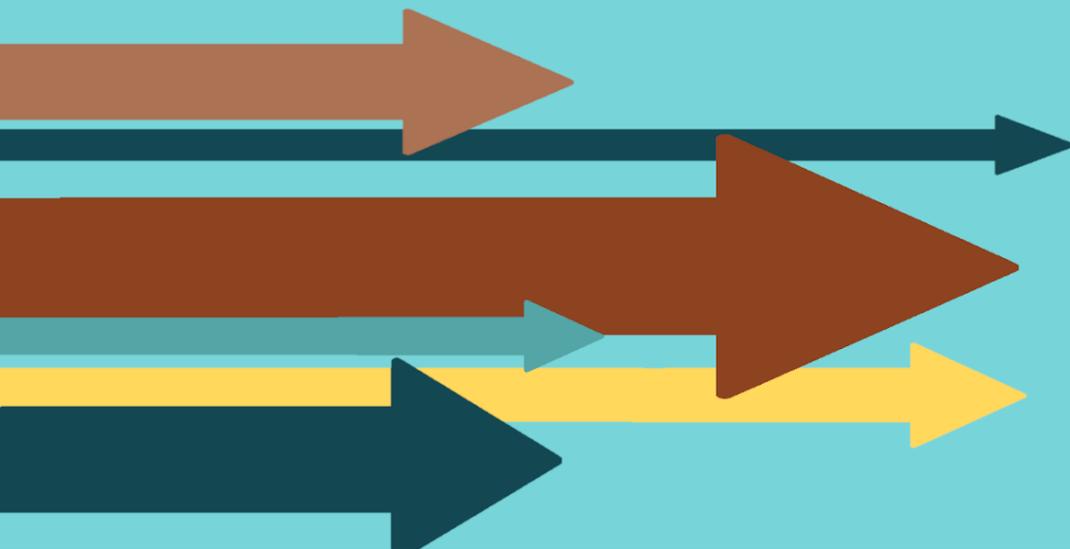




# La thèse d'exercice en médecine

Guide méthodologique



## Réalisation

### **Les internes de santé publique de Dijon :**

Manon Burgat, Nafiz Abdel Carime, Marie Corneloup, Emmanuel Delmas, Melchior de Giraud d'Agay, Magali Koczorowski, Maria Francesca Manca, Mickaël Piccard, Antoine Journé

## Groupe de relecture

### **Réseau d'aide méthodologique de Dijon, Département de médecine générale de l'UFR des Sciences de santé & Unité de Promotion Interne du CHU Dijon-Bourgogne:**

Christine Binquet, Maud Carpentier, Hervé Devilliers, Isabelle Fournel, Adrien Guilloteau, Anne-Sophie Mariet, Katia Mazalovic, Agnès Soudry-Faure, Claire Zabawa

## Direction

Professeure Christine Binquet

1	Introduction	3
2	Réglementation concernant les données de santé	4
3	Recherche documentaire/Outils bibliographiques	9
4	Définir la question de recherche	12
5	Système de gestion de base de données	14
6	Outils d'analyse des données quantitatives	20
7	Outils d'analyse des données qualitatives	22
8	La rédaction de la thèse	24

# 1 INTRODUCTION

---

Ce guide se veut être un outil d'aide méthodologique pour la rédaction de la thèse d'exercice des internes en médecine. Il contient des conseils utiles pour mener son travail de recherche au mieux, et il renvoie vers des ressources extérieures pour compléter les informations, par exemple des tutoriels, des documents, des sites internet.

L'élaboration d'une thèse suppose une série d'étapes qui sont présentées dans ce guide. Ces étapes sont précédées de la recherche et de la rencontre avec le directeur de thèse : cette rencontre est fondamentale, elle permet de définir le domaine d'étude, le calendrier, les moyens à disposition. Dans le cadre de la réforme du 3<sup>ème</sup> cycle, le directeur et le sujet de thèse doivent être définis autant que possible avant la fin de 1<sup>ère</sup> année d'internat (phase socle) et la soutenance doit être planifiée avant la fin de la phase d'approfondissement. Il est donc essentiel de prendre contact assez tôt avec les personnes ressources de votre spécialité.

En parallèle du travail sur les contenus de la thèse, il faut veiller à respecter les démarches et les échéances administratives qui s'échelonnent jusqu'au jour de la soutenance. Le dossier de soutenance de la thèse développé par la scolarité est disponible sur le site de l'UFR des sciences de santé (<https://sante.u-bourgogne.fr/actualite-des-tcem>).

Bonne lecture !

## 2 REGLEMENTATION CONCERNANT LES DONNEES DE SANTE

---

### *La protection des données*

Deux grands textes régissent les autorisations pour traiter les données dans le champ de la santé :

- **La loi Jardé**, ou loi sur la recherche impliquant des personnes humaines.
- **La loi informatique et liberté**, ou loi sur la protection des données.

Deux types de structures, visant au respect de ces lois existent :

<b>Comités de Protection des Personnes (CPP)</b>	Contrôle la conformité aux méthodes et pratiques de recherche référencées dans la loi Jardé. Il en existe au moins un par région.
<b>Commission nationale de l'informatique et des libertés (CNIL)</b>	S'assure que les équipes de recherche (où tout organisme stockant des données) respectent les règles de sécurité des données et les pratiques validées par la loi informatique et liberté.

En somme, si la recherche implique ces deux domaines, deux autorisations doivent être obtenues :

- **Une autorisation d'un CPP** (sélectionné au hasard lors de la demande d'autorisation) : ils sont tenus de répondre sous 45 jours à partir de l'examen en séance. Une absence de réponse équivaut à un refus. Ils peuvent exiger des modifications du protocole de recherche si nécessaire.
- **Une autorisation de la CNIL** : théoriquement tenue de répondre sous 2 mois. Une absence de réponse équivaut à un refus. La demande est renouvelable une fois. En pratique la réponse est souvent plus longue à obtenir, et il peut y avoir des échanges durant de longs mois avant d'obtenir une autorisation.

Vient s'ajouter pour les études portant sur des traitements ou des dispositifs médicaux, une autorisation supplémentaire de l'Agence Nationale de Sécurité du Médicaments, qui se doit de répondre à votre demande en 60 jours.

### Les démarches en fonction du projet

Voici un tableau de l'Inserm récapitulant les démarches réglementaires à suivre devant le type de recherche :

	Loi Jardé / Règlement européen sur les médicaments / Règlement européen sur les DM			Hors Loi Jardé		
	- Recherche interventionnelle (catégorie 1) - Essai clinique de médicament - Investigation clinique de DM	Recherche à risque et contrainte minimales (catégorie 2)	Recherche dénuée de risque (catégorie 3)	Recherche rétrospective  Dossiers ou échantillons disponibles	Recherche en SHS	Evaluation des pratiques professionnelles  Enquête de pratique
Obligations réglementaires	CPP ANSM CNIL  Assurance	CPP CNIL  Assurance	CPP CNIL	Anonymisation du fichier	Anonymisation du fichier	Anonymisation du fichier
Délais réglementaire (hors élaboration)	60 jours à compter du dépôt du dossier*			Aucun sous certaines conditions		

#### Quelques précisions !

- Les recherches de **catégorie 1 et 2** correspondent aux recherches interventionnelles, c'est à dire toute recherche impliquant un geste sur le patient (simples prélèvements compris) ou alors, impliquant un recueil de données sur la vie personnelle ou intime des sujets (ex : essais cliniques impliquant une évaluation d'une prise en charge ou d'un traitement, certaines études cas-témoins sur des sujets très sensibles (IST, alcoolisme et grossesse ...), etc..).
- Les recherches de **catégorie 3** sont les recherches dénuées de risque, (étude cas-témoins ou étude de cohorte n'impliquant aucun geste ni procédure supplémentaire par rapport à la prise en charge habituelle...).

Ex : Étude prospective, purement observationnelle, sans modification de la prise en charge des patients (même pas une « petite » prise de sang) et visant à identifier les facteurs de risque de survenue d'insuffisance rénale aiguë lors d'un traitement par amoxicilline haute dose

→ Les études sur des données existantes sans avoir à revenir au patient sont dite hors champ

Ex : les études rétrospectives monocentriques, les recherches de type méta-analyse et revue de la littérature.



### En pratique

- ★ Les recherches de catégorie 1,2 et 3 nécessitent TOUTES une autorisation d'un CPP.
- ★ Les recherches de catégorie 1 et 2 nécessitent une autorisation de la CNIL si elles ne rentrent pas dans le cadre des méthodologies de référence de la CNIL (données sensibles de type génétique, race, religion...[lien](#)).
- ★ Les recherches de catégorie 1 nécessitent en plus une autorisation de l'ANSM (Agence National de la Sécurité du Médicament).
- ★ Les recherches de catégorie 3 peuvent s'affranchir de l'autorisation de la CNIL si l'équipe de recherche s'engage à respecter une méthodologie de référence de la CNIL ([MR003](#)).
- ★ Les autres recherches s'appuyant notamment sur des enquêtes qualitatives n'ont pas à demander ces autorisations sauf si les questions peuvent induire un risque psychologique pour le patient. L'autorisation d'un CPP ou d'un comité d'éthique est alors requise.

Ces éléments ont des conséquences en terme de **durée de l'étude** :

- **Recherche interventionnelle (catégories 1 et 2)** → > 2 ans pour une valorisation
  - Préparation du protocole (environ 6 mois) ⇒ demande accord de promotion+/- financement- c'est ici que le projet a de grande chance d'être refusé pour un interne
  - Soumissions réglementaires CPP, + le cas échéant ANSM et/ou CNIL (*obligation d'avoir un promoteur* ⇒ 1 an)
  - Début du projet une fois toutes les autorisations obtenues;
  - 6-12 mois minimum pour le recueil et contrôle données ;
  - 6 mois minimum pour la valorisation (rédaction et soumission).
- **Recherche observationnelle (catégorie 3)** → Si l'étude est monocentrique, les seules véritables démarches sont d'obtenir l'autorisation d'un CPP, et de s'engager à respecter la méthodologie de référence MR003 de la CNIL et donc de traiter les données dans des

conditions permettant de garantir la confidentialité et la sécurité des données. Très peu de délai (1~3 mois) est donc nécessaire avant un possible recueil des données.

- Pour les études rétrospectives réalisées à partir des données déjà recueillies dans les dossiers médicaux informatisés ou non, les établissements périphériques et le CHU ont déjà signé des engagements. Les internes dépendant de ceux-ci et utilisant ces données n'ont donc pas besoin de l'autorisation du CPP ou de la CNIL sous réserve qu'ils traitent les données de façon anonymes.

→  Attention dans les deux derniers cas, si toutefois la recherche implique plusieurs centres/CHU ! Des délais supplémentaires peuvent s'ajouter car une demande à la CNIL s'impose et l'autorisation peut être longue à obtenir (>6-12 mois) car il faut démontrer le maintien de la sécurité et de la confidentialité des données à toutes les étapes du recueil et du traitement des données !).

### En conclusion

- ❖ Pour votre thèse d'exercice, on vous conseillera de mener une étude basée sur une recherche qualitative, rétrospective sur données existantes ou éventuellement de **catégorie 3** (et dans ces deux derniers cas monocentrique) afin de limiter les étapes réglementaires..
- ❖ L'autre possibilité est de travailler sur autre chose que des données patient : une revue systématique de la littérature ou une méta-analyse sont aussi des bons sujets de thèse ; les méta-analyses nécessitent toutefois un accompagnement par une personne expérimentée dans ce type d'étude car les analyses ne sont pas toujours très simples .
- ❖ Pour ceux qui veulent aller plus loin, et mener un projet de catégorie 1 ou 2 pour leur thèse, sachez qu'il faut s'y prendre très, très tôt. **Et avant d'entreprendre toutes démarches, adressez-vous OBLIGATOIREMENT à la DRCI (délégation à la recherche clinique et l'innovation) de votre CHU pour plus d'informations.** Si vous avez une appétence pour la recherche : envisagez un master 2 puis une thèse de science, et rapprochez-vous d'une équipe de recherche.

Dans tous les cas, mais en particulier pour les recherches cliniques prospectives, vous serez tenu de respecter les bonnes pratiques cliniques (BPC). Ces bonnes pratiques constituent un ensemble d'exigences éthiques et scientifiques pour la planification, la mise en œuvre, le suivi, le recueil des données, l'analyse et l'expression des résultats d'une étude. Si vous êtes amenés à participer à une recherche clinique dans le cadre de l'un de vos stages ou dans votre vie professionnelle, une formation

aux BPC vous sera demandée. Faites là donc le plus tôt possible. Des sessions de formation sont proposées régulièrement au CHU. Pour plus d'information contacter la DRCI : [recherche@chu-dijon.fr](mailto:recherche@chu-dijon.fr)

Photo by Danielle MacInnes on



### 3 RECHERCHE DOCUMENTAIRE/OUTILS BIBLIOGRAPHIQUES

---

**Objectif** : effectuer une synthèse des connaissances actuelles sur votre sujet. La recherche documentaire portera sur des articles et documents existants, recherchera les interrogations qui restent en suspens dans la littérature.



Une recherche sur Wikipedia ou Google, même si elle peut constituer un point de départ (utiliser alors au moins google scholar – cf. infra) n'est absolument pas suffisante dans le cadre d'une thèse d'exercice. Il est impératif de vous appuyer sur pubmed ± d'autres bases adaptées à votre contexte. N'oubliez pas d'intégrer les recommandations françaises et internationales qui existent dans le domaine qui vous intéresse. La Haute Autorité de Santé peut être une source intéressante d'information ([lien](#))

#### *Les règles de base de la recherche documentaire*

→ **On distingue la recherche documentaire simple de la recherche avancée.**

- La **recherche simple** est génératrice de beaucoup de "bruit". Les mots clés sont recherchés dans le titre mais aussi dans le résumé et le texte de l'article. Cette recherche est nécessaire pour un premier aperçu ; vous allez vite être amenés à cibler.
- La **recherche avancée** est en effet plus précise. Attention au silence documentaire si elle l'est trop. Vous pouvez essayer avec les opérateurs booléens, varier les langues et la date de publication des articles, rechercher des mots clés seulement dans le texte des articles.

→ **On peut avoir besoins de traduire des termes**

Il est utile de traduire sa recherche dans la langue de Shakespeare, pour ne pas se limiter aux publications françaises. Pour cela vous pouvez utiliser Hetop : <https://www.hetop.eu/hetop/>. Ce site vous aidera à trouver la bonne terminologie dans la langue souhaitée.

→ **Il existe des règles d'écritures dans les outils de recherche**

- ne pas mettre d'article devant le terme cherché
- entrer le terme au singulier
- entrer le terme au masculin
- éviter les accents

- utiliser les opérateurs Booléens



Les **opérateurs Booléens** sont des opérateurs logiques permettant de combiner des termes de recherche :

- ★ ET (AND)
- ★ OU (OR, THRU) - THRU (équivalent à THROUGH) sert à spécifier un ensemble de
- ★ valeurs (bornes incluses)
- ★ SAUF (NOT, AND NOT, EXCEPT)

### *Les bases de données*

**Plusieurs bases de données existent et ont chacune leurs spécificités.**

Pubmed
<p>C'est LA base de données bien connue des médecins. Elle est spécialisée en sciences médicales. Elle est également gratuite et libre. Une bonne utilisation de Pubmed est nécessaire pour être le plus pertinent possible dans sa recherche.</p> <p>Pour cela, ne pas hésiter à aller faire un tour sur :</p> <ul style="list-style-type: none"> <li>- Le power point de la BU de Paris Descartes sur Pubmed : <a href="#">lien</a></li> <li>- Le tuto PubMed et Mesh du Collège de liaison des internes de santé publique (ClISP) : <a href="https://clisp.fr/2016/01/15/pubmed/">https://clisp.fr/2016/01/15/pubmed/</a></li> </ul> <p> Le <b>Mesh</b> (Medical subject headings) est le thesaurus de Pubmed. Il permet d'interroger la base de données grâce à l'indexation des articles. En effet, tous les articles sont indexés de mots clés. Utiliser le Mesh améliore la pertinence de la recherche.</p>
Google Scholar
Moteur de recherche de Google permettant la recherche d'articles scientifiques quels qu'ils soient
Lissa
Base de données de la littérature scientifique en santé en langue française.
Cairn
Pour les sciences humaines et sociales en langue française.
Sudoc
Littérature grise= cette littérature correspond à l'ensemble des documents non publiés dans des journaux scientifiques (thèses, mémoires, ....)

Catalogue du système universitaire de documentation, vous pouvez trouver dans cette base notamment les thèses faites sur votre sujet et disponibles dans une autre BU française.

### CISMeF

Catalogue et index des sites médicaux de langue Française.

Vous pouvez rechercher des recommandations professionnelles, des thèses et mémoires, des documents d'enseignement et des documents grand public et associations des patients.

## La gestion de la bibliographie

Logiciel gratuit et simple d'utilisation, **Zotero** permet de stocker ses articles et de les éditer sur Word au format Vancouver. Vos références doivent, le plus souvent, être présentées selon les recommandations établies lors de la conférence de Vancouver. Cela permet d'uniformiser les données bibliographiques.

Exemple de référence aux normes de Vancouver :

O'Brien E, Parati G, Stergiou G, Asmar R, Beilin L, Bilo G, et al. European Society of Hypertension position paper on ambulatory blood pressure monitoring. J Hypertens. 2013 Sep;31(9):1731–68.

Pour cela, il vous faut installer Zotero + Connector (pour Mozilla ou Chrome) sur [zotero.org](http://zotero.org).

### Comment apprendre à bien gérer sa bibliographie ?

- Un tuto est disponible sur Youtube pour apprendre à utiliser Zotero : [lien](#).
- La BU Santé de Dijon propose divers **ateliers documentaires** :
  - ◆ [Zotero Késako ?](#)
    - Atelier 1 : Kit de survie
    - Atelier 2 : Biblio en un clic
  - ◆ [Débuter avec Pubmed](#)
  - ◆ [Réussir sans plagier](#)
  - ◆ [Netvibes : créer et organiser sa veille](#)
- De la documentation est disponible en ligne :
  - ◆ [Les logiciels de gestion bibliographiques](#) (dont tutos pour installer et utiliser Zotero)
  - ◆ [Plein d'autres tutos sur la recherche documentaire](#)
  - ◆ [La recherche documentaire en santé](#) (Cismef, EMC, BDSP)



### Petit plus

Si, après avoir lu et suivis nos liens et les divers ateliers de la BU, vous ne parvenez pas à faire une recherche digne de ce nom, n'hésitez pas à prendre rendez-vous avec le personnel de la BU de Dijon. Sur un créneau de 20 min à une demi-heure, ils seront ravis de vous donner un coup de pouce. Ecrire à [bump.ateliersdocs@u-bourgogne.fr](mailto:bump.ateliersdocs@u-bourgogne.fr) pour plus d'info.

## 4 DEFINIR LA QUESTION DE RECHERCHE

La recherche bibliographique permet de passer d'une thématique générale à la question de recherche ciblée à laquelle la thèse veut répondre. Au départ, vous avez un intérêt, une intuition ou une idée. Si le sujet est intéressant, s'il reste pertinent après votre recherche bibliographique et s'il correspond à un projet réalisable alors vous allez le problématiser.

Thématique	Question de recherche
Un traitement médicamenteux A de la SEP entraîne beaucoup de non observance.	<i>Quelles sont les facteurs de la non-observance du traitement A chez les patients atteints de SEP ?</i>
Les médecins généralistes connaissent peu le parcours des patients ayant des troubles des conduites addictives.	<i>Quels sont les leviers qui permettraient d'améliorer la connaissance du parcours de soin des patients ayant des troubles des conduites addictives par les médecins généralistes ?</i>
...	...

Poser la question de recherche permet d'établir le protocole de recherche, de choisir une méthode d'étude, de définir les données à recueillir. Les objectifs principal et secondaire(s) vont découler de votre question de recherche. L'objectif principal est celui qui permet de répondre à la question de recherche, et le(s) objectif(s) secondaire(s) sont des objectifs que l'on cherche à atteindre en parallèle de l'objectif principal. On ne pourra conclure que sur l'objectif principal d'une étude.

L'objectif de recherche doit être défini avec précision, doit être clair et atteignable : il se construit avec un verbe actif, et on y retrouve la population concernée par l'étude, ainsi que la période d'étude.

Question de recherche	Objectif(s)
<i>Quelles sont les étiologies de la non-observance du traitement A chez les patients atteints de SEP ?</i>	<ul style="list-style-type: none"> <li>- Déterminer les facteurs de la non-observance du traitement A</li> </ul>
<i>Quels sont les leviers permettant d'améliorer la connaissance du parcours de soin des patients ayant des troubles des conduites addictives par les médecins généralistes ?</i>	<ul style="list-style-type: none"> <li>- Décrire les leviers <ul style="list-style-type: none"> <li>- Analyser le parcours de soins</li> <li>- Décrire les acteurs du parcours</li> <li>- etc.</li> </ul> </li> </ul>
...	...



### Petit plus

Cette [vidéo](#) de l'Université de Genève (ndlr : aller regarder leurs [autres vidéos](#), également très instructives) présente brièvement comment poser une question de recherche.

## 5 SYSTEME DE GESTION DE BASE DE DONNEES

---

La création d'une base de données (bdd) exploitable est une activité qui prend du temps lorsqu'elle est réalisée correctement. Ce temps est nécessaire et c'est autant de temps de gagné pour la réalisation ensuite de l'analyse des données.

### *Les incontournables de la base de données*

#### La présentation de la base de données (bdd)

Il existe des règles pour la création et la présentation d'une bdd :

- **Une feuille** (que ce soit sous Excel ou un autre logiciel) ne contient **qu'un seul tableau** de données.
  - ◆ Si plusieurs tableaux doivent être fait pour correspondre par exemple dans une étude sur des mères et des enfants et que vous avez besoin de recueillir les données de façon séparées, il est impératif de les faire sur des feuilles différentes.
  
- La **première ligne du tableau** correspond aux **intitulés des colonnes** (= à quoi correspondent les données) et rien d'autre. Quelques règles pour créer les intitulés de colonnes :
  - ◆ Rester simple et court (max 8 caractères).
  - ◆ Limiter au maximum l'usage des majuscules.
  - ◆ Pas d'accent ou de caractère spécial (&, !, etc...) !
  - ◆ Si plusieurs mots, les séparer par un tiret bas (symbole \_ obtenu avec la touche 8 du clavier - ex : type\_chir).
  - ◆ C'est pratique d'avoir un dictionnaire des variables pour savoir à quelle variable correspond l'intitulé de colonne → on note la correspondance de chaque nom de colonne dans une nouvelle feuille (ou doc Word...). Les noms de variables ne sont pas tous logiquement compréhensibles par tout le monde.
  
- La **première colonne** correspond à l'**identifiant du patient**.
  - ◆ Définition : il s'agit d'un numéro d'identification unique pour chaque patient.
  - ◆ Ceci permet, en l'absence de données nominatives, de différencier les patients malgré l'existence possible de similitude sur l'âge, le sexe, l'année de naissance, .
  - ◆ La plupart des logiciels de gestion des données génèrent cette colonne automatiquement. Ce n'est pas le cas d'Excel.
  - ◆ Pour Excel : Saisir manuellement (non recommandé) ou générer dans la première colonne une vaste série de chiffre (plus que le nombre d'inclusion).

- **Jamais de lignes ou de colonnes *entièrement vides*** au milieu de votre bdd : la première colonne correspond à l'identifiant unique et anonyme de vos patients et la première ligne correspond aux intitulés de vos données.
- On **ne fusionne jamais** de cellules dans une bdd.
- **Une ligne** correspond à **un patient**.
  - ◆ Ne jamais placer les données de plusieurs patients sur une même ligne, cela complexifie grandement toute manipulation future.
- **Une cellule** contient **une seule donnée** et il est préférable que cette donnée soit numérique si vous voulez l'analyser ensuite sans trop de difficulté ; vous ne pouvez pas mettre du contenu numérique et en texte au sein d'une même colonne. Il faut que le codage soit homogène dans la totalité de votre colonne.
- **Jamais de totaux/sous totaux/graphiques** au sein d'une feuille contenant la bdd.
  - ◆ Toujours sur une autre feuille.

## Assurer la sécurité des données

- **Le plus important : anonymiser vos données !**
  - ◆ **Jamais d'identité** du patient dans la base sur laquelle vous travaillez (c'est illégal).
  - ◆ Enlever des tableaux toute référence au nom, prénoms, date de naissance et adresse.
  - ◆ Identifier chaque patient à l'aide d'un numéro unique (cf voir plus haut).
  - ◆ **On ne peut pas garder des données de santé avec les données d'identité sur le même système de sauvegarde.** Par exemple, dans un laboratoire de recherche ou un registre, les base de correspondance entre identité du patient et identifiant sont sécurisée et conservées dans un autre lieu que le serveur qui héberge les données de santé. Pour votre thèse vous pouvez conserver la liste de correspondance sur votre lieu de recueil par exemple dans un tiroir fermé à clé s'il s'agit d'une liste papier, ou sur un PC, sécurisé par mot de passe.
- Si possible, protégez vos données par un **mot de passe**.
  - ◆ Tous les logiciels n'ont pas nécessairement cette option.

- Si des données de santé voyagent (clef USB, PC portable) : elles doivent être déjà anonymisées et idéalement stockées dans un dossier crypté.
- Une fois vos données exploitées, **détruisez-les**.
  - ◆ Si nécessaire, archiver l'original dans un endroit défini avec votre directeur de thèse.
  - ◆ Mais détruire toutes les autres copies.
- On ne stocke JAMAIS les données sur un cloud (Google drive et autres) et on ne les envoie pas par mail.
  - ◆ Éviter de stocker des données sur une clé USB/disque dur externe (trop longtemps).

### COMMENT PROTÈGE-T-ON DES DONNÉES ?

#### Crypter un dossier

- ★ Procédure Windows : [lien](#)
- ★ Procédure Mac : [lien](#)

#### Protéger un dossier par un mot de passe

- ★ Procédure Windows : [lien](#)
- ★ Procédure Mac : [lien](#)

### *Contrôle à la saisie : principes de base*

L'objectif du contrôle à la saisie est d'aider la saisie de l'information afin d'homogénéiser au maximum les données qui seront analysées.

Par exemple :

- Quand on rentre le poids, on le rentre tout le temps en kg ou tout le temps en g mais pas une fois en kg et une fois en g.
- Quand on veut dire que ce patient a subi une intervention et un autre non, on crée une colonne 'Interv' et on aura 2 modalités de remplissage de cette colonne :
  - 0 = Pas d'intervention (ou « PasdInterv »)
  - 1 = intervention (ou « Intervention »)
    - On ne met pas une fois « Intervention » avec majuscule
    - Une fois « intervention » sans majuscule
    - Une fois « traité »
    - Une fois « non traité »
    - Une fois « 0 »

En fait, à chaque fois qu'on écrit une modalité de manière différente, le logiciel de statistique



va comprendre qu'il s'agit d'une modalité différente et non pas juste d'une variante d'écriture.

### **En pratique, comment on contrôle la saisie sur Excel ?**

Pour **Excel**, il est fortement conseillé de s'aider de l'outil "**Validation des données**" (VDD) dans l'onglet "Donnée". Cette fonction permet de limiter les valeurs qui pourront être entrées dans telle ou telle cellule. Si on entre dans une cellule une valeur autre que celle qui a été paramétré, le système empêche toute saisie ultérieure jusqu'à correction de l'erreur.

Vous pouvez choisir si ce sera du texte ou des chiffres, définir un intervalle de valeur, définir une liste de termes... Les applications sont très nombreuses.

[Liens vers un tutoriel pour la VDD](#)

Une autre fonction pratique est "**Protéger la feuille**" dans l'onglet "Révision". Elle permet de rendre la page sélectionnée non modifiable jusqu'à la levée de la protection. Vous pouvez définir ou non un mot de passe pour cette levée, permettant ainsi de définir qui est autorisé à modifier la feuille.

Malgré tout, il ne faut pas se leurrer, Excel n'est pas un logiciel initialement conçu pour la saisie sécurisée des données. Les éléments cités plus haut permettent de limiter les gros accidents, mais ils nécessitent d'être vigilant dans leur configuration pour éviter d'autres couacs. Notamment si vous devez trier vos données, faites une sauvegarde avant, et faites attention à bien prendre en compte toute la plage de données ; sinon vous risquer de perdre la correspondance patients/données. Le mieux reste encore d'utiliser un logiciel conçu pour la saisie des données (voir infra).



### **Petit plus**

Exemple de tutoriels pour diverses fonctions Excel (il en existe de nombreux autres sur le net) :

- ★ <https://www.pcastuces.com/pratique/astuces/1187.htm>
- ★ <https://www.youtube.com/watch?v=4zRTng2U7Gs>

### **Et les autres logiciels ?**

#### **Des logiciels similaires : LibreOffice BASE/OpenOffice BASE**

Il sont les équivalents gratuits du logiciel Access de Microsoft. Installer Java est nécessaire à leur bon fonctionnement. Téléchargement : [LibreOffice](#) et [OpenOffice](#).

Tutoriel : [Tutoriel tout en image](#). Bien que l'exemple traité ne touche pas au domaine de la santé, ce petit tutoriel montre pas à pas, en image, toutes les fonctionnalités basiques du logiciel, que ce soit la construction d'une table, la création d'un formulaire de saisie et l'émission d'un rapport selon des critères définis. Il y a même des tutoriels plus avancés pour ceux qui souhaiteraient aller plus loin.

## Les logiciels spécialisés : Epi-info et Wepi (épiconcept)

Certains logiciels de création de base de données permettent la création d'une bdd à partir de plusieurs formulaires de saisie (le top !). Il est nécessaire de créer un formulaire de saisie des données (comme un questionnaire *googleform* mais en privé) qui pourra ensuite être rempli pour chaque patient. La bdd se met automatiquement à jour à chaque nouveau remplissage de formulaire de saisie.

### **Epi-info**

Logiciel gratuit, créé par le Centers for disease control and prevention (CDC), disponible en français.

Téléchargement : [https://www.cdc.gov/epiinfo/support/fra/fr\\_downloads.html](https://www.cdc.gov/epiinfo/support/fra/fr_downloads.html)

Tutoriels :

- Tutoriel sur Youtube : <https://www.youtube.com/watch?v=TH-LNGZUDh4>
- Epi Info™ 7 User Guide : <https://www.cdc.gov/epiinfo/pdfs/UserGuide/EI7Full.pdf>

### **Wepi (épiconcept)**

Vous pouvez en trouver la description et les explications sur le site <https://www.wepi.org/>.

C'est sécurisé, français, développé par et pour des épidémiologistes. Il y a un accès à un ensemble de fonctionnalité pour créer et gérer un questionnaire en ligne, un peu comme dans Google forms sauf que là les données sont sécurisées. C'est gratuit, et grâce aux tutoriels on arrive à faire beaucoup de choses.

Tutoriels téléchargeables à partir de <http://www2.voozаноо.net/fr/node/29> - aller à Wepi

**Un mot sur les autres options surtout si vous avez déjà des connaissances en programmation ou sur les systèmes de gestion de base de données ou si vous souhaitez développer des compétences dans ce domaine**

### **SQL**

L'emploi de ce type de logiciel demande une formation spécifique (cf formation à l'université – SAS, formation sur internet de type MOOC ...etc.).

### **Microsoft Access**

Un grand classique pour la création et la gestion de base de données. Très complet, il permet la création de tables, formulaires et requêtes. Toutefois cela reste un logiciel payant et pouvant être bien plus difficile à prendre en main que de prime abord.

Téléchargement (payant) : <https://products.office.com/en/access>

Tutoriel : <https://www.youtube.com/watch?v=4FXcl0hmn40>

En résumé, pour la gestion de votre BDD, les choix principaux sont les suivants :

- Excel : simple mais peu fiable ; fonction tri des données dangereuse
- Epi-Info : prise en main plus lente qu'Excel, mais plus rapide qu'Access ; permet de faire directement des analyses statistiques ; à préférer si >50 patients et/ou 15-20 variables
- Wepi : très complet en terme de création de formulaires de saisie et accessible en ligne ( ! ) ; ne permet que de recueillir des données
- Access : complexe, mais fiable ; très utile pour des bdd complexes (plusieurs suivis par patients, plusieurs mesures à chaque temps )

## 6 OUTILS D'ANALYSE DES DONNEES QUANTITATIVES

Il existe différents logiciels d'analyse statistique gratuits et accessibles, présentés dans le tableau ci-dessous.

Logiciel et site	Auteurs	Plateforme	Analyses	Type d'interface	Aide
<b>pvalue</b> <a href="#">lien</a>	<i>Kevin Zarca</i>	Application web	Analyses descriptives, univariées, multivariées, courbes de Kaplan Meier, analyses de survie, gestion des données manquantes	Application web avec interface sans programmation	<a href="#">Site de l'auteur</a>
<b>jamovi</b> <a href="#">lien</a>	<i>The jamovi project</i>	Windows Mac OS Linux	Statistiques descriptives Tests bivariés Analyses de variance Analyses factorielles, en composantes principales Régression linéaire Régression logistique Analyses de survie	Logiciel avec interface sans programmation	<a href="#">Manuel</a> du logiciel (anglais)  <a href="#">Chaine youtube</a> (anglais)
<b>AnaStats</b> <a href="#">lien</a>	<i>AnaStats Scop ARL</i>	Microsoft office excel	Statistiques descriptives Tests bivariés Analyses de variance Test de concordance (Kappa) Coefficient alpha de Cronbach	Utilitaires sous formes de fichiers excel	Chaque utilitaire contient une description du test et le manuel d'utilisation

<b>R</b> <a href="#">lien</a> et <b>R Studio</b> <a href="#">lien</a>  (installer R, puis R Studio)	<i>GNU Project</i>	Windows Mac OS Linux	Toutes les statistiques possibles et imaginables	Logiciel avec interface et programmation	<a href="#">Vidéos YouTube</a> du MOOC « Introduction à la statistique avec R » par le Pr Bruno Falissard – Université Paris Saclay
---	--------------------	----------------------------	--	--	---

Ne téléchargez pas tous ces logiciels ; faites un choix en fonction de vos besoins. La plupart du temps vous n'avez besoin que d'un logiciel réalisant les descriptifs classiques, et les analyses bivariées. Si une analyse multivariée est proposée par votre directeur, demandez-lui de l'aide pour sa réalisation (ou de vous diriger vers une personne ressource). Un certain nombre de personnes sont formées aux statistiques au sein des services cliniques ; ils pourront vous aider avec les logiciels dont ils ont l'habitude mais qui ne sont pas évoqués dans le tableau ci-dessus car non disponibles sans frais (SAS, Stata, SPSS...) ou trop complexes si on ne les utilise pas au quotidien.

Excel est pratique pour les données descriptives, après c'est vraiment galère (et pas toujours fiable). Donc à éviter !!!

Vous pouvez également utiliser [BiostaTGV](#) pour le calcul de votre nombre de sujet nécessaire (NSN).

## 7 OUTILS D'ANALYSE DES DONNEES QUALITATIVES

---

### *L'analyse qualitative en bref*

L'analyse qualitative traite les données issues des entretiens individuels, des focus groupes, des observations sur le terrain. Pour les données qualitatives, les règles d'anonymisation et de protection de données sont les mêmes que pour les données quantitatives.

L'analyse qualitative peut être inductive, quand les thèmes émergent des données, ou déductive, quand on applique une théorie aux données recueillies.

L'analyse devrait démarrer dès le premier entretien, focus groupe ou observation, pour pouvoir adapter le guide d'entretien ou la grille d'observation, mais aussi pour déterminer quand le recueil des données peut être arrêté (saturation des données).

Il existe de nombreuses approches à l'analyse qualitative (analyse thématique, théorie ancrée, analyse phénoménologique...) mais en général les étapes de l'analyse qualitative sont les suivantes :

- Codage des verbatims (=la transcription mot par mot des entretiens, y compris le non verbal, comme les rires, les pauses...)
- Regroupement des codes dans des thèmes
- Description des thèmes (analyse descriptive)
- Interprétation des données
- Comparaison avec les données issues de la littérature, formulation d'hypothèses, de recommandations...

D'un point de vue méthodologique (et idéal), il est intéressant de réaliser un double codage sur une partie des verbatims. On peut par exemple demander à une autre personne d'analyser un verbatim (codage avec ses propres codes, son propre regroupement, sa propre interprétation). Cela permet d'avoir un autre regard et de fournir deux éléments :

- une réassurance sur le fait que notre interprétation se recoupe avec celle d'une personne différente;
- un regard forcément divergent sur certains points (ajout de codes différents, différence d'interprétation), et qui permettra d'enrichir l'ensemble de l'analyse avec des éléments pertinents qui n'auraient pas été remarqués.



Un [article d'Exercer](#) présente une introduction à la recherche qualitative.

## Quels outils pour l'analyse qualitative ?

### Les logiciels généraux de traitement de texte : Word et similaires

L'analyse de données qualitatives peut être faite manuellement, sur Word ou sur un autre logiciel (ex. suite Libre Office, Open Office...). Dans ce cas, vous pourrez copier-coller des parties de texte dans les cases des thèmes correspondants, ou surligner le texte avec des couleurs différentes selon le thème...

### Les logiciels spécialisés

Logiciel et site	Auteurs	Plateforme	Analyses	Interface	Aide
<b>QDA Miner Lite</b> (gratuit) <a href="#">lien</a>	Provalis Research	Windows	Données textuelles Images fixes	Interface graphique simple disponible en français	Sur le <a href="#">site</a> et sur la <a href="#">chaîne Youtube</a> de l'auteur
<b>CATMA</b> (gratuit) <a href="#">lien</a>	Université de Hambourg	Web application	Données textuelles : analyse, annotation, modèles, possibilité de collaborer	Web application en anglais	Sur le <a href="#">site</a> de l'auteur
<b>ATLAS.ti</b> (un mois d'essai gratuit) <a href="#">lien</a>	ATLAS.ti	Windows Mac	Données textuelles, photos, audios, vidéos...	Interface graphique simple	Sur le <a href="#">site</a> de l'auteur
<b>Dedoose</b> (un mois d'essai gratuit) <a href="#">lien</a>	Dedoose	Windows Mac	Données textuelles, photos, audios, vidéos...	Interface graphique simple (anglais)	Sur le <a href="#">site</a> de l'auteur

NVivo est en train de devenir la référence, mais il est payant. Pour limiter son coût vous pouvez acheter une licence étudiant pour 6 mois ou un an... et l'inclure dans vos frais réels lors de votre déclaration de revenus au service des impôts.

## 8 LA REDACTION DE LA THESE

---

La thèse consiste en un mémoire ou un article dactylographié rédigé en français ou en anglais.

Elle se compose de différentes parties :

### ➤ Introduction

Le but de cette première partie est de poser le cadre dans lequel s'inscrit votre travail de recherche. Partant de la présentation du cadre général, vous focalisez au fur et à mesure votre argumentaire vers le problème qui vous intéresse, celui qui sert de justification à votre étude.

C'est ici que doit être valorisé une bonne part de votre travail bibliographique. Le lecteur peut ainsi s'assurer que l'auteur est à jour dans les connaissances liées au sujet et que l'objet de l'étude repose sur des bases solides.

### ➤ Objectif

Cette courte partie sert à trois choses :

1. D'abord exposer l'objectif (ou la question) principal de l'étude. Une étude ne peut avoir **qu'un seul objectif principal**. Le travail présenté est principalement conçu pour répondre à cette question.
2. On peut associer, suivant les justifications qu'on apporte, un ou plusieurs objectifs secondaires. L'idée étant que même si le travail présenté n'a pas été conçu spécifiquement pour répondre à ces questions, il peut apporter quelques éléments de réponses.
3. Enfin présenter le schéma d'étude employé pour répondre à l'objectif principal.

Le maître mot de cette partie est d'être **clair et concis**. Le lecteur doit pouvoir identifier en très peu de temps la problématique de votre recherche et les grandes lignes de la méthode employée pour y répondre.

## ➤ Méthodes

Ce chapitre très structuré sert à décrire en détail les processus et les outils employés pour mener à bien votre recherche. On peut la diviser en plusieurs sous-parties :

### **1. Décrire la source des données et estimer le nombre de sujets nécessaires**

Si votre étude exploite un échantillon d'individus, il faut décrire ce dernier et la population dont il est issu. Si un échantillon sous forme de données patients est extrait d'une base de données, il faut décrire cette base et présenter rapidement son contenu. Si l'étude est une revue de la littérature ou une méta-analyse, il faut citer et présenter les bases documentaires employées pour récupérer vos articles ainsi que la stratégie de recherche (mots clés, langues, période → tout doit être tracé).

→ Quelle que soit la situation, il faut toujours apporter une justification sur l'intérêt de l'usage de telle ou telle source. Le lecteur peut ainsi commencer à évaluer si les données employées ne sont pas biaisées et jusqu'à quel point les résultats de l'étude sont généralisables dans un autre environnement.

Il faut aussi que vous présentiez votre estimation du nombre de sujets nécessaires, si vous travaillez sur des individus (ou données patients).



Le **nombre de sujets nécessaires** est le nombre minimum d'éléments à inclure dans votre échantillon afin d'avoir les effectifs nécessaires pour répondre à votre objectif de recherche. Ces estimations sont réalisables par la plupart des logiciels statistiques (et certains logiciels spécialisés dans le domaine).

Dans le cas d'une recherche qualitative, le recueil de données s'arrête lorsque les nouvelles données n'apportent plus d'informations (=saturation). Pour cette raison, l'analyse et le recueil des données sont un processus continu en recherche qualitative.

## 2. Sélection des éléments à inclure - Critères d'inclusion et non inclusion

Sont décrites ici les modalités selon lesquelles vous incluez un individu/dossier patient/article à votre étude. Ceci passe par la définition claire des critères d'inclusion et de non inclusion.

<b>Critères d'inclusion</b>	Correspondent aux caractéristiques nécessaires pour que l'élément puisse être inclus dans l'étude. Tous les critères doivent être présents pour que l'élément soit inclus.
<b>Critères de non inclusion</b>	S'emploient sur les éléments n'ayant pas rempli les conditions citées au-dessus. Si au moins un des critères de non inclusion est présent, l'élément n'est finalement pas inclus dans l'étude.

Le lecteur peut alors estimer si les éléments nécessaires au bon déroulement de votre étude ont été inclus et si la plupart des éléments pouvant porter préjudice (biais) ont été écartés.

## 3. Critères de jugement

A chaque objectif défini correspond un critère de jugement. Il s'agit de la variable dont la description/comparaison précise au cours de l'analyse permet de répondre à votre objectif d'étude. On distingue :

- Le **critère de jugement principal** qui permet de répondre à l'objectif principal. En général, il n'y en a qu'un.
- Les critères de jugement secondaires qui permettent de répondre aux autres objectifs.

Il faut définir clairement et décrire précisément chacune de ces variables et leurs modalités de mesure. Bien sûr, il faut aussi justifier pourquoi ce sont ces variables qui ont été choisies. C'est de celles-ci que dépendent les conclusions de votre étude.

#### **4. Les autres données recueillies**

Sont décrites ici toutes les catégories des caractéristiques/variables que vous avez recueilli pour obtenir les résultats de votre étude. N'hésitez pas à fournir plus de détails sur les variables présentant le plus d'intérêt à vos yeux.

Ce descriptif fonctionne toujours sur le même principe : permettre au lecteur de s'assurer que les données pertinentes ont été récupérés / que les éléments superflus ont été écartés.

#### **5. Mode de recueil des données**

Sont détaillées les méthodes utilisées pour récupérer vos données.

Ex : Pour des individus, cela peut être par un interrogatoire au cabinet et/ou par consultation du dossier patient ; pour des données issues d'une base, la méthode d'extraction et les modalités choisies pour sélectionner les bons dossiers ...

#### **6. Le plan d'analyse des données**

Pour finir, on détaille les méthodes employées pour exploiter et analyser les données, au regard des critères de jugements définis. Vous devez préciser le plan d'analyse adopté pour répondre à chacun des objectifs que vous avez fixés, répondre à l'objectif principal restant la priorité.

Dans un premier temps, ceci passe toujours par un descriptif des variables recueillies au sein de notre échantillon en fonction de leur distribution (moyenne  $\pm$  écart-type ou médiane et intervalle interquartile pour les variables continues ; proportion/fréquence pour les variables en catégories etc.). C'est ensuite **et seulement dans la mesure du nécessaire** qu'on réalise des analyses plus complexes (comparaisons simples - analyses bivariées, comparaisons plus complexes - analyse multivariée, etc.).

Soyez clair dans la description de votre cheminement et précis dans les termes employés. Une analyse bien argumentée et structurée est pour le lecteur un gage de résultats de qualité.

#### **7. Considérations éthiques et réglementaires**

Dans le cas où vous menez une étude prospective ou une enquête qualitative, il est important d'indiquer les modalités d'information et d'obtention du consentement (ou de la non opposition) des

participants et les autorisations que vous avez obtenu (ou qui ont été obtenues par l'équipe qui gère l'étude en pratique). Si vous avez reçu un avis d'un comité d'éthique ou de la CNIL, mettez-le.

## ➤ Résultats

Dans cette partie très structurée, vous présentez de façon exhaustive les résultats de vos analyses sans les interpréter. Le but ici est de donner aux lecteurs un compte-rendu de vos données et analyses dépourvus de votre interprétation. Soyez donc **objectif** et réservez votre avis pour la partie suivante.

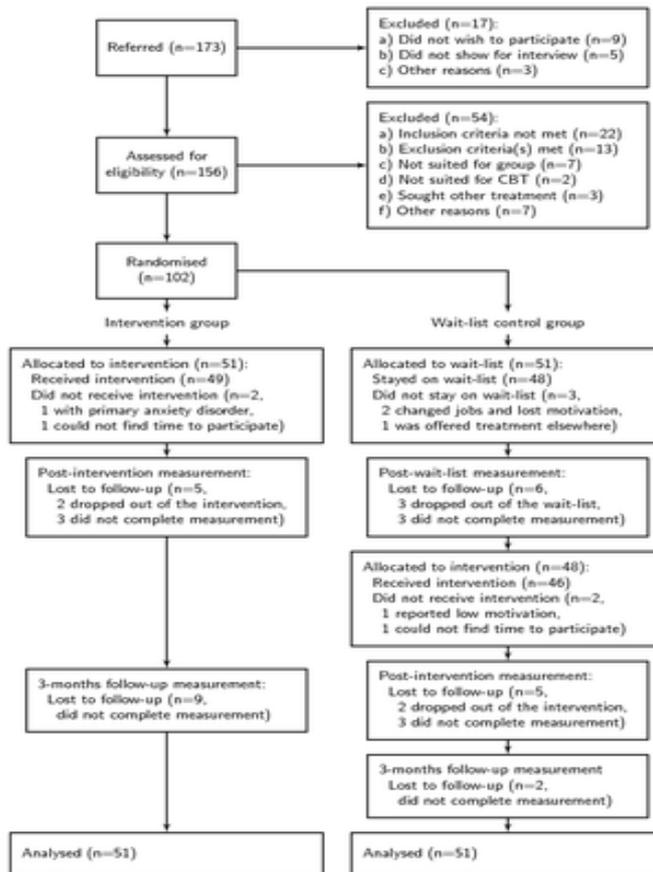
La plupart des figures et tableaux sont aussi inclus dans cette partie. Ces éléments permettent de présenter de façon synthétique vos résultats et de concentrer votre rédaction sur les éléments les plus pertinents.

Cette partie se divise en plusieurs sections :

### *1. Description de votre échantillon*

Commencez d'abord par présenter la construction de votre échantillon étape par étape. Il n'est pas rare que, partant d'un effectif initial, vous ayez été obligé de retirer de l'étude un certain nombre d'éléments. Il faut préciser pour chaque retrait les motifs de ce dernier. Le mieux est de passer par la présentation d'un diagramme de flux (ou Flow chart) qui schématise ce processus.

Figure 1: Flowchart of participants' progress through the phases of the trial



[Source](#)

Puis vous allez décrire votre échantillon à proprement parlé. Ceci passe par la réalisation d'un tableau contenant une description de votre échantillon à partir des distributions des variables recueillies. Ces distributions peuvent être décrites en terme d'effectif et de pourcentage pour les variables en catégories (on parle aussi de variable qualitative) ou de moyenne  $\pm$  écart-type ou médiane et intervalle interquartile (25<sup>ème</sup> percentile – 75<sup>ème</sup> percentile) pour les variables continues (dites aussi quantitatives) selon que leur distribution est normale (gaussienne) ou non. Pour détecter une distribution non gaussienne, il suffit de regarder la moyenne et la médiane. Si ces deux statistiques sont éloignées alors il y a de grande chance pour que la distribution ne soit pas normale.

Le texte accompagnant ce tableau devra commenter les constats principaux et non reprendre la totalité des informations du tableau.

## **2. Présentation des analyses en lien avec l'objectif principal**

Commencez toujours par présenter les résultats issus des analyses les plus simples avant de commenter les résultats plus complexes. N'hésitez pas à vous aider de graphiques et tableaux si nécessaire.

Les résultats principaux doivent être présentés le plus clairement possible. Le lecteur doit pouvoir les identifier rapidement dans votre argumentaire.

## **3. Présentation des autres analyses**

Une fois les résultats répondant à l'objectif principal choisis, vous pouvez faire de même pour les secondaires. Les mêmes conseils généraux sont applicables ici.

### **➤ Discussion**

Probablement la partie la plus complexe, mais aussi la plus intéressante à rédiger. Elle correspond à l'interprétation de vos résultats et à leur mise en relation avec les données de la littérature.

Rédigez d'abord quelques lignes sur vos résultats principaux et leur interprétation. Mettez-les en perspective par rapport aux découvertes déjà faites sur le sujet. Si, à la lumière de ce qui a été avancé vous avez des hypothèses, avancez-les prudemment ici.

Présentez ensuite de façon honnête et complète les forces et les faiblesses de votre étude. Mettez vos atouts en valeur. Si vous parvenez à justifier en quoi les faiblesses de votre étude ont peu d'impact sur la validité des résultats obtenus, c'est encore mieux (mais dans l'absolu pas toujours possible).

### **➤ Conclusion**

Enfin finissez votre rédaction par quelques lignes synthétisant vos résultats principaux et les possibilités éventuelles d'aller plus loin grâce aux notions abordées dans votre étude.

## Nota bene !

La scolarité demande que les conclusions soient signées par le président du jury et par le doyen de l'UFR au moins un mois avant la soutenance de la thèse. Ces conclusions ne correspondent pas à la courte conclusion que nous trouvons généralement à la fin des articles : **la conclusion** met en exergue en quelques lignes ce qui a été montré dans l'étude et ouvre sur les perspectives.

**Les conclusions** pour le doyen sont plus longues (une page, une page et demi) et résument les différentes parties de la thèse.

Un **résumé** doit aussi être inséré dans le dossier de thèse : il correspond à un abstract et il doit être accompagné par des **mots-clés** (à choisir en fonction de leur pertinence ou de la terminologie en vigueur dans les disciplines biomédicales - MESH). Il doit être signé par le président du jury et il doit lui aussi être déposé un mois avant la soutenance.

Pour la constitution du dossier de thèse et les échéances administratives, vous pouvez télécharger le dossier de soutenance de thèse à ce lien :

<https://sante.u-bourgogne.fr/actualite-des-tcem>.

Pour aller plus loin, vous pouvez aussi consulter l'ouvrage « initiation à la recherche de Paul Frappé : <https://www.decite.fr/livres/initiation-a-la-recherche-9782919616275.html>; il est disponible à la BU.

# Bon travail !

Réalisé par les internes de santé publique de Dijon – octobre 2023

