



Sciences Économiques et Sociales de la Santé  
& Traitement de l'Information Médicale

**Roch GIORGI**

Directeur SESSTIM

Responsable Équipe QuantIM - UMR1252 - SESSTIM

Méthodes de Prédiction du Risque Santé basée sur un Algorithme : Transparence et Complétude

**avril 2022**

*[Cliquez ici pour voir l'intégralité des ressources associées à ce document](#)*

# Méthodes de Prédiction du Risque Santé basée sur un Algorithme : Transparence et Complétude

Roch Giorgi

Séminaire interne du SESSTIM  
11 avril 2022, Marseille

SESSTIM, Faculté des Sciences Médicales et Paramédicales, Aix-Marseille Université, Marseille, France  
<http://sesstim.univ-amu.fr/>



Sciences Economiques et Sociales de la  
Santé & Traitement de l'Information Médicale

Inserm / IRD / Université AIX-MARSEILLE



# Risque en Santé

- Risque

- Survenue de la maladie

- Consommation de tabac : est un facteur de risque du cancer du poumon

- Fréquence de survenue de la maladie (incidence) dans la population donnée, pendant une période déterminée

- Estimée dans un échantillon représentatif de la population d'intérêt

- Facteur de risque (néfaste / protecteur)

- L'exposition à ce facteur modifie la survenue de la maladie

- Quantification de l'effet par différents indicateurs

- Estimation selon une méthodologie appropriée

# Etude de Framingham

- Cohorte ayant débutée en 1948 aux Etats-Unis
- Inclusion des sujets selon des critères bien spécifiques
- Suivi des sujets avec un recueil prospectif, actif, des informations (données de santé environnementales, comportementales)
- Etude du risque cardiovasculaire

# Etude de Framingham

Score de Framingham <http://medicalcul.free.fr/framingham.html>



-----

<input checked="" type="radio"/>	Homme	
<input type="radio"/>	Femme	
Age :	48	ans
Cholestérol Total :	1,4	g/l
Cholestérol HDL :	0,5	g/l
PA Systolique :	85	mmHg
PA Diastolique :	120	mmHg
<input type="checkbox"/>	Fumeur	
<input type="checkbox"/>	Diabétique	

Calculer

Risque à 10 ans : 7%

## Interprétation :

Le score de Framingham permet une estimation de l'éventualité de survenue d'un évènement cardio-vasculaire dans les 10 prochaines années, chez un patient ne présentant pas déjà une pathologie cardio-vasculaire connue.

Plusieurs versions de ce score existent. La version présentée ici est basée sur les références citées ci-dessous.

## Références :

Wilson, D'Agostino, Levy *et al.* Prediction of Coronary Heart Disease using Risk Factor Categories; Circulation; 1998; 97:1837-1847.

Framingham Heart Study: <http://www.framinghamheartstudy.org>.

# Etude de Framingham

Score de Framingham <http://medicalcalcul.free.fr/framingham.html>



Homme  
 Femme

Age : 48 ans

Cholestérol Total : 1,4 g/l

Cholestérol HDL : 0,5 g/l

PA Systolique : 85 mmHg

PA Diastolique : 120 mmHg

Fumeur  
 Diabétique

Calculer

Risque à 10 ans : 7%

Homme  
 Femme

Age : 48 ans

Cholestérol Total : 1,4 g/l

Cholestérol HDL : 0,5 g/l

PA Systolique : 85 mmHg

PA Diastolique : 120 mmHg

Fumeur  
 Diabétique

Calculer

Risque à 10 ans : 10%

## Interprétation :

Le score de Framingham permet une estimation de l'éventualité de survenue d'un évènement cardio-vasculaire dans les 10 prochaines années, chez un patient ne présentant pas déjà une pathologie cardio-vasculaire connue.

Plusieurs versions de ce score existent. La version présentée ici est basée sur les références citées ci-dessous.

## Références :

Wilson, D'Agostino, Levy *et al.* Prediction of Coronary Heart Disease using Risk Factor Categories; Circulation; 1998; 97:1837-1847.

Framingham Heart Study: <http://www.framinghamheartstudy.org>.

# Etude de Framingham

Score de Framingham <http://medicalcalcul.free.fr/framingham.html>



Homme  
 Femme

Age : 48 ans

Cholestérol Total : 1,4 g/l

Cholestérol HDL : 0,5 g/l

PA Systolique : 85 mmHg

PA Diastolique : 120 mmHg

Fumeur  
 Diabétique

Calculer

Risque à 10 ans : 7%

Homme  
 Femme

Age : 48 ans

Cholestérol Total : 1,4 g/l

Cholestérol HDL : 0,5 g/l

PA Systolique : 85 mmHg

PA Diastolique : 120 mmHg

Fumeur  
 Diabétique

Calculer

Risque à 10 ans : 10%

Homme  
 Femme

Age : 48 ans

Cholestérol Total : 1,4 g/l

Cholestérol HDL : 0,5 g/l

PA Systolique : 85 mmHg

PA Diastolique : 120 mmHg

Fumeur  
 Diabétique

Calculer

Risque à 10 ans : 16%

## Interprétation :

Le score de Framingham permet une estimation de l'éventualité de survenue d'un événement cardio-vasculaire dans les 10 prochaines années, chez un patient ne présentant pas déjà une pathologie cardio-vasculaire connue.

Plusieurs versions de ce score existent. La version présentée ici est basée sur les références citées ci-dessous.

## Références :

Wilson, D'Agostino, Levy *et al.* Prediction of Coronary Heart Disease using Risk Factor Categories; Circulation; 1998; 97:1837-1847.

Framingham Heart Study: <http://www.framinghamheartstudy.org>.

# Etude de Framingham

Score de Framingham <http://medicalcalcul.free.fr/framingham.html>



Homme  
 Femme

Age : 48 ans

Cholestérol Total : 1,4 g/l

Cholestérol HDL : 0,5 g/l

PA Systolique : 85 mmHg

PA Diastolique : 120 mmHg

Fumeur  
 Diabétique

Calculer

Risque à 10 ans : 7%

Homme  
 Femme

Age : 48 ans

Cholestérol Total : 1,4 g/l

Cholestérol HDL : 0,5 g/l

PA Systolique : 85 mmHg

PA Diastolique : 120 mmHg

Fumeur  
 Diabétique

Calculer

Risque à 10 ans : 10%

Homme  
 Femme

Age : 48 ans

Cholestérol Total : 1,4 g/l

Cholestérol HDL : 0,5 g/l

PA Systolique : 85 mmHg

PA Diastolique : 120 mmHg

Fumeur  
 Diabétique

Calculer

Risque à 10 ans : 16%

**ATTENTION**

- Population nord-américaine
- Ne prend pas en compte tous les facteurs de risque
- Ne prend pas en compte la durée de l'exposition aux facteurs de risque

## Interprétation :

Le score de Framingham permet une estimation de l'éventualité de survenue d'un évènement cardio-vasculaire dans les 10 prochaines années, chez un patient ne présentant pas déjà une pathologie cardio-vasculaire connue.

Plusieurs versions de ce score existent. La version présentée ici est basée sur les références citées ci-dessous.

## Références :

Wilson, D'Agostino, Levy *et al.* Prediction of Coronary Heart Disease using Risk Factor Categories; Circulation; 1998; 97:1837-1847.

Framingham Heart Study: <http://www.framinghamheartstudy.org>.

# Facteur de Risque et Causalité

- Identification d'un mécanisme physiopathologique associé à une maladie permettant de formuler une hypothèse causale
  - Recherche de l'existence de faits épidémiologiques en accord
- Détection de l'augmentation d'une maladie
  - Recherche de facteurs candidats causaux

# Facteur de Risque et Causalité

- Causalité ?
  - Difficile à établir
  - Identifier la relation entre un candidat cause et un candidat effet
  - Démarche pragmatique
  - Causes souvent multi-factorielles
  - Démarche globale

# Facteur de Risque et Causalité

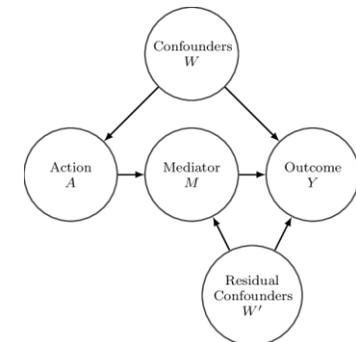
- Causalité ?

- Difficile à établir
- Identifier la relation entre un candidat cause et un candidat effet
- Démarche pragmatique
- Causes souvent multi-factorielles
- Démarche globale

## Critères de Bradford Hill

- Force de l'association
- Cohérence (externe, interne)
- Temporalité (cause précède effet)
- Relation dose-effet (gradient)
- Preuve expérimentale
- Spécificité de l'association
- Cohérence biologique
- Plausibilité biologique
- Analogie

## Graphique Acyclique Orienté Inférence causale

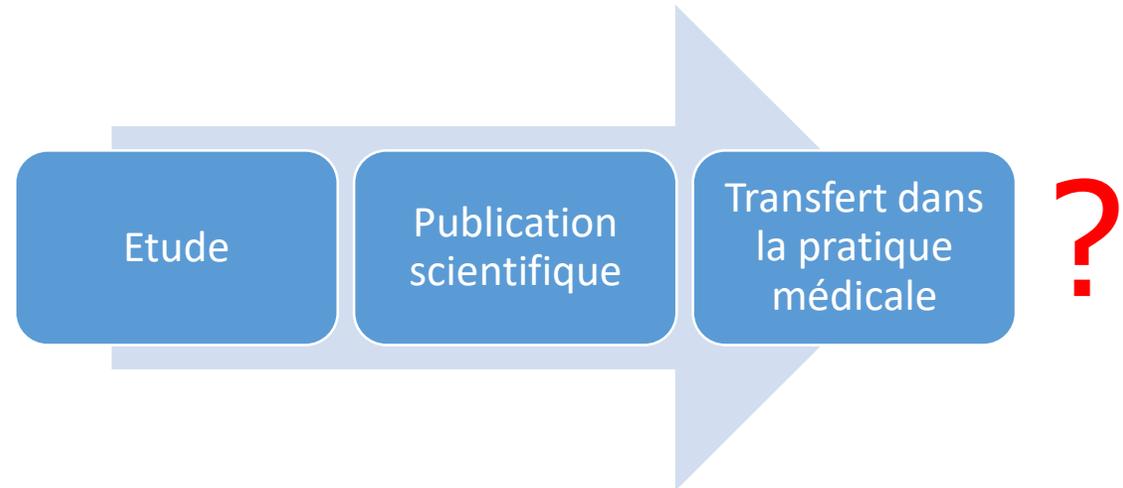


# Modélisation de Prédiction du Risque Santé

- Schéma d'étude
- Qualité de l'information recueillie
- Modèle biostatistique / économétrique / statistique / mathématique
- Variables / Features
- Validation d'un modèle
  - Discrimination, calibration
- Reproductibilité
- Extrapolation

# Modélisation de Prédiction du Risque Santé

- ✓ Schéma d'étude
- ✓ Qualité de l'information recueillie
- ✓ Modèle biostatistique / économétrique / statistique / mathématique
- ✓ Variables / Features
- ✓ Validation d'un modèle
  - Discrimination, calibration
- ✓ Reproductibilité
- ✓ Extrapolation



# Transfert dans la Pratique Médicale

- Manque d'informations sur le développement du modèle
  - De la description des patients étudiés à la méthodologie utilisée pour développer le modèle
  - Effectifs insuffisants
  - Prédicteurs examinés
  - Traitement des données manquantes
  - Stratégie de développement des modèles
  - ...
- Nombre de modèles mis en œuvre, utilisés dans la pratique
  - <<< nombre de modèles de prédiction publiés

# TRIPOD Statement

- **T**ransparent **R**eporting of a multivariable prediction model for **I**ndividual **P**rognosis **O**r **D**iagnosis
  - Guidelines de rédaction des articles scientifiques
  - Dans la même ligne directrice que
    - **CONSORT** Statement: **CON**solidated **S**tandards **O**f **R**eporting **T**rials
    - **STROBE** Statement: **ST**rengthening the **R**eporting of **OB**servational studies in **E**pidemiology
    - **REMARK** Statement: **RE**porting recommendations for tumour **MARKER** prognostic studies
    - **STARD** Statement: **ST**andards for the **R**eporting of **D**iagnostic accuracy studies
    - **GRIPS** Statement: **G**enetic **RI**sk **P**rediction **S**tudies
- ⇒ Obligatoire pour de nombreuses revues biomédicales, instances décisionnelles

# TRIPOD Statement

Section/topic	Item	Development or validation?	Checklist item	Page
<b>Title and abstract</b>				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	
<b>Introduction</b>				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model, or both.	
<b>Methods</b>				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	
	5b	D;V	Describe eligibility criteria for participants.	
	5c	D;V	Give details of treatments received, if relevant.	
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	
Predictors	7a	D;V	Clearly define all predictors used in developing the multivariable prediction model, including how and when they were measured.	
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	
Sample size	8	D;V	Explain how the study size was arrived at.	
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	

Gary S Collins, Johannes B Reitsma, Douglas G Altman, Karel G M Moons. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015;162(1):55-63.

Also published in BJOG, British Journal of Cancer, British Journal of Surgery, BMC Medicine, British Medical Journal, Circulation, Diabetic Medicine, European Journal of Clinical Investigation, European Urology, and Journal of Clinical Epidemiology.



# Modèle de Régression – Apprentissage Automatique

- Modèle de régression de prédiction du risque santé
  - Equation reliant un ensemble de variables prédictives à une variable réponse
  - Facilement décrite pour être mise en œuvre par d'autres personnes
- Modèle de prédiction du risque santé basé sur un algorithme
  - Souvent présenté comme une solution en présence de grande base de données, grand nombre de variables, sources de données différentes et hétérogènes
  - Plus dépendante des ordinateurs pour la mise en œuvre du modèle sous-jacent
  - Performances prédictives peuvent être constamment enrichies de nouvelles données observationnelles ( $\neq$  formule chimique d'un médicament fixe dans le temps)
  - Souvent qualifié de « boîte noire »

# Méthodes de Prédiction du Risque Santé basée sur un Algorithme et Utilisation en Pratique pour la Santé

- Inquiétudes quant à une surestimation de leurs apports
- Conséquences d'une prédiction erronée ou inexacte peuvent être considérables pour la santé
- Responsabilité de l'algorithme auto-apprenant ou responsabilité du médecin utilisateur ?

# Méthodes de Prédiction du Risque Santé basée sur un Algorithme et Utilisation en Pratique pour la Santé

- Besoin de compétences et de connaissances contextualisées
  - Effectuer des évaluations comparatives, équitables, neutres
  - Adapter les modèles aux différentes situations avant leur utilisation dans la pratique médicale quotidienne
  - Transparence des algorithmes
  - Intelligence artificielle « explicable »
- ⇒ Rapport complet, transparent des points clés des études basée sur des preuves scientifiques

Objectif de **TRIPOD-AI** Statement, en cours de développement

Collins G, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open* 2021;11(7):e048008.



Institut  
Sciences de  
la Santé Publique  
d'Aix-Marseille  
Aix-Marseille Université



## Artificial Intelligence for Public Health AI4PH

Institut  
Sciences de  
la Santé Publique  
d'Aix-Marseille  
Aix-Marseille Université

**SESStim**  
UMR 1262  
Sciences Economiques et Sociales de la  
Santé & Traitement de l'Information Médicale  
Inserm / IRD / Université AIX-MARSEILLE

DESU : ouvert en 2021  
M2 : ouverture en 2022  
<https://sesstim.univ-amu.fr>

