentiellement permettre de faciliter l'émergence de nouvelles stratégies en comparant leur effet avec

1 - Institut Curie - 35 rue Dailly - 92210 Saint-Cloud

2 - Université de Versailles-St-Quentin - EA 7285 - Risques cliniques et sécurité en santé des femmes et en santé périnatale - 55 avenue de Paris - 78000 Versailles

3 - CHI Poissy-Saint-Germain-en-Laye - 10 rue du Champ Gaillard - 78303 Poissy cedex

4 - Centre Georges-François Leclerc - Université de Bourgogne - 1 rue du Professeur Marion - 21000 Dijon

l'effet attendu d'un régime de référence dont l'efficacité aura été prédite au niveau individuel et non pas au niveau d'une population.

Mots clés : prédiction, modélisation, calibration, discrimination, validation

Déclaration publique d'intérêt

Les auteurs ne déclarent aucun conflit d'intérêt avec le sujet présenté.

INTRODUCTION

De par ses fonctions, le médecin est souvent amené à intervenir dans des processus de décision de natures très diverses. Nombre de ces décisions, telles l'identification d'une pathologie, l'interprétation d'imagerie médicale ou l'évaluation des chances de succès d'une thérapeutique peuvent être rattachées à des problèmes de classement (en anglais : *classification*). Le processus de classement consiste à attribuer une classe préalablement définie (préexistante) à un individu ou à une situation donnée, à ne pas confondre avec la classification (en anglais : *clustering*), dont l'objet est la création *ex nihilo* de nouvelles classes par groupement d'individus présentant des caractéristiques semblables. Pour mener à bien ces tâches de classement, le médecin peut utiliser deux voies distinctes mais non exclusives :

- un raisonnement formel, basé sur une théorie clairement établie et formulée,
- une réflexion empirique, résultat de la synthèse des expériences passées.

La médecine dispose en fait d'un arsenal restreint de théories de type déterministe comparé à d'autres domaines appliqués tels que la physique ou la chimie. En effet, la multiplicité des facteurs intervenant dans les processus pathologiques, la complexité de leurs interactions et la variabilité du vivant rendent difficile leur mise en équation. Cet état limite *de facto* la possibilité de recours à la voie purement théorique et conduit à extraire principalement les connaissances et à construire les facultés de classement à partir des expériences passées. Ceci a d'ailleurs

été à l'origine de l'avènement de l'expérimentation et de l'inférence statistique [1], définie comme l'ensemble des méthodes permettant de tirer des conclusions fiables à partir de données d'échantillons statistiques provenant de populations réelles ou hypothétiques.

Cependant, l'inférence statistique n'offre qu'une utilité marginale dans les procédures de synthèse conduisant aux classements car elle repose sur des hypothèses de travail qui doivent elles-mêmes être au préalable extraites des expériences passées. Elle intervient donc éventuellement en phase de confirmation, à la suite du classement, pour en vérifier les hypothèses et non pour les élaborer.

La première étape permettant de refléter l'expérience passée consiste à rechercher des facteurs de risque d'un événement donné : facteurs de risque de non-réponse à une chimiothérapie néoadjuvante, facteurs de risque de rechute locale ou à distance, facteurs de risque de complication des traitements adjuvants...

Un problème médical a rarement une solution unique : probabilité d'un diagnostic par exemple, balance bénéfice/risque d'un traitement... Plusieurs méthodes d'analyse ou de classement permettent de rendre compte de l'incertitude. Par exemple, la régression logistique qui consiste à modéliser le ratio de probabilité (probabilité d'un événement / (1 - probabilité de cet événement)) en une équation linéaire rend bien compte de la probabilité d'un événement. Pour extraire des règles de classement d'un ensemble d'exemples existants, d'autres algorithmes existent mais sont souvent moins performants que la régression logistique [2].

Le continuum entre la recherche de facteurs de risque jusqu'à la création de prédicteurs et d'interfaces web donne la perspective d'une médecine personnalisée.

I. CRÉATION D'UN PRÉDICTEUR

I.1. Du modèle univarié au prédicteur

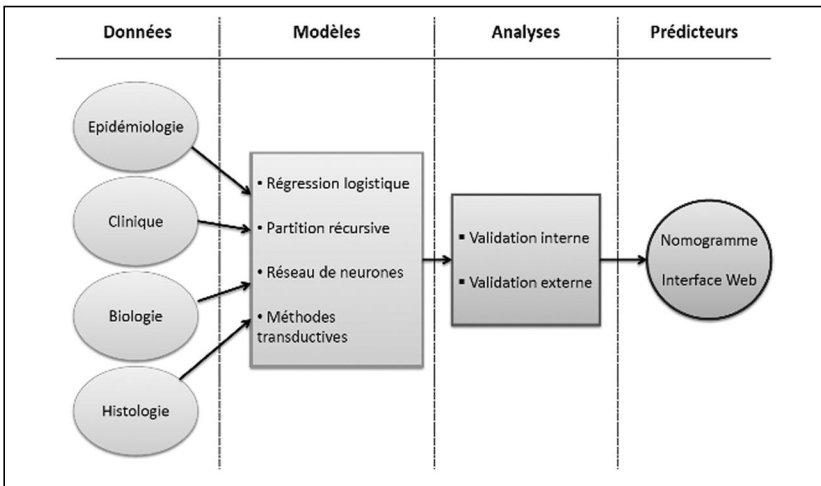
L'association entre une caractéristique clinique et un événement lié à une maladie peut être testée dans différents modèles. Tester l'association entre une caractéristique clinique et un événement sans tenir compte des autres facteurs se fait par une analyse univariée. Dans les analyses univariées, on se contente de regarder les odds ratios (OR) bruts sans chercher à ajuster les résultats sur l'existence d'éventuelles variables de confusion. Tester une association de façon univariée ne

permet en fait que rarement d'établir un lien causal. L'analyse univariée a valeur de travail cognitif et exploratoire mais ne peut que rarement être utilisée pour la prédiction.

L'utilisation de modèles multivariés permet, au moins partiellement, de prendre en compte les tiers facteurs qui pourraient biaiser les résultats d'une étude univariée. Ces modèles peuvent être utilisés pour prédire des évènements comme nous allons le démontrer.

Depuis que nous développons cette thématique, nous avons mis au point une stratégie de développement de prédicteur. Nous utilisons surtout la régression logistique mais nous avons aussi développé des modèles de partition récursive et des réseaux de neurones. La création se fait selon le modèle décrit dans la figure 1.

Figure 1 - Stratégie de développement d'un prédicteur



I.2. Modèles mathématiques

Il est envisageable de combiner les facteurs cliniques, histologiques et biologiques pour améliorer notre capacité de prédiction.

La méthode la plus simple est la régression logistique. Le principe du modèle de régression logistique est la modélisation du logarithme népérien du ratio des pourcentages par une équation linéaire :

$$\ln \left(\frac{P}{1-P} \right) = a + b_1 x_1 + b_2 x_2 + \dots + b_i x_i = a + b X$$

où les coefficients a et b_1, b_2, \dots, b_i sont à déterminer et X est un vecteur de données connues x_1, x_2, \dots, x_i .

À partir de la modélisation :

$\ln(P / (1/P)) = a + b_1 x_1 + b_2 x_2 + \dots + b_i x_i = a + b X$
 on obtient

$$P = \frac{1}{1 + e^{-(aX+b)}}$$

Il est donc possible de définir une probabilité individuelle à partir du modèle multivarié.

En pathologie mammaire, cette approche permet d'utiliser l'information de cas traités pour prédire la réponse à la chimiothérapie néoadjuvante, la probabilité d'envahissement ganglionnaire, de récurrence locale ou à distance...

II. MÉTHODE DE VALIDATION D'UN MODÈLE

Tout modèle doit être validé. La validation d'un modèle repose sur 2 mesures essentiellement : la discrimination et la calibration [3]. La discrimination est la capacité d'un classement à faire des prédictions individuelles dans le bon ordre. Elle peut être quantifiée par l'aire comprise sous la courbe d'une courbe ROC (*receiver operating characteristic*) aussi appelée index de concordance. L'index de concordance est la probabilité que, pour 2 patientes désignées au hasard, la patiente ayant l'issue la plus défavorable ait la prédiction d'issue la plus défavorable. L'index de concordance varie entre 0 et 1, 1 indiquant une concordance parfaite, 0,5 indiquant une concordance équivalente au hasard, et 0 une parfaite discordance. La calibration est l'agrément entre les fréquences observées pour l'issue étudiée et les probabilités prédites (courbes de calibration). Une courbe de calibration est construite en indiquant en abscisse des moyennes de prédiction et en ordonnée des moyennes observées pour des groupes de patientes définis par quantile. Un modèle parfait va avoir pour courbe de calibration une droite passant par l'origine et ayant une pente de 45°.

Il existe plusieurs solutions pour valider un modèle : la meilleure est sans nul doute la validation prospective dans un centre indépendant. Une validation prospective dans un centre indépendant démontre l'exportabilité d'un modèle. Elle se heurte à un écueil majeur : l'évolution des thérapeutiques. En effet, les séries sur lesquelles il est possible de faire des études car elles ont un recul mature

comprennent des patientes qui ont reçu des thérapeutiques qui ne correspondent plus exactement aux standards actuels. Il est donc difficile d'imposer à des patientes des thérapeutiques pour lesquelles, a *contrario*, nous disposons de prédictions individuelles mais qui sont dépassées par de nouvelles molécules ou de nouvelles techniques qui ont montré leur supériorité à l'échelon d'une population.

Des techniques de validation interne sont disponibles. Par exemple, les techniques de *bootstrapping* qui consistent à fractionner la population en deux groupes pour créer un algorithme sur un groupe et le valider sur l'autre groupe. Cela permet de connaître dans quelle mesure un modèle est trop optimiste.

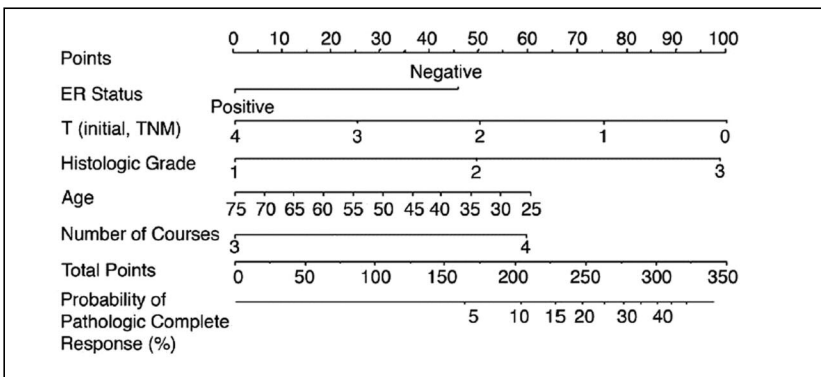
III. APPLICATIONS DES PRÉDICTEURS

III.1. Illustration par la prédiction de la chimiosensibilité des cancers du sein

Prédiction individualisée - développement d'interfaces web

Notre équipe a créé et validé un modèle multivarié de régression logistique pour prédire la réponse histologique complète (PCR) à la chimiothérapie néoadjuvante (Figure 2) [4]. L'utilisation en pratique se fait par un nomogramme : ce nomogramme est basé sur la stadification clinique, le statut des récepteurs aux œstrogènes, le grade et le type

Figure 2 - Nomogramme pour prédire la réponse à la chimiothérapie néoadjuvante (modèle de régression logistique)



histologique et le nombre de cures de chimiothérapie préopératoires. Le nomogramme avait une bonne calibration et une bonne discrimination tant sur le « training set » que sur le « validation set » (index de concordance : 0,77 et 0,79).

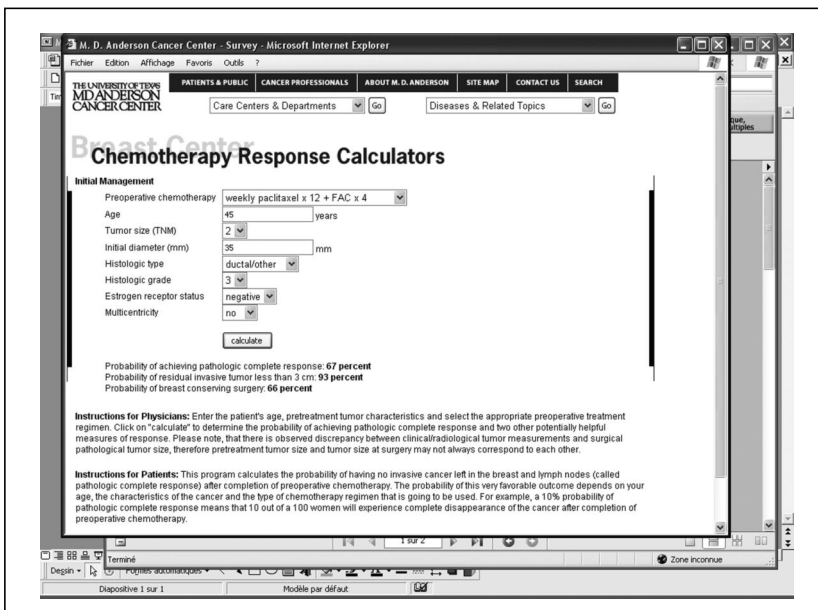
À chaque variable correspond un nombre de points sur l'axe du haut, ces points sont ensuite additionnés et au total correspond une probabilité de réponse complète à la chimiothérapie.

L'utilisation d'un nomogramme est cependant peu pratique et ergonomique. Pour la présentation, nous avons choisi d'utiliser un nomogramme et développé une interface web pour rendre l'accès plus convivial (Figure 3). Le praticien peut ainsi rentrer les données clinico-pathologiques de la patiente et déterminer sa probabilité de réponse complète à la chimiothérapie.

L'utilisation des interfaces web peut permettre d'informer au mieux les patientes et aboutir à une prise en charge individualisée.

Nous avons cependant démontré le caractère limité aux tumeurs HER2 négatif. Les données de 200 patientes atteintes de cancer du sein

Figure 3 - Interface web dérivée du nomogramme pour prédire la réponse complète à la chimiothérapie



traitées par chimiothérapie préopératoire ont été recueillies. Nous avons calculé la prédiction du taux de PCR avec les deux nomogrammes (celui rapporté plus haut et celui de Colleoni *et al.* [5]) puis nous avons comparé les prédictions avec les résultats (calibration). Soixante pour cent des patientes atteintes de tumeurs HER2 positives avaient reçu du trastuzumab concomitant avec les taxanes. Les performances du modèle ont été quantifiées en termes de discrimination et de calibration [6]. Dans l'ensemble de la population, les index de concordance étaient bons (0,74 et 0,75) mais les nomogrammes sous-estimaient le taux de PCR ($p = 0,026$ et $0,0005$). Lorsque les patientes HER2 positives traitées avec le trastuzumab étaient exclues, les index de concordance étaient excellents : 0,78 pour les deux nomogrammes avec des calibrations satisfaisantes, c'est-à-dire aucune différence significative entre les prédictions et les résultats observés. *A contrario*, lorsque le sous-groupe des patientes HER2 positives traitées avec le trastuzumab en préopératoire a été analysé, la discrimination et la calibration étaient mauvaises, démontrant ainsi *de facto* l'impact d'une thérapeutique ciblée.

III.2. Illustration par la probabilité d'envahissement des ganglions non sentinelles en cas de ganglion sentinelle métastatique

En cas d'envahissement métastatique du ganglion sentinelle (GS) chez les patientes ayant un cancer du sein, un curage axillaire (CA) est recommandé. Cependant, dans 40 à 70 % des cas les ganglions non sentinelles (GNS) sont indemnes. Pour ce groupe de patientes, un CA complémentaire apparaît donc inutile. Dans ce contexte, des prédicteurs clinicopathologiques (c'est-à-dire modèles mathématiques prédictifs intégrant plusieurs variables) ont été développés afin de prédire le risque individuel d'envahissement des GNS. Coutant *et al.* ont réalisé une validation des neuf prédicteurs développés à ce jour pour prédire le statut des GNS en cas d'envahissement du GS sur une cohorte prospective, multicentrique et indépendante de 561 patientes ayant au moins un GS métastatique [5]. Les auteurs ont montré que le nomogramme du *Memorial Sloan-Kettering Cancer Center* (MSKCC) et le score de Tenon étaient les deux prédicteurs les plus performants, tant sur les critères de fiabilité que d'utilité clinique, et ont confirmé leur reproductibilité et leur exportabilité pour prédire le statut des GNS en cas d'envahissement du GS. Le nomogramme du MSKCC est un modèle continu, c'est-à-dire fournissant pour chaque patiente une

probabilité d'envahissement des GNS. Le score de Tenon est un modèle binaire qui sépare en fonction d'un seuil la cohorte en deux groupes (haut risque et bas risque).

Une étude prospective multicentrique financée par un PHRC (programme hospitalier de recherche clinique) a été menée pour démontrer l'exportabilité du nomogramme du *Memorial Sloan-Kettering Cancer Center* et du score de Tenon : l'étude NOTEGS.

Elle était basée sur l'hypothèse que l'utilisation combinée du score de Tenon et du nomogramme du MSKCC a un triple intérêt :

- 1) sélectionner les patientes à très faible risque d'envahissement des GNS (environ 1/3 des patientes) ne justifiant pas d'un CA complémentaire, permettant ainsi de diminuer la morbidité liée au CA (lymphocèle, lymphœdème, douleurs chroniques et difficultés de mobilisation du membre supérieur homolatéral au CA) ainsi que le surcoût lié à cette réintervention et à la prise en charge des complications précoces et chroniques éventuelles ;
- 2) permettre par le caractère combiné de la prédiction une augmentation des performances prédictives en diminuant le taux de faux négatifs ;
- 3) augmenter la pertinence de ces deux prédicteurs sur le plan clinique. En effet, l'utilisation combinée d'un modèle continu et d'un modèle binaire peut s'avérer très utile pour la pratique clinique : un résultat binaire (positif/négatif) est très pertinent sur le plan pratique alors qu'une quantification du risque est plus informative.

L'objectif principal était d'évaluer de manière prospective l'utilisation combinée du nomogramme du MSKCC et du score de Tenon pour sélectionner, en cas de GS métastatique, les patientes à faible risque d'envahissement des GNS et chez qui un CA pourrait être évité. Le critère de jugement principal était le taux de faux négatifs chez les patientes ayant une probabilité d'envahissement des GNS avec le nomogramme du MSKCC $\leq 10\%$ et un score de Tenon $\leq 3,5$ (c'est-à-dire à bas risque) : pourcentage de patientes ayant un envahissement des GNS prélevés lors du CA complémentaire alors qu'elles étaient considérées comme à faible risque par l'utilisation combinée des 2 prédicteurs.

Les objectifs secondaires étaient de : i) déterminer l'impact, sur la fiabilité du prédicteur combiné, de l'examen extemporané en tenant compte de sa réalisation ou non, et de sa sensibilité ; ii) permettre, du fait du nombre de patientes important à inclure dans l'étude, de diffuser une charte de qualité de la procédure du GS.

Cette étude s'est heurtée à l'abandon du curage systématique en cas de GS métastatique mais elle illustre parfaitement la modélisation statistique en pathologie mammaire.

Parmi les 2 936 patientes analysables, au moins un GS était métastatique chez 696 patientes (23,7 %). Parmi elles, 178 n'ont pas eu de CA (en raison des résultats des essais Z011 et IBCSG 23-01) dont 47,5 % des cas étaient à faible risque ($p < 0,001$). Parmi les patientes ayant eu un CA, $n = 518$ (74,4 %), 67 (13 %), 437 (84 %) et 14 (3 %) patientes avaient un risque combiné bas, haut et indéterminé respectivement. Le taux de faux négatifs chez les patientes à faible risque était de 16,4 % (11/67) [intervalle de confiance à 95 % : 8,5 %-27,5 %]. Cette étude n'a donc pas validé l'hypothèse initiale. Toutefois, ce taux se comparait favorablement avec les patientes du groupe à haut risque : 33,9 % (148/437) [intervalle de confiance à 95 % : 29,6 %-38,4 %] avaient des ganglions non sentinelles métastatiques ($p = 0,004$). Analysés individuellement, le score de Tenon score et le nomogramme du MSKCC avaient des performances similaires. Cette étude démontre donc que le score de Tenon et le nomogramme du MSKCC ont une discrimination satisfaisante mais une calibration insuffisante en cas de modifications des critères d'application.

IV. PERSPECTIVES

L'utilisation d'un nomogramme peut permettre de comparer une issue observée à une issue prédite. Cela nous a permis par exemple de démontrer que les cancers du sein associés à la grossesse (CSAG) sont aussi chimiosensibles que les cancers du sein en dehors de la grossesse [8]. Sur une série de 48 patientes ayant un CSAG, le taux de réponse complète à la chimiothérapie était de 21 % pour un taux attendu de 21 % en dehors de toute grossesse. Dans une population aussi complexe, l'utilisation d'un nomogramme est une méthode séduisante pour tester une hypothèse.

D'une manière générale, cette approche permet de comparer une issue prédite à une issue observée avec 2 perspectives possibles : pour démontrer la supériorité d'un traitement A sur un traitement B, on réalise classiquement un essai comparant 2 groupes de patientes : un groupe recevant (A) et un groupe recevant (B). Si à l'échelle de la population le bras qui reçoit (A) a une issue statistiquement plus favorable, on en conclut que A est supérieur à B et cela conduit à la

recommandation de prescrire A à l'ensemble de la population. Cependant, on peut imaginer que chez certaines patientes B est plus efficace que A mais que A est plus efficace que B sur un plus grand nombre de patientes. Identifier le bénéfice individuel peut permettre d'optimiser l'efficacité des thérapeutiques.

Il est envisageable de développer de nouveaux designs d'essais cliniques [9]. En effet, la plupart des essais de phase 2 se basent sur une hypothèse. Les modèles mathématiques et l'utilisation de techniques d'intelligence artificielle peuvent potentiellement permettre de faciliter l'émergence de nouvelles drogues en comparant leur effet avec l'effet attendu d'un régime de référence dont l'efficacité aura été prédite au niveau individuel et non pas au niveau d'une population.

Les modèles statistiques peuvent permettre le développement de modèles combinés en intégrant les nouvelles données haut débit, en particulier les données de génomique et de protéomique. Ces données telles qu'elles sont rapportées aujourd'hui sous forme de signature sont peu utiles car peu reproductibles et liées à la méthodologie d'exploitation. Ces outils vont probablement devenir indispensables car de plus en plus de régimes de chimiothérapie sont disponibles et la réalisation de grands essais est de plus en plus lourde et coûteuse.

Bibliographie

- [1] Fisher R. Statistical methods for research workers. Ed. Oliver and Boyd. 1925, Edinburg (UK). 239.
- [2] Kattan MW. Comparison of Cox regression with other methods for determining prediction models and nomograms. *J Urol* 2003;170(6Pt2):S6-9.
- [3] Harrell Jr FE. Regression modeling strategies, with applications to linear models. Logistic regression, and survival analysis. New York: Springer 2001.
- [4] Rouzier R *et al.* Nomograms to predict pathologic complete response and metastasis-free survival after preoperative chemotherapy for breast cancer. *J Clin Oncol* 2005;23:8331-9.
- [5] Colleoni M *et al.* A nomogram based on the expression of Ki-67, steroid hormone receptors status and number of chemotherapy courses to predict pathological complete remission after preoperative chemotherapy for breast cancer. *Eur J Cancer* 2010;46:2216-24.
- [6] Frati A *et al.* Comparison of two nomograms to predict pathologic complete responses to neoadjuvant chemotherapy for breast cancer: evidence that HER2-positive tumors need specific predictors. *Breast Cancer Res Treat* 2011;132:601-7.
- [7] Coutant *et al.* Comparison of models to predict nonsentinel lymph node status in breast cancer patients with metastatic sentinel lymph nodes: a prospective multicenter study. *J Clin Oncol* 2009;27:2800-8.
- [8] Rouzier R *et al.* Pregnancy-associated breast cancer is as chemosensitive as non-pregnancy-associated breast cancer in the neoadjuvant setting. *Ann Oncol* 2011;22:1582-7.
- [9] Jia Z *et al.* Generation of "virtual" control groups for single arm prostate cancer adjuvant trials. *PLoS One* 2014;9:e85010.