

Services d'Aide Médicale Urgente

Enjeux et défis

Ce chapitre a pour vocation de présenter l'espace des problèmes ainsi que la richesse des sujets traités et à traiter dans les services d'aide médicale urgente (SAMU). Dans un premier temps, nous présentons la diversité des systèmes de santé à l'échelle internationale. Nous proposons plusieurs critères sur lesquels nous appuyer pour caractériser ces systèmes. Dans un second temps, nous décrivons le fonctionnement détaillé de notre cas d'étude : le modèle français. Enfin, nous nous intéressons aux problèmes spécifiquement rencontrés par les SAMU et la manière dont la communauté scientifique s'y est intéressée et les a adressés.

1.1 La diversité des services de soins d'urgence

Les services de soin d'urgence diffèrent d'un pays à un autre et parfois même au sein d'un même pays. Dans cette partie, nous nous intéressons à ces différences et à la manière de les caractériser.

1.1.1 Des modèles formels et informels

En fonction du niveau de développement d'un pays, la maturité du système de santé ne sera pas la même. En effet, cela va dépendre des infrastructures à disposition, du niveau de vie et d'éducation de la population, de l'accessibilité au système de soin, de l'équipement disponible, du personnel formé, etc. C'est pourquoi (Sikka et al., 2005) proposent de distinguer deux types d'EMS (Emergency Medical System) : les modèles informels et formels.

Les modèles informels se caractérisent par une absence de système pré-hospitalier. Il n'y a pas d'organisation établie pour la réponse à l'urgence ou alors ce qui existe est très rudimentaire. On constate plusieurs niveaux dans les modèles informels : un premier avec une absence totale d'organisation pré-hospitalière où le patient

doit s'appuyer sur sa famille, ses amis, ou toute autre personne en dehors de ses proches pour le transporter vers un établissement de santé en cas de besoin. Certains systèmes évoluent lorsque des acteurs dont le rôle premier n'est pas le soin d'urgence commencent à porter secours. Ce peut être le cas lorsque les pompiers ou la police commencent à transporter des patients vers l'hôpital alors que ce n'est pas leur cœur de métier (Mock et al., 2002). Un premier pas dans l'amélioration des systèmes informels consiste à commencer à former des citoyens à l'usage des premiers secours.

Les modèles formels se caractérisent par une organisation coordonnée entre plusieurs ressources capables de répondre et assurer le transport des patients depuis le terrain jusqu'à l'établissement de santé adapté à son besoin lorsque cela est nécessaire. La plupart comprennent un numéro d'accès, un système de prise d'appels, un système de répartition, de réponse, de soin et de transport.

Dans la suite de ce manuscrit, nous réduisons notre périmètre à l'étude des systèmes formels qui disposent de moyens organisés et coordonnés pour la gestion des appels d'urgence.

1.1.2 La chaîne des urgences : Amont et Aval

L'enchaînement des actions faisant intervenir plusieurs ressources afin de porter secours au patient s'appelle **la chaîne des urgences** (Yersin, 2006). Celle-ci peut être décomposée en plusieurs étapes. Sikka et al. proposent une décomposition en quatre étapes (Sikka et al., 2005) comme suit :

1. Notifier les autorités compétentes lorsqu'il y a une urgence.
2. Réceptionner cette information puis envoyer des moyens adaptés pour porter secours à l'appelant.
3. Appliquer les soins de première nécessité sur place.
4. Transporter le patient dans une structure adaptée disposant de l'équipement et des équipes nécessaires pour traiter le patient.

Toutefois, nous avons préféré retenir pour nos travaux la décomposition en 8 étapes issues du référentiel métier du système français qui nous semble plus précise (France et al., 2015) (cf. figure 1.1). Dans cette figure, nous présentons notre vision de la chaîne des urgences qui vient préciser celle proposée par Sikka et al. et qui s'appuie sur le référentiel métier du système français.

1. **Appeler** : Composer un numéro d'urgence qui notifie les autorités compétentes lorsqu'il y a urgence.
2. **Qualifier et Orienter** : Réceptionner cette information, lui affecter une priorité et transférer l'appel au praticien adéquat.
3. **Réguler** : Diagnostiquer et décider des moyens adaptés pour porter secours.
4. **Engager des moyens** : Mobiliser les moyens adaptés et transmettre les informations aux équipes de secours.
5. **Rejoindre et Stabiliser** : Se rendre auprès du patient et stabiliser son état.

6. **Orienter le patient** : Identifier l'établissement de santé le plus adéquat pour envoyer le patient en prenant en compte les places disponibles ainsi que la volonté du patient.
7. **Transporter le patient** : Transporter le patient jusqu'à l'établissement de santé choisi.
8. **Hospitaliser** : Procéder à l'admission du patient dans le service idoine.

Dans cette chaîne des urgences, on identifie deux macro-étapes. Une première qui concerne la réception, le tri et l'orientation de l'appel, jusqu'à la décision d'envoyer des moyens. Toutes ces activités ont lieu au niveau du centre d'appels. En France, ce dernier est appelé Centre de Réception et de Régulation des Appels (CRRA). Nous nommerons cette partie **l'amont de la chaîne des urgences**. La deuxième macro-étape se compose de tout ce qui concerne l'intervention sur place, le transport jusqu'au patient, sa stabilisation ainsi que son transfert jusqu'à ce qu'il soit admis dans un service de santé si cela est nécessaire. Nous nommerons cette partie **l'aval de la chaîne des urgences**.

La gestion du soin urgent peut être décomposée en deux macro-étapes : l'amont et l'aval. Elles différencient les activités au niveau du centre d'appels et les actions des équipes de secours sur place. Au sein de ces deux grandes catégories, on trouve plusieurs décompositions possibles avec un niveau de détail pouvant être différent.

Dans la suite de cette étude, nous nous sommes appuyés sur la décomposition du référentiel métier français et nous nous sommes plus spécifiquement intéressés aux différences que l'on peut retrouver au sein de l'Union Européenne (UE).

1.1.3 L'accès aux services d'urgences : numéro unique ou numéros spécialisés ?

La première étape de la chaîne des urgences est donc *Composer un numéro d'urgence qui notifie les autorités compétentes*. Cette première étape présente déjà des différences notables d'un pays à un autre. L'*European Emergency Number Association* (EENA) est une organisation non gouvernementale dont la mission est de contribuer à

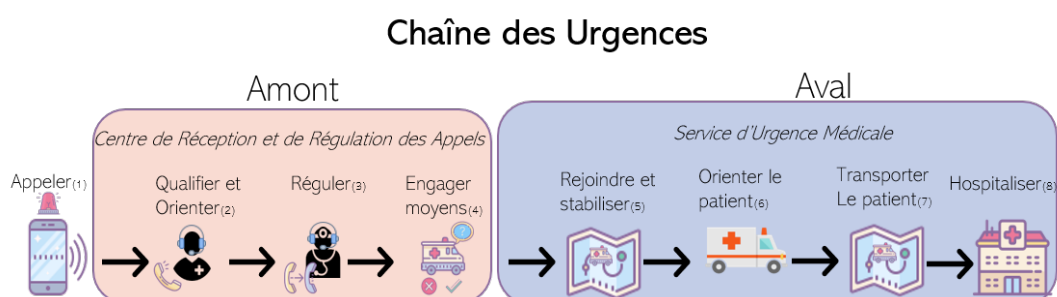


Figure 1.1 – La chaîne des urgences

l'amélioration de la gestion de l'urgence pour les citoyens. Une de ses activités consiste notamment à réaliser une veille technologique sur les différentes modalités d'appels dans plusieurs pays. Dans son dernier rapport ¹ de décembre 2019, elle mentionne différentes pratiques au sein de l'UE concernant l'accès aux soins d'urgence. En ce qui concerne l'accès aux centres d'appels d'urgences, trois émergent plus particulièrement :

1. **Un numéro unique** pour toutes les urgences, qu'elles soient médicales, civiles ou criminelles. 11 pays de l'UE adoptent ce mode de fonctionnement : L'Irlande, le Portugal, les Pays-Bas, le Danemark, la Suède, la Finlande, l'Estonie, la Roumanie, la Bulgarie, Chypre et Malte.
2. **Un numéro distinct** pour chaque type d'urgence. Ce mode de fonctionnement concerne 12 autres pays de l'UE : la France, l'Espagne, l'Italie, L'Autriche, la République tchèque, la Pologne, la Lituanie, la Lettonie, la Slovaquie, la Hongrie, la Croatie et la Grèce.
3. **Un numéro commun** simplifié pour les urgences médicales et les pompiers, la sécurité civile ayant gardé son propre numéro. C'est le cas pour 4 pays, la Belgique, le Luxembourg, l'Allemagne et la Slovénie.

Parmi ces trois moyens de contacter les secours, on constate que 41% des pays de l'UE disposent d'un numéro unique, 44% ont un numéro distinct dépendant du type d'urgence et 15% utilisent un modèle mixte avec un numéro commun pour les urgences médicales et les pompiers et un numéro particulier pour contacter la police. Il n'y a donc pas de système qui prédomine à l'échelle des pays européens, et encore moins d'harmonisation.

1.1.4 Quels sont les différents modèles de décrochés des appels

Une fois l'appel composé en utilisant le bon numéro, il est acheminé vers le centre d'appels adéquat pour être décroché. Plusieurs questions se posent : vers quel centre d'appels doit-il être envoyé s'il y en a plusieurs ? Quelles sont les ressources présentes sur la plateforme téléphonique ? Est-ce que ce sont des ressources généralistes, capables de décrocher tous les types d'urgence ? Ou est-ce que sont des ressources spécialistes d'un type d'urgence en particulier ? L'*European Emergency Number Association* (EENA) ² recense six modèles différents de prise en charge des appels. L'étude *EUROCALL* menée dans 21 centres d'appels de la sécurité civile de 18 pays différents montre l'importance de l'accès au centre d'appels (Nikolaou et al., 2017). En effet, le fait d'avoir une organisation mono ou bi-niveau aura un impact en terme d'accès.

1. **Le modèle décentralisé** présenté dans la Figure 1.2. Lorsqu'un appelant est dans le besoin, il choisit le numéro qu'il doit composer pour être mis en

1. Public Safety Answering Points – 2019 Global edition - <https://eena.org/knowledge-hub/documents/>

2. <https://eena.org/knowledge-hub/documents/>

relation avec le centre d'appels qui est le plus à même de répondre à son besoin : police, pompier ou service médicaux. C'est le cas de la France, de l'Autriche, de l'Allemagne, de l'Italie et de la Norvège.

2. **Le premier modèle centralisé (version V1)** est présenté dans la Figure 1.3. Lorsqu'un appel arrive, il est d'abord pris en charge par un opérateur de niveau 1 qui demande à l'appelant avec quel centre d'appels d'urgence (*gathering*) il souhaite être mis en relation, puis il effectue le transfert (*dispatching*). La phase de récupération des données et d'envoi de moyens est gérée par les spécialistes. C'est le cas notamment de l'Irlande.
3. **Le deuxième modèle centralisé** présenté dans la Figure 1.4. Lorsqu'un appel arrive, il est aussi décroché par un opérateur de niveau 1, mais celui-ci commence déjà à récupérer de l'information et c'est lui qui décide à quel(s) effecteur(s) (pompiers, policiers ou urgentistes) il doit transférer l'appel. C'est le cas de la Roumanie.
4. **Le modèle centralisé et unifié** présenté dans la Figure 1.5. Ce modèle est le même que le précédent sauf que tous les acteurs des centres d'appels sont dans un lieu commun. C'est le cas de certaines régions d'Espagne, de la Belgique.
5. **Le modèle généraliste** présenté dans la Figure 1.6. Dans ce modèle il n'y a qu'un centre d'appels géré par des civils sans spécialisations. Ce sont eux qui gèrent le décroché, la récupération des informations et l'envoi de moyens. La Finlande fonctionne avec ce modèle.
6. **Le modèle inter-connecté** présenté dans la Figure 1.7. Ce modèle est une variante collaborative où tous les centres d'appels sont connectés entre régions. S'il n'y a pas de ressources disponibles pour décrocher dans le centre d'appels le plus proche, l'appel est redirigé vers un autre centre d'appels. C'est le cas de la Bulgarie, de la République Tchèque et de la Suède.

1.1.5 Des stratégies pré-hospitalières différentes : *Stay and Play* ou *Scoop and Run* ?

Concernant la prise en charge préhospitalière du patient dans une situation d'urgence, on distingue deux catégories distinctes : Les soins vitaux de base (Basic Life Support,

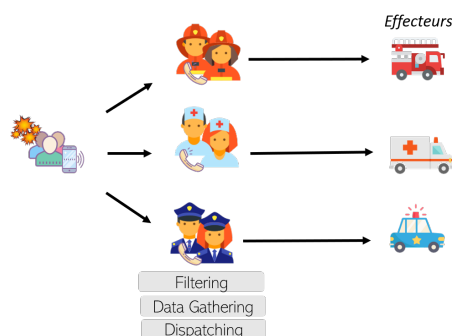


Figure 1.2 – Le modèle de décroché décentralisé

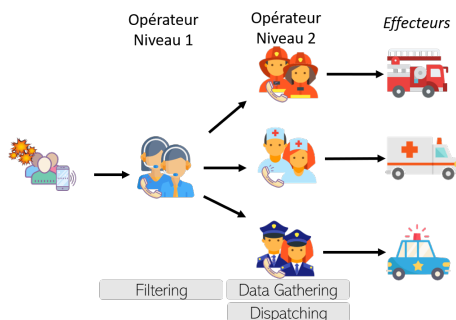


Figure 1.3 – Le modèle de décroché centralisé : première version

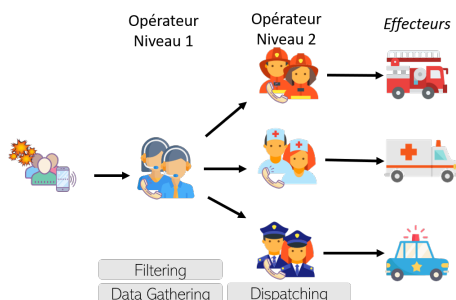


Figure 1.4 – Le modèle de décroché centralisé : deuxième version

BLS), et les soins vitaux avancés (Advanced Life Support, ALS). Le maintien des fonctions vitales de base (BLS) suppose la prise en charge des patients blessés par des méthodes non invasives. Cela suppose le dégagement des voies respiratoires, l'apport d'oxygène, la réanimation cardio-pulmonaire, l'arrêt de l'hémorragie externe, l'immobilisation des fractures et de la colonne vertébrale. L'objectif principal du BLS est de maintenir les fonctions cardiaques et respiratoires pendant le transport du patient vers le centre de traumatologie, sans causer de dommages supplémentaires (Beuran et al., 2012). Les soins vitaux avancés (ALS) impliquent l'utilisation de procédures invasives pour la prise en charge initiale des patients traumatisés : procédures plus sophistiquées pour la gestion des voies aériennes, la surveillance et la défibrillation cardiaques, le cathétérisme intraveineux et l'administration de médicaments (Beuran et al., 2012). Ce faisant, ces différences de pratique engendrent des philosophies de traitement du soin d'urgence qui divergent. On trouve deux stratégies aux extrêmes du spectre du soin préhospitalier qui sont :

- Le *Rejoindre et stabiliser* connue sous le nom "*Stay And Play*" (SAP) ou "*Stay And Stabilize*" qui implique le soin et la stabilisation de l'état du patient sur place avant de le transporter vers un établissement de santé. Ce mode de fonctionnement implique plus souvent l'utilisation d'ALS.
- Le *Charger et courir* connue sous la désignation "*Scoop And Run*" (SAR) ou "*Load And Go*" où l'objectif est de transporter le patient le plus rapidement possible dans un établissement de santé pour lui prodiguer les soins nécessaires. Cela nécessite de gérer les blessures qui mettent immédiatement la vie en danger dans l'ambulance pendant le transport.

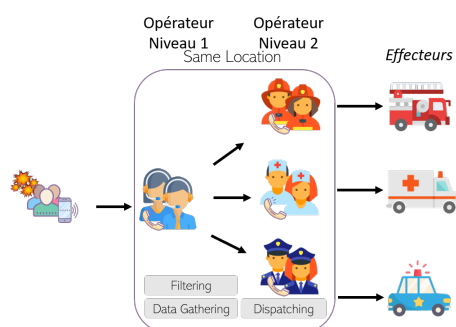


Figure 1.5 – Le modèle de décroché centralisé et unifié

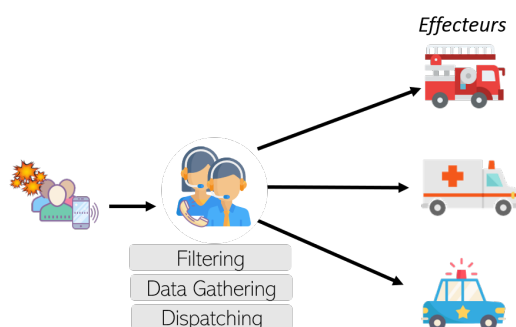


Figure 1.6 – Le modèle de décroché généralisé

Si on prend l'exemple du modèle français, on se trouve dans un cas de SAP. En effet, les véhicules du Service Mobile d'Urgence et de Réanimation (SMUR) sont gérés par une équipe médicale avec à son bord un médecin urgentiste. De plus, le modèle français va plus loin, car le patient est suivi par un médecin à chaque étape de sa prise en charge lorsque la médicalisation sur place est nécessaire. De la prise d'appels à son arrivée aux urgences en passant par la stabilisation du patient sur le lieu de l'urgence, il y a un médecin pour prendre en charge le patient.

Le modèle SAP a la particularité d'inclure un médecin à chaque étape de l'urgence préhospitalière. (Bieler et al., 2017) dans leur étude, comparent les modèles allemands et autrichiens au modèle anglo-saxon des Etats-Unis et du Royaume Uni. En Allemagne et Autriche, un médecin urgentiste est envoyé sur le terrain dans 95% des cas nécessitant une intervention. L'étude de Javaudin et al. (Javaudin et al., 2020) permet de documenter le système français et ses avantages. A contrario, de l'autre côté de la Manche et de l'Atlantique, ce sont des équipes para-médicales qui s'occupent de stabiliser le patient sur place et durant le transport vers un établissement de santé. L'objectif de cette étude est de mesurer les bénéfices potentiels de chacun des modèles. Les résultats laissent le débat entier sur l'intérêt de la présence d'un médecin, car il n'y a pas de différences significatives entre les deux modèles sur le temps de prise en charge ou sur le taux de mortalité dans les 24h ou durant hospitalisation. La conclusion est qu'il existe toujours un besoin d'analyse pour objectiver la procédure de prise en charge du patient la plus appropriée en termes de soins d'urgence. Malgré la grande quantité d'études existantes sur le sujet, ((Hannay et al., 2014) (Levy et al., 2007) (McCoy et al., 2013) (Smith et al., 2009)) nous ne

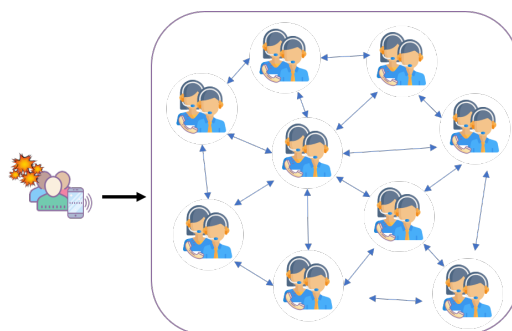


Figure 1.7 – Le modèle de décroché interconnecté

sommes toujours pas capables de statuer sur le modèle le plus bénéfique du point de vue de la survie des patients. Une étude sur plusieurs centres d'urgence européens montre néanmoins que les deux modèles sont utilisés en Europe et qu'il n'y a pas d'uniformité entre les pays. Le contenu de l'étude ne permet pas de conclure sur une meilleure performance d'un système par rapport à l'autre (Cnossen et al., 2019).

Dans cette partie, nous avons fait plusieurs constats. Les systèmes de santé pré-hospitaliers sont multiples et hétérogènes sur plusieurs niveaux :

- On trouve des modèles informels pour lesquels la gestion du soin d'urgence a un niveau de maturité assez faible. On trouve aussi des modèles formels qui sont plus structurés, organisés, et qui disposent de plus de moyens pour gérer le soin d'urgence.
- Au sein des modèles formels, on définit la chaîne des urgences qui se découpe en deux parties : amont et aval correspondant respectivement aux activités du centre d'appels et des services d'urgence médicale mobiles.
- Concernant l'accès à cette partie amont, les modalités varient selon les pays, certains adoptent un numéro unique, d'autre un numéro distinct, ou encore un modèle mixte avec un numéro commun pour les services médicaux et les pompiers, la police étant indépendant.
- On trouve aussi plusieurs modes de décrochés qui peuvent être plus ou moins centralisés avec des niveaux d'opérateurs téléphoniques différents.
- Enfin, les types de réponses pour la gestion du patient en aval avec le *Stay and Play* et le *Scoop and Run*.

Dans la suite de notre étude, nous nous focaliserons plus particulièrement sur le modèle français qui est un modèle formel. Le mode d'accès est avec un numéro distinct pour chaque type de besoin, et le mode de décroché peut varier au sein du territoire mais avec une prédominance pour le décroché décentralisé. Dans ce manuscrit, nous nous intéresserons donc aux modèles « *Stay and Play* » dont fait parti le modèle français.

1.2 Le modèle français du SAMU : Service d'Aide Médicale Urgente

Puisque cette étude a permis d'observer trois SAMU de la région Occitanie (Toulouse, Albi et Rodez), nous nous sommes intéressés plus spécifiquement au fonctionnement du modèle français que nous présentons dans cette section.

1.2.1 Le SAMU français : Naissance et Histoire

Le SAMU tel que nous le connaissons apparaît pour la première fois en 1955. Son succès se généralise et il se multiplie à travers la France dès 1964. La première forme du SAMU et de son Service Mobile d'Urgence et de Réanimation (SMUR) est un véhicule d'urgence dans lequel se trouvent un infirmier et un médecin. En 1965, les services d'urgence et de réanimation sont créés et dissociés des SMUR. C'est en 1968 que la première salle de régulation est créée à Toulouse par le Pr Louis Lareng : c'est la naissance du SAMU tel qu'on le connaît aujourd'hui. Il faudra attendre 1978 pour que le 15, numéro unique et national d'urgence, soit créé. La loi n°86-11 du 6 janvier 1986 officialise et rend obligatoire le SAMU dans chaque département. Elle définit aussi les différentes missions du SAMU comme le montre la Figure 1.8 :

- Réception et tri des appels via un numéro unique (15) auquel se rajoute le 112 dans certains départements.
- Assurer une écoute constante et apporter un conseil médical.
- Assurer l'envoi et la coordination des moyens de secours les plus adaptés dans les délais les plus rapides.
- Assurer la coordination des SMUR et de leurs Unités Mobiles Hospitalières (UMH).
- Vérifier la disponibilité des lits d'hospitalisation publics ou privés afin d'orienter le patient vers la structure la mieux adaptée à sa pathologie.
- Participer à l'élaboration et à l'organisation des différents plans de secours (plan blanc, plan rouge).
- Assurer la formation des personnels médicaux et paramédicaux aux soins d'urgence via les Centres d'Enseignement des Soins d'Urgences (CESU).

En 2009, la création de la Permanence Des Soins Ambulatoires (PDSA) par la loi HPST³ (Hôpital, Patient, Santé, Territoire) crée un nouveau flux d'appels dans les centres 15 : les appels de médecine ambulatoire. La PDSA est accessible durant les horaires de fermeture des médecins généralistes, soit le soir de 20 h à 8 h et le samedi de 12 h jusqu'à 8 h le lundi suivant.

Cet ajout ainsi que la démocratisation des SAMU ont engendré une augmentation structurelle du nombre d'appels dans les centres 15 pouvant aller de 5 à 15% par an (Penverne et al., 2009). Le livre blanc du SAMU-Urgences de France (SUdF) souligne⁴ que les Centres de Réception et de Régulation des Appels (CRRA) du

3. Loi n° 2009-879 du 21 juillet 2009 portant réforme de l'hôpital et relative aux patients, à la santé et aux territoires

4. (FRA 2015) SAMU Urgences de France Paris FRA. Livre blanc : organisation de la médecine d'urgence en France : un défi pour l'avenir. Français. Tech. rep. 2015, 48p

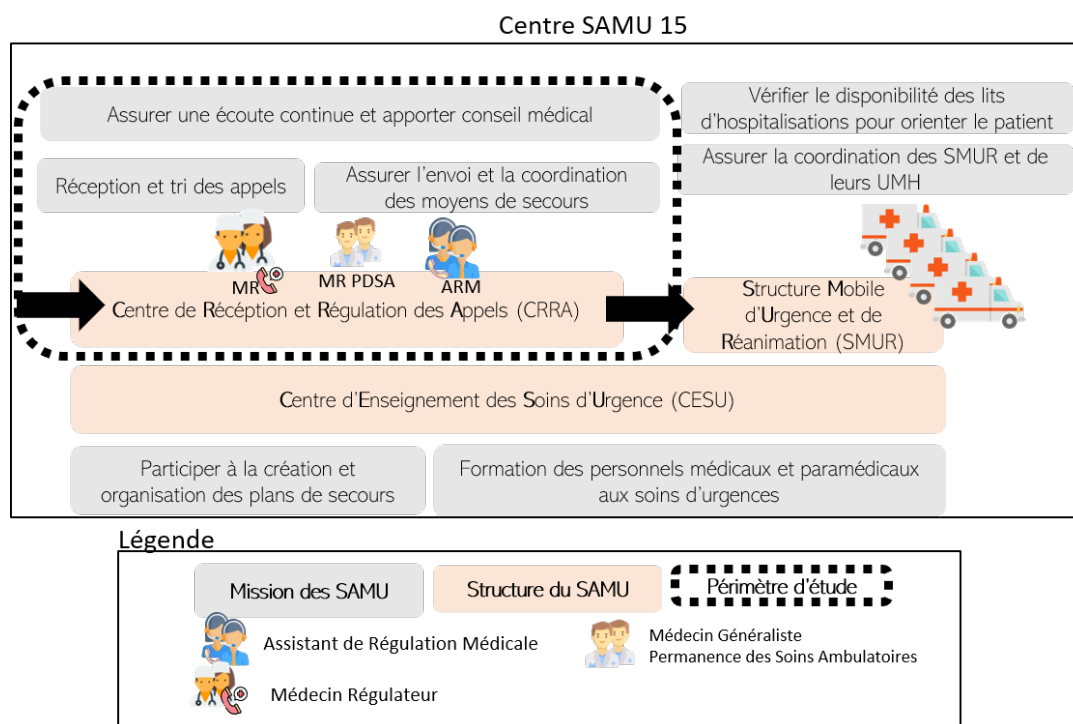


Figure 1.8 – Missions et structure du SAMU

SAMU font face à une augmentation importante du nombre d'appels entrants allant jusqu'à 10% par an. Il faut par ailleurs prévoir une nouvelle augmentation de la volumétrie des appels dans les prochaines années dues à l'intégration de nouveaux canaux de communication tels que les lunettes connectées, les appels en visio, les e-mails et texto, etc. Ces canaux de communication commencent à être testés dans certains centres d'appels et les discussions pour les intégrer aux SAMU sont déjà en cours depuis la sortie du pacte de refondation des urgences de 2019. Il faut aussi prendre en compte les appels E-call qui augmenteront en flèche d'ici 2025 (Oorni et al., 2017). En effet, depuis janvier 2020, toutes les voitures neuves doivent être équipées d'un système de communication directe avec les plateformes d'urgence. Ces changements vont obliger les SAMU à s'interroger sur la pertinence de leur organisation pour garantir un niveau de service à la hauteur de leurs missions et des enjeux associés.

Chaque SAMU se compose de trois parties :

- Une plateforme de gestion des appels, le CRRA (Centre de Réception et Régulation des Appels), qui est l'objet de ce travail de recherche.
- Un Centre d'Enseignement des Soins d'Urgence (CESU).
- Un ou plusieurs SMUR (Structure Mobile d'Urgence et de Réanimation) qui sont répartis dans le département pour assurer la meilleure couverture possible du territoire. Chaque SMUR dispose d'une ou de plusieurs équipes médicales d'intervention, pouvant être spécialisée (pédiatrie par exemple). Les équipes médicales sont généralement composées d'un médecin urgentiste, d'une infirmière (anesthésiste ou pas) et d'un ambulancier.

Pour décrire le modèle des SAMU français, nous nous appuyons sur trois sources principales :

- *Le référentiel et guide d'évaluation des SAMU Centres 15* publiés par *SAMU et Urgences de France* (SUdF) et la *Société française de Médecine d'Urgence* (SFMU) de mars 2015. (France et al., 2015)
- Le rapport MARCUS publié en décembre 2019 par Yann Penverne (médecin urgentiste) et le Lieutenant-colonel Romain Moutard (pompier sécurité civile) sur *la modernisation de l'accessibilité et de la réception des communications d'urgence pour la sécurité, la santé et les secours*. (Moutard et al., 2019)
- Les observations terrain effectuées dans les trois SAMU partenaires du projet.

Le rapport MARCUS est le résultat d'un programme interministeriel lancée par l'Inspection Générale de l'Administration et L'Inspection Générale des Affaires Sociales (IGAS) pour faire l'état des lieux du fonctionnement des services de secours, de soins et de sécurité. Ce programme commandé par le ministère de la santé et le ministère de l'intérieur a pour objectif d'émettre des préconisations concrètes d'améliorations des centres d'appels et de fournir une feuille de route afin d'améliorer la qualité du service et l'accessibilité à la population des services de secours. Les trois objectifs principaux recherchés par cette étude sont :

- Décrocher tous les appels adressés aux différents services d'urgence.
- Les décrocher plus rapidement.
- Garantir l'orientation de chaque appel vers le service pertinent.

De cette étude (rapport MARCUS), résulte un certain nombre de constats et de recommandations sur lesquels nous nous appuyerons afin de valider la pertinence de notre démarche d'un point de vue métier. Un premier constat effectué est que l'activité opérationnelle de réponse aux appels d'urgence n'est pas standardisée. Il apparaît donc nécessaire de pouvoir décrire le traitement des appels d'urgence en utilisant une sémantique et un vocabulaire commun et homogène.

Dans la suite de ce chapitre nous présentons les différents éléments qui constituent et caractérisent les SAMU. Ces éléments permettent de comprendre leur fonctionnement et leur complexité tout en introduisant le vocabulaire utilisé dans le domaine. De plus, nous proposons d'enrichir ces descriptions avec les observations terrain effectuées dans nos SAMU partenaires. Ces observations permettent de venir mettre en avant les différences qui existent entre les SAMU et corroborent les challenges relevés par le rapport MARCUS.

1.2.2 Les différents flux

Plusieurs types et flux d'appels transitent dans les centres d'appels d'urgences. Ils sont représentés dans la figure 1.9 Les centres d'appels du SAMU sont mixtes. Cela signifie qu'ils gèrent à la fois des appels entrants, et des appels sortants.

- Dans **les appels entrants** on considère, les appels de la population (aussi appelé primo appels), mais aussi les appels des professionnels de l'urgence (police, pompiers, ambulances) et les appels des centres d'appels d'autres régions.

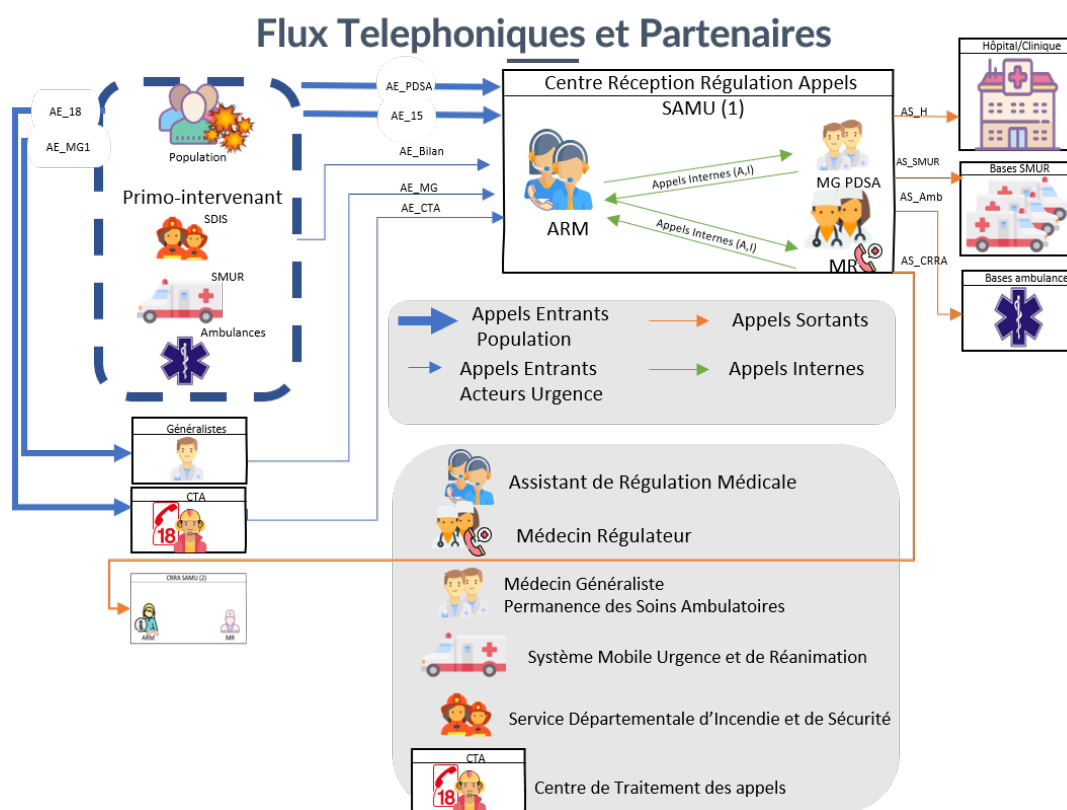


Figure 1.9 – Flux et partenaires du SAMU

Il y a aussi les appels bilans qui sont effectués par les primo-intervenants lorsqu'ils arrivent auprès du patient.

- Dans **les appels sortants**, on considère les communications vers les cliniques, hôpitaux et autres établissements de santé, les appels aux équipes de primo-intervenants pour les envoyer sur place, les appels vers les centres d'appels voisins.
- On considère aussi **des appels internes** entre les opérateurs du centre d'appels. Ceci peut arriver par exemple lorsqu'un Assistant de Régulation Médicale (ARM) appelle un Médecin Régulateur (MR) pour lui transmettre un appel particulier.

1.2.3 Le cycle de vie d'un appel entrant

Les SAMU ont défini plusieurs états pour un appel d'urgence afin de décrire au mieux son cycle de vie. La figure 1.10 résume tous les états de l'appel.

- **Un appel initié** est un appel qui est composé vers le SAMU.
- **Un appel non entrant** est un appel qui est rejeté par l'autocommutateur de manière automatisée.
- **Un appel entrant** est un appel initié qui arrive à joindre l'autocommutateur.
- **Un appel abandonné** est un appel entrant qui est interrompu par l'appelant (erreur, faux numéros, etc.).

- **Un appel traité automatiquement** est un appel entrant dirigé vers un serveur vocal interactif avec traitement automatique de l'appel.
- **Un appel perdu** est un appel présenté qui n'est ni décroché, ni traité automatiquement, ni rejeté et qui aurait dû faire l'objet d'un traitement.
- **Un appel présenté** est un appel entrant en mesure d'être décroché par un agent du CRRA.
- **Un appel décroché** est un appel présenté pris en charge par un agent du CRRA. une fois que l'appel est décroché, la communication est établie.

Le processus de gestion de l'appel nous permet de définir plusieurs durées caractéristiques du cycle de vie de l'appel. Ces différentes durées, une fois identifiées, serviront dans le calcul d'indicateurs de performance dont nous parlerons ultérieurement. Ces durées sont notamment représentées dans la figure 1.9.

- **Durée de l'appel** : Cette durée englobe toute la durée de vie de l'appel depuis son arrivée dans le centre d'appels jusqu'au moment où l'appel est raccroché à la fin de son traitement.
- **Temps d'attente** : c'est la durée d'attente de l'appel avant d'être décroché par un opérateur. On commence à mesurer ce temps à partir du moment où l'appel a passé l'autocommutateur et le disque d'accueil du SAMU. Ainsi, ce temps d'attente peut être décomposé en deux temps : **Le temps de réponse** et **le temps de sonnerie**. **Le temps de réponse** correspond à l'accueil vocal du SAMU. C'est ce que l'on appelle le « disque du SAMU » : un message automatique de bienvenue est passé à tous les appelants. **Le temps de sonnerie** correspond au moment où le téléphone de l'appelant sonne en attendant d'être décroché par un opérateur.
- **Durée de communication** : Cette durée est mesurée dès que l'appel est pris en charge par un opérateur. Elle s'étend jusqu'à ce que l'appel soit fini et raccroché. Il se peut que l'appel soit entrecoupé de temps d'attente notamment avant d'être transféré vers un médecin.
- **L'acheminement** correspond à la durée entre le moment où le patient compose l'appel et où celui-ci est visible dans le centre d'appel. Ce temps dépend majoritairement de l'opérateur téléphonique.

Ces termes présentés dans le référentiel métier de SUdF permettent d'uniformiser le vocabulaire utilisé pour décrire le traitement des appels du SAMU. C'est en effet un point très important permettant de faciliter l'échange. Nous avons pu le vérifier lors des observations menées sur le terrain. De plus, le rapport MARCUS souligne l'importance de mettre en place une sémantique homogène pour caractériser le traitement de l'appel d'urgence entre les métiers. Ceci constitue une des premières étapes pour se diriger vers une uniformisation des pratiques et permettre davantage de collaboration entre les SAMU.

1.2.4 Les ressources

Dans un CRRA, on trouve différents agents qui participent à la réception et au traitement de l'appel. Ce sont les acteurs internes du centre d'appels. On distingue :

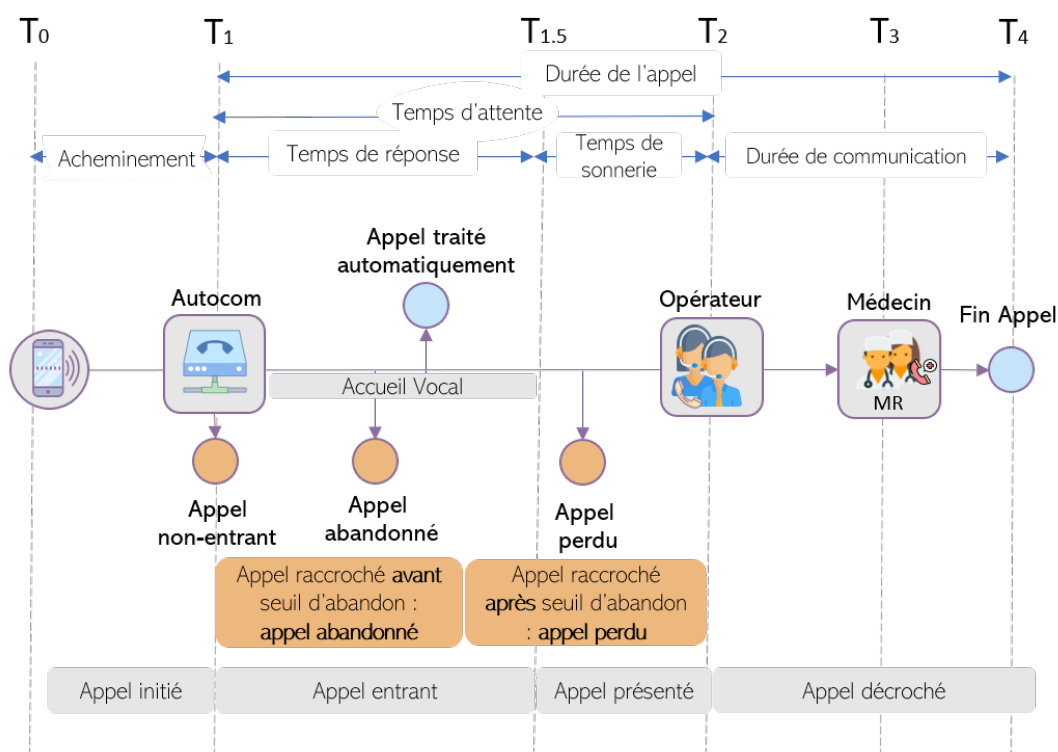


Figure 1.10 – Cycle de vie d'un appel : les étapes.

- Les **Assistants de Régulation Médicale (ARM)** qui décrochent et qualifient l'appel.
- Les **Médecins Régulateurs (MR)** qui traitent les appels nécessitant une régulation et décident de l'envoi de moyens adaptés.
- Les **Médecins Généralistes** de la PDSA qui traitent les appels de médecine générale.
- Les **Gestionnaires des Moyens (GM)** (facultatif) qui s'occupent de l'envoi effectif des moyens. À titre d'exemple, ils sont en charge d'appeler les compagnies d'ambulances, de contacter les pompiers pour les envoyer, de notifier les équipes SMUR de leur départ, etc.
- Les **Superviseurs** (facultatif) qui garantissent un management opérationnel conforme aux recommandations de l'Agence Nationale d'Appui à la Performance (ANAP). Leur rôle est d'assurer un usage des ressources cohérent vis-à-vis de la charge de travail.

Le management des ressources varie d'un département à un autre. Par exemple, parmi les SAMU observés, il existe des différences notables. Toulouse dispose d'un GM et d'un superviseur, ce qui n'est pas le cas pour Rodez et Albi. Rodez fonctionne avec un seul ARM et un seul MR pour gérer l'Aide Médicale Urgente (AMU). Durant les horaires de la PDSA, un ARM et un MR sont ajoutés pour gérer ces appels. Toulouse et Albi ont un effectif constant d'ARM sur 24 heures pour décrocher les appels, qu'ils soient AMU ou PDSA.

Puisque le CRRA fait partie du réseau des urgences, il interagit aussi avec des partenaires externes. Une partie des intervenants externes va générer des appels

entrants. C'est le cas pour les centres d'appels de la police ou de ceux de la gestion des incendies (pompiers). De même, les SAMU des départements voisins peuvent être amenés à appeler le centre d'appels pour solliciter l'aide de leur voisin. Les médecins généralistes, SOS Médecins, les infirmier(e)s libéraux et les maisons médicales de garde sont aussi des partenaires qui peuvent appeler le SAMU. Dans les appels entrants, on trouvera aussi tous les appels de suivi des cas en cours, les appels bilans passés par les primo-intervenants.

Ensuite, on retrouve les partenaires en aval de la chaîne des urgences. Tous les intervenants de terrain en font partie : les équipes du SMUR, les compagnies d'ambulances privées et les équipes d'interventions des pompiers. Les établissements de santé et les cliniques privées font aussi partie de ce réseau et peuvent être les destinataires d'appels sortants. Ils sont contactés afin d'obtenir des informations sur les places disponibles dans leurs services. Ces informations permettent au MR d'orienter efficacement le patient vers le meilleur établissement.

1.2.5 Processus de traitement d'un appel

Lorsqu'un appel arrive au CRRA, le processus de traitement d'un appel est lancé et suivra les différentes étapes correspondant à la chaîne des urgences décrite plus haut :

1. *Qualifier et Orienter* : Lorsqu'un ARM reçoit un appel, il effectue une première sélection. Dans le cas où l'appel n'est pas une demande médicale, l'ARM peut donner des renseignements administratifs. Sinon, ses tâches peuvent se décomposer en plusieurs sous-tâches comme précisées dans le référentiel métier de SuDF ((France et al., 2015)(chap. 3.6.1 p25) : «*Un ARM (1) reçoit l'appel initial (2) localise et identifie le patient (3) circonscrit la demande (4) en détermine le degré d'urgence et (5) engage si besoin des moyens avant (6) d'informer le MR et (7) d'éventuellement lui transférer l'appel.*». L'ARM engage des moyens lui-même seulement dans des cas bien précis qui sont identifiés dans des procédures métiers. C'est le cas notamment pour les Arrêts Cardio-Respiratoires (ACR).
2. *Réguler* : La régulation concerne le MR qui va pouvoir analyser la situation et diagnostiquer le patient par téléphone. Il effectue un interrogatoire médical pour ensuite décider des mesures à prendre : conseil médical ou envoi de moyens.
3. *Engager des moyens* : En fonction de la gravité de l'appel et de son diagnostic, le MR va décider des moyens à mettre en œuvre pour répondre à la situation du patient. Le type de moyen dépend de la priorité de l'appel, les échelles de priorités seront présentées un peu plus loin. Une fois que le MR a pris la décision d'envoyer une équipe SMUR, il identifie la base qui est géographiquement la plus proche et qui dispose d'une équipe disponible pour leur transmettre la notification d'intervention avec les informations nécessaires. Une fois cette étape finie, le processus bascule dans la partie aval (ou externe) de la gestion de l'appel.

4. *Rejoindre et stabiliser* : Une fois la notification reçue, l'équipe SMUR prépare l'intervention avec le véhicule adapté à la situation. Les Unités Mobiles Hospitalières (UMH) sont des ambulances disposant du matériel et des médicaments nécessaires pour transporter un patient alité. Les Véhicules Légers (VL) contiennent du matériel médical et des traitements pour la prise en charge et la stabilisation du patient, néanmoins ils ne permettent pas de transporter le patient. Enfin, l'hélicoptère sanitaire permet d'accéder très rapidement à des lieux reculés ou à des régions difficiles d'accès ou impraticables par voie routière. Une fois sur place, l'équipe de soignants stabilise le patient puis contacte le CRRA pour transmettre un bilan au MR afin qu'il puisse orienter le patient.
5. *Orienter* : Une fois le patient stabilisé et le bilan transmis, le MR dispose de plus d'information pour décider du devenir du patient. S'il décide d'envoyer le patient dans un établissement de santé, il doit s'assurer de la place disponible pour l'accueil du patient. Le choix de l'établissement de santé dépendra des places disponibles, de la spécialité nécessaire, mais aussi du choix du patient.
6. *Transporter* : Étape de transfert du patient du lieu de l'intervention vers l'établissement de santé.
7. *Hospitaliser* : L'équipe SMUR reste avec le patient jusqu'à son admission. Puis elle retourne vers sa base et son statut est de nouveau disponible, et peut au besoin être renvoyée en intervention.

Prioriser les appels

En fonction de leur priorité, les appels suivront tout ou partie des étapes présentées ci-dessus. Toutefois, les questions qui peuvent se poser maintenant sont notamment : dans quel ordre traiter les appels ? Quels sont les appels prioritaires par rapport aux autres ? Selon quels critères ? Est-ce que les appels entrants sont plus importants que les appels sortants ? Est-ce que tous les appels entrants sont plus importants que les appels sortants ? etc.

Le référentiel métier du SAMU (France et al., 2015) apporte des éléments de réponse à certaines de ces questions. Effectivement, il y a une hiérarchie dans les appels entrants. Des niveaux de priorité sont définis. La « priorisation » de l'appel se décline selon 3 niveaux de priorité, de P0 à P2.

- P0 : Pour les appels présentant d'emblée un caractère d'extrême urgence et autorisant un départ réflexe « flash » (SMUR, VSAB médicalisé) sans régulation médicale préalable (selon un protocole préétabli), suivi d'une régulation prioritaire ;
- P1 : Pour les appels nécessitant une régulation médicale prioritaire, immédiate, par un urgentiste ;
- P2 : Pour les appels relevant de la PDSA ou de transferts de patients vers une structure médicale, pouvant être temporisés.

La Haute Autorité de Santé (HAS) propose une classification différente (Santé, 2011). Bien qu'elle reconnaisse la pertinence de l'échelle de priorité proposée par SAMU

et Urgences de France, la HAS propose une décomposition en 4 niveaux. Les deux premiers concernant l'Aide Médicale Urgente (AMU) et les deux suivants concernant les médecins généralistes.

- Le niveau R1 correspond à une urgence absolue et vitale. On y envoie une équipe SMUR.
- Le niveau R2 correspond à une urgence sans détresse vitale. Il exige une intervention médicale prompte (temps d'intervention inférieur à 30 minutes).
- Le niveau R3 correspond à la Permanence des Soins Ambulatoires (PDSA). Ce niveau requiert un avis médical dans les 12 à 24 h.
- Le niveau R4 correspond à un conseil médical. Il peut donner lieu à la programmation d'un rendez-vous avec le médecin traitant.

Les ARM sont en mesure d'identifier les appels relevant de l'AMU (urgences de niveaux R1 et R2) de ceux relevant de la PDSA (degrés R3 et R4) et de les transférer vers le bon médecin qu'il soit urgentiste ou généraliste. Ils peuvent satisfaire de plus aux requêtes administratives distinctes du conseil médical.

Dans la suite, nous utiliserons le référentiel de SUdF auquel nous rajouterons un niveau P3 correspondant aux appels qui ne donnent pas lieu à un suivi par un médecin.

1.2.6 Les systèmes d'information des SAMU

La gestion informatique des SAMU se décompose en deux parties supportées par deux logiciels métiers différents. La partie de gestion des appels téléphoniques qui s'occupe de la gestion des files d'attente, du paramétrage du disque d'accueil, du routage des appels vers les ARM est supportée par le Logiciel de Téléphonie (LT). La gestion des dossiers de régulation est faite grâce au Logiciel de Régulation Médicale (LRM).

Les Logiciels de Téléphonie

Tous les SAMU de France ne sont pas équipés du même système de téléphonie. Il en existe de plus ou moins performants décomposés en quatre types d'après l'étude menée pour la mise en place du projet SI-SAMU⁵ :

- Des SAMU équipés de téléphonie très peu évoluée, avec des téléphones fixes et des serveurs non sauvegardés. Cela implique que la récupération des données de téléphonies est impossible, car il n'y a pas d'endroit où les données sont stockées.
- Des SAMU équipés de téléphonie standard, avec des téléphones fixes, mais une sauvegarde des serveurs. C'est notamment le cas pour le SAMU de Rodez où le décroché des appels se fait sur le téléphone fixe. Les ARM n'ont pas de visibilité sur la file d'attente des appels si ce n'est un voyant lumineux lorsqu'il y a un ou plusieurs appels en attente d'être décroché.

5. <https://esante.gouv.fr/projets-nationaux/si-samu>

- Des SAMU équipés de téléphonie sans fil avec sauvegarde et serveurs avec sauvegarde locale assurant la redondance des informations. C'est le cas par exemple pour le SAMU d'Albi.
- Des SAMU équipés de téléphonie avancée, interfacée avec un ordinateur pour permettre de décrocher et de gérer l'appel depuis le poste informatique. Le SAMU de Toulouse est dans ce cas, il dispose d'un logiciel de téléphonie interfacé avec le poste informatique. De ce fait, le décroché peut se faire directement depuis l'ordinateur. De plus, ils ont une visibilité sur les appels présents dans la file d'attente.

La gestion des dossiers patients

Lorsqu'un patient appelle le SAMU, il est affilié automatiquement à un dossier Patient. Il en existe plusieurs types (D, DR, DRM) en fonction de la situation du patient et de la réponse qui lui est apportée. Dès qu'un appel arrive au SAMU, le LRM crée un Dossier D en lien avec cet appel. Ce dossier sera qualifié de Dossier de Régulation (DR) lorsqu'il contient les informations renseignées par l'ARM lors de la qualification de l'appel. Le DR concerne les demandes d'information à caractère médical, médico-social ou sanitaire. Un dossier est qualifié de Dossier de Régulation Médicale (DRM) lorsqu'il a fait l'objet d'un acte de régulation médicale par un MR.

À la suite de nos observations sur le terrain, nous avons pu constater que la communication des données entre ces deux logiciels n'est pas évidente. Le lien entre un appel dans le LT et un dossier dans le LRM n'est pas automatique. Il n'existe pas forcément d'identifiants uniques permettant de les relier. Toutefois, nous pouvons espérer qu'avec le projet SI-SAMU d'uniformisation des systèmes d'information du SAMU, cela ne soit plus un problème. Intéressons nous un instant à ce fameux projet SI-SAMU.

Le projet SI-SAMU : Programme de modernisation des systèmes d'information et télécommunication des SAMU

Le projet SI-SAMU est un projet mené par l'agence du numérique en santé (appelée agence des systèmes d'information partagés de santé (ASIP) avant 2019) qui propose de concevoir un nouveau système d'information commun pour tous les SAMU. Ce projet ambitieux cherche à répondre à plusieurs objectifs :

- Contribuer à des soins et un service de qualité homogènes sur l'ensemble du territoire.
- Assurer la régulation médicale et le pilotage à tous les échelons, local, régional et national des appels d'urgences.
- Être en capacité de gérer une crise sanitaire d'envergure nationale.
- Améliorer l'interaction avec les systèmes d'information des partenaires comme avec les logiciels de santé.
- Garantir la confidentialité des données de santé.
- Offrir de nouvelles opportunités pour la recherche clinique.
- Concilier l'efficacité de la dépense publique et la performance de l'activité du SAMU.

SI-SAMU veut regrouper le LRM, le LT, les flux multicanaux (comme les ondes hertziennes), et le système de géolocalisation sous un seul logiciel unique et commun à tous les centres d'appels du territoire. Cette uniformisation a pour ambition une mise en réseau des SAMU, une interopérabilité accrue entre les acteurs du réseau des urgences, un meilleur pilotage de l'activité nominale, mais aussi de la gestion de crise. Cette solution métier est très attendue comme le montre ce témoignage du Dr Marc Giroud, cofondateur de SAMU France en 1975 :

« Si la pandémie de grippe A H1N1 a souligné l'immense intérêt des centres de régulation médicale en 2009, elle a aussi mis en évidence l'insuffisance et le cloisonnement de leurs solutions de télécommunication et de leurs systèmes d'information, dont le rôle est pourtant critique pour prendre en charge des patients en situation de détresse vitale. Le travail conduit sur le SI-SAMU par l'ASIP Santé depuis 2014 est sans précédent. Non seulement il s'est fait à un niveau d'expertise et de concertation inédit, mais il a suscité l'adhésion de la communauté métier élargie. Une adhésion croissante des urgentistes hospitaliers, dont certains avaient initialement manifesté des réserves et qui sont aujourd'hui impatients d'accéder aux progrès attendus, en termes de qualité et de sécurité ».

Ce projet, commencé en 2014, est aujourd'hui en 2020 dans la phase des tests pilotes sur certains sites. D'après le calendrier initial, le déploiement du SI-SAMU devrait se généraliser sur tout le territoire entre 2020 et 2024.

1.2.7 Évaluation des SAMU

Nous avons auparavant souligné les difficultés du SAMU à décrocher tous les appels. Un nombre non négligeable d'appels abandonnés ou perdus a été relevé durant l'année 2018. À noter qu'un seuil d'abandon en deçà duquel l'appel est considéré comme une erreur de l'appelant est fixé à 15 secondes pour le SAMU. Cela signifie que les *appels abandonnés* avant 15 secondes ne sont pas considérés comme des appels qui auraient nécessité une réponse, là où les appels décrochés après 15 secondes sont des *appels perdus* qui auraient dû recevoir une réponse (France et al., 2015).

Plusieurs indicateurs de performances (KPI) sont utilisés pour évaluer les centres d'appels. On peut distinguer plusieurs catégories d'indicateurs suivant le triptyque classique coût, qualité et délais (Olsen, 1971). Parmi eux on retrouve les suivants :

Les indicateurs de délais

1. **La Durée Moyenne de Communication** (DMC) est le rapport entre le temps de communication et le nombre d'appels, elle correspond au temps moyen de communication passé sur chaque appel par agent. Cela permet d'évaluer leur productivité. (cf. équation 1.1)

$$DMC = \frac{\text{temps de communication total}}{\text{nb d'appels traités}} \quad (1.1)$$

2. **Le temps de réponse** est défini comme la durée entre l'arrivée de l'appel sur la plateforme téléphonique et la prise en charge du patient par des effecteurs sur place (Aboueljinane, 2014). Il peut être calculé de plusieurs manières. On peut considérer le temps de réponse moyen, mais aussi le **Taux de couverture dans un délai T** (Ingolfsson et al., 2003) où T est une valeur prédéfinie qui est soit fixée par les auteurs (cf. équation 1.3), ou fixée par la loi en vigueur dans le pays. On retrouve aussi des **Taux de non-couverture au-delà d'un délai T** qui se définissent comme le pourcentage d'appels pris en charge au-delà du seuil T (cf. équation 1.4). La valeur de T varie d'un pays à un autre suivant les recommandations gouvernementales ou régionales. En Italie, 100% des appels doivent être gérés en moins de 8 min dans les villes et leurs périphéries. En Amérique du Nord, 90% des appels doivent être couverts en moins de 9 minutes (Ingolfsson et al., 2003).

$$\text{Temps de réponse} = \text{Instant de prise en charge du patient} - \text{Arrivée de l'appel} \quad (1.2)$$

$$\text{Taux de couverture dans un délai } T = \frac{\text{Nbre d'appels traités en moins de } T \text{ sec}}{\text{Nbre appel total}} \quad (1.3)$$

$$\text{Taux de non couverture dans un délai } T = \frac{\text{Nbre d'appels traités en plus de } T \text{ sec}}{\text{Nbre appel total}} \quad (1.4)$$

3. **Le temps d'accès au traitement** est défini comme le délai entre l'arrivée de l'appel dans le centre et l'arrivée du patient dans un établissement de santé. Il peut être mesuré de différentes manières : Su et al. utilisent la durée entre la demande d'envoi d'une équipe d'effecteur et l'admission du patient à l'hôpital (Su et al., 2003). Lee et al. le définissent d'une manière similaire au temps de couverture avec un seuil T. Ils mesurent alors la part des appels dont le temps d'accès au traitement est inférieur à ce temps T (Lee et al., 2012).
4. **Le temps de service** correspond à la durée ininterrompue pendant laquelle une ressource est engagée à réaliser un service. Pour la partie aval, cela correspond au temps passé en service par l'équipe d'effecteurs entre le moment où elle est notifiée de l'intervention et le moment où elle est de nouveau disponible. Pour la partie amont dans le centre d'appels, le temps de service correspond au temps de communication, c'est le temps où les ressources sont en ligne ou en train d'effectuer une activité en lien avec l'appel. Cet indicateur est important car il permet d'assurer que la charge de travail est bien répartie. Goldberg, montre l'importance sociale de cet indicateur : une équité dans la répartition de la charge de travail permet de diminuer le sentiment de pénibilité et d'injustice ressenti par le personnel (Goldberg, 2004). Ce qui amène directement à définir un taux d'utilisation des ressources.

5. **Le temps de déclenchement** correspond au délai entre l'arrivée de l'appel au CRRA et la notification de la mission aux effecteurs. Cet indicateur permet de mesurer l'efficacité du maillon amont de la chaîne des urgences, la capacité du centre d'appels à traiter et à décider d'envoyer des moyens (Aboueljinane, 2014). Pour le maillon aval, c'est l'indicateur suivant qui sera utilisé.
6. **Le temps de transport** au lieu d'intervention correspond au délai entre la notification de la mission aux effecteurs et l'arrivée de l'équipe sur place. Ce délai est sujet à des perturbations extérieures comme les conditions de circulation, ou la difficulté d'accès et de localisation de la zone d'intervention (Su et al., 2003).
7. **Le temps d'attente** se définit à la fois pour la partie amont et la partie aval. Pour la partie amont, c'est le temps avant qu'un appel ne soit décroché. Pour la partie aval, il correspond au temps passé à attendre avant qu'une équipe d'effecteurs soit affectée à un appel. De même, **la taille de file d'attente** correspond au nombre d'appels en attente de l'affectation d'une équipe d'effecteurs. Silva et al. utilisent cet indicateur pour évaluer la coordination entre les activités de l'amont et l'aval. De plus, ils utilisent la valeur maximale de la file d'attente pour évaluer la résilience du système lors d'une augmentation des appels (Silva et al., 2010). La résilience peut être définie comme la capacité d'un système à revenir à son état initial après avoir subi une perturbation (Leflar et al., 2013).

Les indicateurs de qualité

1. **Le taux d'accueil** est défini comme le nombre d'appels répondus sur le nombre d'appels entrants. Il exprime la capacité du centre d'appels à être capable de décrocher tous les appels. D'après SUdF en 2015, il n'y avait pas de recommandation professionnelle concernant l'objectif du taux d'accueil. Toutefois le rapport MARCUS de 2019 précise que suite au comité stratégique du 7 novembre 2019, la cible de 99,95% de taux d'accueil est fixée. Cet indicateur de taux d'accueil permet de mesurer l'accès de la population aux centres d'appels d'urgences. Il est généralement calculé selon la formule 1.5.

$$Taux\ Accueil = \frac{Nb\ d'appels\ décrochés}{nb\ d'appels\ entrants - nb\ d'appels\ abandonnés} \quad (1.5)$$

2. **La qualité de service** aussi appelé temps de décroché est le pourcentage d'appels décrochés en moins de x secondes. x peut prendre plusieurs valeurs selon le pays. L'état français fixe que 99% des appels doivent être décrochés en moins de 60 secondes (Penverne et al., 2017). La NENA⁶ pour le 911 aux USA définis que 95% des appels doivent être décroché en moins de 20 secondes. (cf. équation 1.6)

$$QS_x = \frac{Nb\ d'appels\ décrochés\ avant\ x\ sec}{nb\ d'appels\ entrants - nb\ d'appels\ abandonnés} \quad (1.6)$$

6. National Emergency Number Agencies

3. **Le nombre d'appels servis par base/équipe** permet de compléter la mesure du temps de service ainsi que le taux d'utilisation pour assurer une certaine équité entre les équipes. Cela permet que ce ne soit pas toujours la même équipe qui parte en intervention, cela permet aussi de lisser le taux d'utilisation entre les équipes. Cet indicateur est transposable aux personnels du CRRA avec le nombre d'appels décrochés/traités par ressource (Penverne et al., 2009). Les recommandations de SUDF sont qu'un ARM devrait traiter une moyenne de 7,2 appels par heure. De la même manière, la charge cible pour un MR est de 6 dossiers de régulations médicales par heure⁷.
4. **Le pourcentage de non-réponse** correspond à l'échec de prise en charge d'un appel dans un temps limite fixé. Il est complété par le taux d'abandon, qui correspond aux appels qui ne sont pas décrochés. Il faudrait aussi rajouter les appels qui n'arrivent pas dans le système, car les files d'attente sont saturées avant le décroché. (Inakawa et al., 2010)
5. **Le taux de survie** est un indicateur à part entière qui est plus difficile à mesurer que les différents délais. Il correspond au pourcentage de patients qui survivent après avoir fait appel au service de secours. C'est un indicateur qui reste très difficile à quantifier dans la mesure où beaucoup de facteurs extérieurs peuvent intervenir dans le décès potentiel d'un patient. Évaluer les impacts des modifications organisationnelles sur le taux de survie est donc assez complexe, ce qui explique le manque de représentation de cet indicateur dans la littérature. (Sacco et al., 2005)
6. **La collaboration** est une famille d'indicateurs définis par (Belaidi et al., 2007) qui évalue un SAMU au sein du réseau de soin des urgences vis-à-vis des autres acteurs du réseau. Les concepts d'évaluation de la collaboration sont définis, mais ne sont pas forcément mis en application.

Les indicateurs de coût

1. **Le taux d'utilisation** ou encore **taux de charge** concerne tant les ressources du centre d'appel que les effecteurs. Il correspond à la fraction entre le temps de service et le temps total de travail (cf. équation 1.7). Les taux d'utilisation des équipes comme des ressources du CRRA sont généralement bas pour maximiser les chances d'être disponibles lors d'une urgence. Dans le système français, le taux de charge cible pour les médecins devrait être fixé autour de 25% d'après les recommandations de la MeaH (Hospitalier (MeaH), 2008). Concernant les ARM, le taux de charge cible est se situe entre 20 et 40 %⁸.

$$\text{Taux d'utilisation} = \frac{\text{temps de service}}{\text{temps total de travail}} \quad (1.7)$$

7. Samu-Urgences de France. Définitions & standardisation des données : Activité des SAMU Centre 15. 2009

8. Samu-Urgences de France. Définitions & standardisation des données : Activité des SAMU Centre 15. 2009 : http://www.samu-de-france.fr/documents/actus/155/559/definitions_apl_drm_2009.pdf; http://www.samu-de-france.fr/documents/actus/2155/2559/thesaurus_apl_drm_2009.xls

2. **La distance parcourue** permet d'évaluer le coût kilométrique des interventions, cela permet d'évaluer l'efficacité du déploiement des équipes. Elle correspond à la distance totale parcourue par une équipe SMUR depuis son départ de la base jusqu'à son retour après intervention. (Koch et al., 2003)
3. **La rentabilité** correspond à l'évaluation du coût de mise en place des modifications organisationnelles proposées. C'est un indicateur plutôt complexe dans le contexte d'un service d'urgence. Le but de cet indicateur n'est pas tant de mesurer la rentabilité économique des services d'urgences que d'être en mesure d'évaluer le coût et les bénéfices des solutions proposées pour l'améliorer. Bien souvent les améliorations proposées ne sont pas ou peu évaluées du point de vue économique. Aboueljinane et al. l'ont déjà relevé à travers leur revue de littérature sur le sujet (Aboueljinane et al., 2013). La littérature est assez succincte sur cette question sûrement à cause de la difficulté de chiffrer les améliorations proposées, bien que des tentatives de classifications de coûts standards aient été proposées par (Savas, 1969) ou encore par (Lerner et al., 2007).

La Figure 1.11 présente une synthèse des indicateurs sur la chaîne des urgences.

1.3 Revue de littérature sur la gestion des SAMU

Maintenant que nous avons fixé les éléments de caractérisation du système d'étude, intéressons-nous aux principaux travaux de recherche qui ont été conduits sur le sujet. Les SAMU ont suscité l'intérêt de la sphère académique de par leurs problématiques qui ouvrent un champ d'investigation à la fois scientifiquement riche, mais aussi très humain (Aringhieri et al., 2017). La complexité des interactions humaines lors d'une situation d'urgence, croisée avec la nécessité d'agir rapidement et le tout au sein d'un réseau multi-acteurs (Belaidi, 2009) font qu'il y a de nombreux défis à relever. De nombreux travaux scientifiques ont étudié les centres d'appels d'urgences en prenant en compte leurs différences et leurs particularités. Parmi les problématiques adressées, on trouve notamment la définition du nombre de ressources (Addis et al., 2012) pour décrocher les appels, mais aussi la définition d'indicateurs de performance adéquats pour évaluer l'activité (Passmore et al., 2013; Penverne et al., 2017), le dimensionnement des flottes de véhicules pour répondre à la demande de soin (Aboueljinane et al., 2014). Il faut ensuite traiter la localisation et la répartition de ces véhicules pour avoir un maillage optimal du territoire (Bélanger et al., 2016). Ce sont des exemples de challenges auxquels font face les SAMU et que nous allons étudier plus en profondeur dans la suite.

Toutefois, tous les problèmes ne concernent pas l'intégralité de la chaîne des urgences. En effet, on observe des problèmes bien distincts entre la partie amont liée à la gestion des appels et la partie aval liée à la gestion des effecteurs et de la prise en charge du patient. La figure 1.12 cherche à représenter les types de problèmes traités en fonction de l'étape de la chaîne des urgences. Une manière de les classer et de découper en

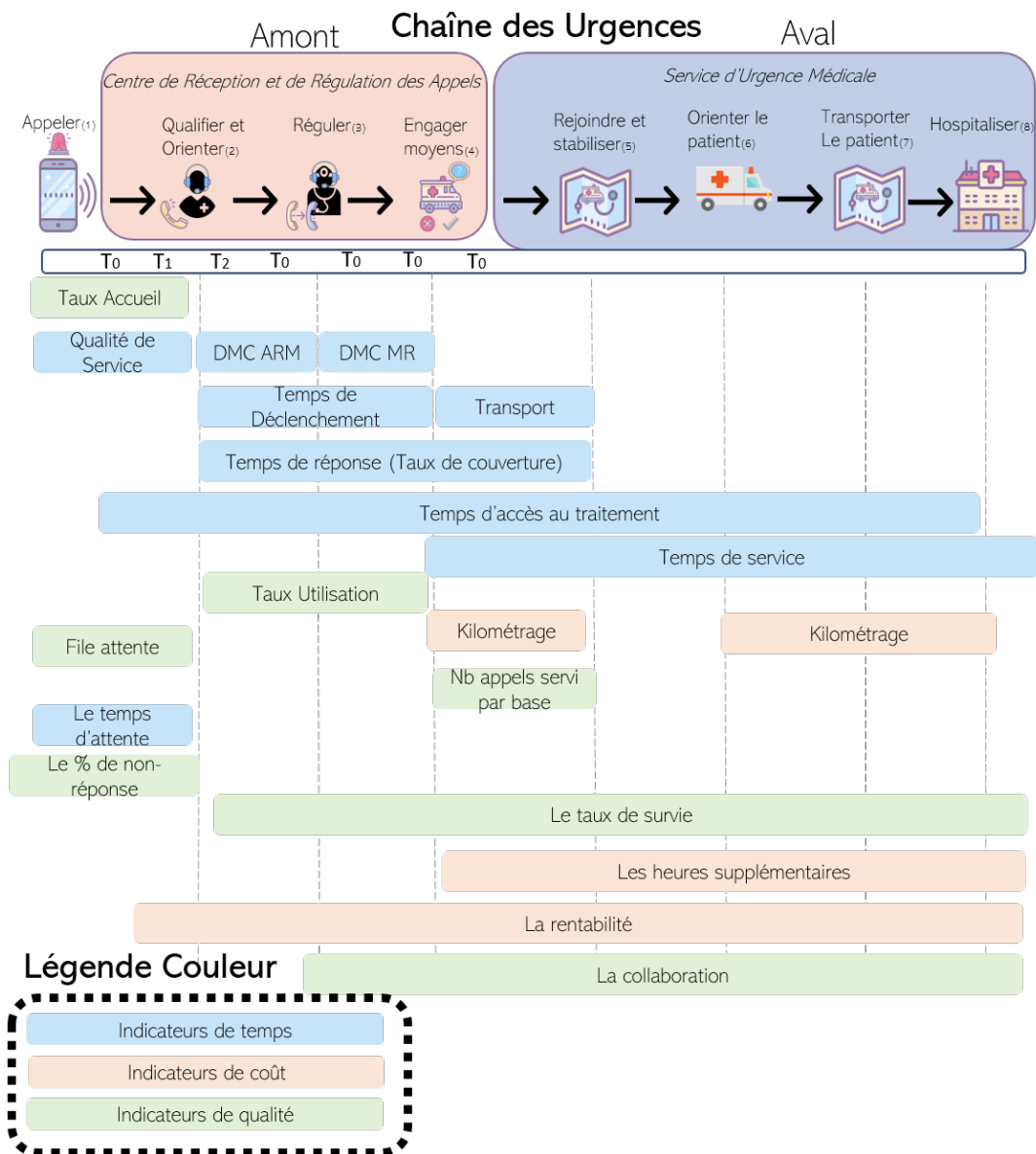


Figure 1.11 – Indicateurs de performance

trois niveaux de décisions différents : stratégique, tactique et opérationnel (Gendreau et al., 1997). Le niveau stratégique correspond à un horizon de temps assez long avec des décisions structurelles qui nécessitent un investissement conséquent. On retrouve les problèmes de localisation des bases SMUR, de dimensionnement des flottes de véhicules et de leurs équipes. On retrouve aussi les problèmes prospectifs comme la gestion des nouveaux supports de communication. Le niveau tactique s'intéresse notamment à la prévision des appels pour pouvoir définir les plannings des ressources, les affectations des équipes aux différentes bases. Le niveau opérationnel concerne le pilotage de l'activité quotidienne. Dans cette catégorie on va trouver les problèmes de répartition (*dispatching*), de routage (*routing*) et de relocalisation des équipes d'intervention pour la partie aval. Pour la partie amont on peut y ranger les règles de décroché et de priorisation des appels.

1.3.1 Les problèmes stratégiques

Les problèmes de dimensionnement

Les services médicaux d'urgences ont besoin de ressources hautement qualifiées que ce soit au niveau amont du centre d'appels ou au niveau aval des intervenants terrains afin d'apporter le meilleur soin possible au patient. C'est pourquoi les problèmes de dimensionnement des équipes de soin (Passmore et al., 2013) sont primordiaux dans la conception et la gestion des systèmes de soins urgents. En effet, il faut trouver un juste équilibre entre un nombre suffisant de ressources de manière à garder un temps d'attente bas et une éventuelle sous-utilisation de ces ressources hautement qualifiées avec un coût salarial important. (Addis et al., 2012) Dans les problèmes de dimensionnement, on trouve aussi ceux qui concernent les flottes de véhicules (Savas, 1969 ; Trudeau et al., 1989). Ces problèmes cherchent à répondre à la question « Combien d'équipes et de véhicules faut-il pour être capable de répondre à la demande d'intervention ? ». Ces problèmes sont complexes, car ils sont soumis à une grande incertitude. En effet, le dimensionnement des équipes dépend directement de la prévision du nombre d'interventions potentielles (Setzler et al., 2009). Plusieurs chercheurs se sont penchés sur ce problème pour identifier le nombre de bases nécessaires pour couvrir une zone géographique le plus efficacement possible. Nous pouvons citer notamment : (Aringhieri et al., 2017) (Iskander, 1989), (Su et al., 2003), (Swoveland et al., 1973), (Trudeau et al., 1989). Savas, 1969 a utilisé la simulation afin de tester l'effet de l'ajout ou de la suppression d'équipes et de bases potentielles sur la qualité et le coût du service d'urgence. L'objectif était de réduire le temps de réponse moyen des équipes de l'hôpital Kings County à New York. Il montre que l'ajout de nouvelles équipes dans la base initiale (soit l'hôpital) permet de réduire le temps de réponse moyen de 18 secondes. Au-delà de dix équipes, il n'y a plus de différence sur le temps moyen. De plus, il montre qu'en jouant sur les emplacements des équipes il est possible d'améliorer le temps de réponse jusqu'à 19% du temps initial. Cette étude est une des premières à utiliser la simulation pour un service d'urgence. Une autre étude (Gunes et al., 2005) a évalué l'impact qu'aurait l'extension de planning d'un hélicoptère de 5 à 7 jours pour un état du nord des États-Unis. Leurs résultats montrent que cela induirait

une diminution du taux d'utilisation de l'hélicoptère de 43% avec un impact très faible et non significatif sur le temps de réponse. (Aboueljinane, 2014), (Inakawa et al., 2010) et (Ingolfsson et al., 2003) ont développé des modèles de simulation pour mesurer l'effet de la modification du nombre d'équipes et de bases pour les services d'urgence de la région du Val-de-Marne en France, de Seto au Japon et d'Edmonton au Canada respectivement. Les résultats ne sont pas les mêmes d'une étude à l'autre. À Edmonton, il serait plus intéressant de rajouter deux nouvelles équipes plutôt qu'une nouvelle base, cela permettrait d'améliorer le pourcentage des appels répondus en moins de 9 minutes de 87% à 90%. Au Japon et dans le Val-de-Marne, les études montrent que l'ajout de nouvelles bases et la répartition des équipes dans celles-ci permettent d'obtenir de meilleurs résultats que l'ajout de nouvelles équipes.

Les problèmes de localisation

Les problèmes de localisation ont pour but de décider de l'emplacement des bases SMUR d'où partent les véhicules d'intervention (Savas, 1969) (Aboueljinane et al., 2013). Comme on a pu le constater dans le paragraphe précédent, ils vont bien souvent de pair avec les problèmes de dimensionnement. L'objectif est d'avoir un maillage optimal pour maximiser la couverture du territoire. Cela signifie que l'on doit être capable de joindre chaque point du territoire dans un délai défini au préalable. La localisation va être liée aux emplacements des bases. Les problèmes de dimensionnement répondent à la question « Combien d'infrastructures dois-je déployer ? ». La localisation, quant à elle, répond à la question « Où dois-je les déployer ? ». Il est nécessaire de trouver un compromis pour maximiser la couverture tout en minimisant le nombre de bases. Ces problèmes de localisation et de dimensionnement vont donc avoir un impact sur plusieurs types d'indicateurs :

- Des indicateurs de temps : Si la localisation est bien faite et que la couverture du territoire est bonne, le temps de prise en charge des interventions sera réduit, car il sera plus rapide de se rendre sur place.
- Des indicateurs de coût : Si la localisation est optimale, elle permet d'avoir une bonne couverture du territoire en minimisant le nombre d'infrastructures. Cela permet de réduire les coûts associés au nombre de kilomètres parcourus.

(Başar et al., 2012) proposent de classer les problèmes de localisation suivant trois critères principaux : le type de problème, la modélisation, et les méthodologies de résolution utilisées. Pour le type de problème, ils les ont répartis en quatre sous-groupes comprenant le type d'urgence, la structure du modèle (s'il est déterministe, stochastique ou robuste), la variation du modèle dans le temps (s'il est statique ou dynamique) et enfin le nombre d'objectifs à atteindre. Concernant la modélisation, elle est caractérisée par la définition de la fonction objectif, les paramètres utilisés ainsi que le type de programmation mathématique utilisée (entière, dynamique, non linéaire, etc.) Quant aux méthodologies elles se répartissent dans plusieurs catégories : les méthodes exactes, les heuristiques, les méta heuristiques et la simulation.

La gestion des nouveaux supports

Dans les années à venir de nouveaux supports de communication vont être utilisés par les centres d'appels d'urgence que ce soit des supports vidéo, SMS, e-mail, ou encore les appels E-call issus des voitures. L'avancée des technologies de communication pourrait avoir un impact sur le fonctionnement des SAMU. Gupta et al. listent les différents challenges et problèmes en lien avec la diversification des canaux de communication notamment dans le cadre de la gestion de crise : « Comment faire évoluer l'architecture du 911 pour intégrer les technologies sans fil ? La géolocalisation ? », « Comment gérer la continuité de la prise en charge d'une série de SMS ? », « Comment éviter la transmission de virus par SMS et MMS dans de telles situations ». « Comment intégrer la gestion de support image et vidéo ? », « Est-ce que l'utilisation de capteurs dans les maisons notamment de personnes âgées est envisageable ? ». Toutes ces questions se posent à l'heure où la technologie de l'information et de la récupération des données se développe extrêmement rapidement. La grande quantité de données que va générer la multiplication des canaux de communication induit de nouveaux besoins pour la visualisation, l'analyse et l'utilisation des données. Il y a aussi un besoin d'interopérabilité entre les différents acteurs pour la transmission des données pertinentes. Un autre challenge concerne la localisation de l'appelant pour localiser le lieu de l'accident, mais aussi localiser le centre qui doit gérer l'appel. Notamment pour le cas des téléphones portables qui sont plus difficiles à localiser lorsqu'ils sont proches de frontières ou de grands axes. Un autre problème est la gestion des appels redondants dans le cas de crises majeures. Comment les traiter, les rediriger vers d'autres acteurs de la sécurité civile comme la croix rouge par exemple ? Une preuve de concept a été faite sur le nouveau modèle envisagé du NG-9-1-1 où le transfert d'appels d'un centre d'appels vers un autre a été testé avec succès (Gupta et al., 2010).

Markakis et al. posent différentes questions comme « Comment intégrer les médias sociaux à la gestion des appels d'urgence ? » ou encore « Comment préparer les services d'urgences la prochaine génération de réseau (Next Generation Networks, NGN) ». Cet article propose une plateforme pour intégrer les nouveaux réseaux de communication au service des urgences. Cela permet alors de rassembler les communications téléphoniques, radiophoniques, smartphones, tablettes, E-Call, mais aussi les différents médias sociaux existants. (Markakis et al., 2017). (Weinlich et al., 2018) développent et testent un système de géolocalisation par GPS et Wi-Fi afin de contacter rapidement les urgences. À des fins de test, il a été déployé dans 11 pays différents. Les résultats de leur travail montrent qu'un tel système améliore significativement l'arrivée des urgences (jusqu'à 2 heures peuvent être gagnés) surtout pour des patients désorientés ou avec une difficulté à s'exprimer.

Oorni et Goulart rappellent qu'à partir de 2020, tous les véhicules commercialisés devront être équipés du module E-call permettant de joindre directement les centres d'appels d'urgence par pression du bouton ou automatiquement lors du déclenchement des airbags. Avec cette évolution, les auteurs se demandent si les plateformes de gestion des appels d'urgences sont prêtes et équipées pour recevoir les E-Call. À travers cet article les auteurs présentent le concept des appels E-call et la manière

dont il est implémenté en comparant deux systèmes. Cet article étudie plus le côté architecture système des centres d'appels d'urgence (Oorni et al., 2017).

Les problèmes de dimensionnement et de localisation sont très liés et bien souvent traités ensemble dans la littérature scientifique. En effet, il est évident qu'il faut connaître le nombre de ressources disponibles avant de pouvoir les répartir. La difficulté de ces problèmes réside dans le fait de trouver une réponse optimale pour satisfaire des critères de temps tout en ayant des contraintes de moyens.

L'apparition de nouveaux supports de communication soulève des challenges au niveau de l'architecture du réseau, mais aussi au niveau de la gestion et du traitement de ces sources d'information. Qui devra s'en occuper ? Et comment les intégrer aux outils déjà existants sont des questions auxquelles des réponses restent à apporter.

1.3.2 Les problèmes tactiques

Les problèmes de prévisions des appels

Pour résoudre les problèmes mentionnés précédemment il est nécessaire d'avoir une vision assez précise du nombre d'appels, du nombre d'interventions qu'il peut y avoir par heure, jour, semaine, etc. Ainsi, un autre pan de la recherche s'intéresse à la prédiction du volume d'appels et aux lois probabilistes suivies par l'arrivée des appels. Pour déterminer au mieux le nombre d'appels dans le futur, il est aussi intéressant de connaître les paramètres extérieurs et environnementaux qui peuvent influencer l'arrivée des appels. Les vacances scolaires peuvent générer une augmentation des appels d'urgence dans certaines régions. Ceci est dû au tourisme et à l'augmentation de la population dans une zone donnée sur une courte période. Les épidémies saisonnières comme la grippe auront aussi un impact sur l'arrivée des appels. La démographie d'une région aura aussi une influence sur le nombre d'appels : si un département est très urbanisé avec une population plutôt dense, le nombre d'appels d'urgence augmentera en conséquence.

Ainsi, les problèmes de prévision des appels touchent aussi à l'épidémiologie, la géographie et la sociologie. Pour être capable de définir un modèle de prévision des appels fiable, il faut être en mesure de prendre en compte tous ces paramètres. C'est pourquoi une pratique courante est de travailler avec des données historiques (Sariyer, 2018 ; Viglino et al., 2017) ce qui permet d'avoir l'assurance que le comportement de l'arrivée des appels est pertinent (puisqu'il a déjà eu lieu). Cette méthode est pertinente lorsqu'il s'agit de développer de nouvelle manière de fonctionner. Pour une gestion opérationnelle, il reste limité. C'est pourquoi des modèles de prévision des appels basés sur les données historiques ont été développés. Une bonne prévision des appels est très importante, car c'est sur celle-ci que les responsables d'équipes vont se baser afin de constituer les plannings des ressources notamment. (L. H. Brown et al., 2007)

On trouve plusieurs types de modèles de prévision. Les premiers modèles rencontrés sont ceux qui se basent sur des statistiques descriptives, (Hall, 1971) ils sont néanmoins très vite limités par la quantité et la qualité des données disponibles. Certains d'entre eux sont particulièrement intéressants, car ils prennent en compte des données socio-démographiques pour effectuer leurs prévisions (Aldrich et al., 1971). McConnel et al. ont également développé un modèle de régression prenant en compte l'âge de la population (McConnel et al., 1998).

Un autre type de modèle de prévision se base sur des séries temporelles, ce qui permet de corriger certaines faiblesses des techniques classiques de régression. Le modèle de (Baker et al., 1986) a la particularité de prendre en compte les appels urgents ainsi que les appels non urgents dans ses prévisions en utilisant un modèle de lissage exponentiel de type Holt Winters. (Channouf et al., 2007) quant à eux testent plusieurs modèles de série temporelle pour être capable de générer des appels heure par heure pour les systèmes de santé de Calgary en Alberta (Canada). Une évolution importante dans les prédictions des appels vient avec l'étude de (Setzler et al., 2009) qui pour la première fois se rendent compte que l'arrivée des appels dépend du jour de la semaine et de l'heure de la journée. Ils développent alors un réseau de neurones afin de prévoir l'arrivée des appels. Les performances de leur réseau sont comparées aux données historiques afin d'évaluer la précision de cette méthode. Une autre étude propose de prédire les appels dans le temps et dans l'espace afin d'avoir une prévision géographique d'où l'appel pourrait survenir (Vile, 2013). Pour cela l'auteur utilise une technique non paramétrique pour l'analyse des séries temporelles connue sous le nom d'analyse à simple spectre (*single spectrum analysis*), il teste ce modèle sur un service d'ambulance en Écosse et leur méthode permet d'obtenir des prévisions sur le long terme de meilleure qualité ainsi que des prévisions sur le court terme de qualité comparable aux méthodes déjà existantes. D'autres études comme celle de (Grekousis et al., 2019) utilisent une approche par réseaux de neurones pour prévoir la localisation géographique d'un appel. Ils proposent de considérer les appels d'urgence comme des ensembles de points interconnectés dans l'espace qui créent des trajectoires dans le temps. Ils combinent ensuite leur réseau de neurones à un algorithme évolutif pour prédire la localisation future de la demande en soin.

Les problèmes de planning

Les problèmes de planning se décomposent globalement en deux parties, la création du planning et l'affectation des équipes sur ce planning. Dans la littérature on retrouve ces deux problèmes respectivement sous le nom de *scheduling* et *rostering*. (Ernst et al., 2004a)(Ernst et al., 2004b) proposent une revue de littérature autour de ces problèmes en analysant les applications, les méthodes et les modèles utilisés pour résoudre ces problèmes. Dans les systèmes de santé, les problèmes de planning du personnel infirmier et médical ont été très largement étudiés (voir la revue de littérature de (Cheang et al., 2003)).

On trouve cependant un peu moins de travaux concernant les problèmes de planning directement en lien avec les EMS. Bradbeer et al. proposent d'utiliser un algorithme évolutif pour fournir un planning et des affectations acceptables en prenant en compte le nombre et le placement des effecteurs comme données et non comme paramètres

(Bradbeer et al., 2000). On trouve des études qui cherchent à résoudre à la fois le problème de dimensionnement, de localisation, et de planification comme dans les travaux effectués par (Erdoğan et al., 2010). Dans leurs travaux, ils appliquent dans un premier temps une recherche de voisinage pour résoudre le problème de localisation, puis ils utilisent les résultats dans des modèles de programmation en nombres entiers pour résoudre le problème de planification. Rajagopalan et al. (Rajagopalan et al., 2011) présentent également une approche en deux étapes pour l'affectation des plannings aux équipages. Au cours de la première étape, ils résolvent un modèle dynamique de couverture en utilisant un algorithme tabou, tandis qu'un modèle de programmation en nombres entiers est présenté au cours de la deuxième étape pour résoudre le problème d'affectation des plannings. Enfin, Li et Kozan (Li et al., 2009) proposent des modèles en deux étapes qui utilisent aussi des techniques de programmation non linéaires en nombres entiers. Dans la première étape, on détermine les plages horaires travaillées et le nombre d'équipes nécessaires sur chacune d'elles. Les résultats de la première étape sont utilisés comme entrée pour la deuxième étape, qui détermine un planning équilibré pour les effecteurs.

Néanmoins, il y a assez peu de travaux sur l'effet des plannings sur la partie amont, on note quelques études comme celle de Montassier et al. (Montassier et al., 2015) qui ont travaillé sur l'impact de la planification du travail sur les performances des centres d'appels d'urgence en France. Ils ont notamment montré que la durée passée en poste pouvait avoir un impact négatif sur le temps de traitement des appels.

Le planning des opérateurs du centre d'appels est à faire en adéquation avec le volume d'activité du centre d'appels. Il doit aussi respecter les contraintes légales qui peuvent exister dans le milieu : comme le nombre d'heures consécutives maximales autorisées ou encore des temps de pause de 10 minutes toutes les deux heures. (Moutard et al., 2019). (Sariyer, 2018) a par ailleurs travaillé sur les données historiques d'un centre d'appels d'une grande ville en Turquie. Il développe une méthode basée sur le problème du vendeur de journaux pour définir des plannings optimisés pour le centre d'appels. Cette méthode leur permet de constituer des plannings prenant en compte la variation de la demande à l'échelle de la semaine.

En France, le nombre d'ARM nécessaire est défini par le volume des appels entrants et selon un taux d'occupation maximal cible des ARM. Pour rappel, la Mission d'Expertise et d'Audit Hospitaliers (MeaH) (Hospitalier (MeaH)., 2008) préconise un taux d'occupation des ARM entre 20 et 40% de leur temps de travail. Penverne et al. conseillent un taux d'occupation des ARM de 23% (Penverne et al., 2017). La recommandation pour fixer le nombre d'ARM est qu'ils devraient décrocher en moyenne 7,2 appels par heure. En ce qui concerne les MR, leur nombre est fixé en fonction du nombre de dossiers de régulations médicales. En effet, un MR ne doit pas dépasser 6 dossiers de régulations médicales par heure (France et al., 2015).

La prévision des appels est un problème fondamental de l'étude des centres d'appels d'urgence, car ce sont ces prévisions qui serviront de base pour résoudre les problèmes de dimensionnement des ressources, de localisation des équipes d'effecteurs, de planning des équipes, etc. Une difficulté de ce domaine d'étude est qu'il faut être en mesure de prévoir les appels aussi

bien dans le temps que dans l'espace. Ainsi, il est tout aussi important de prévoir quand l'appel va arriver, que la zone géographique dont il sera issu. Aringhieri et al. soulèvent le manque de solutions technique à ce sujet dans la littérature actuelle et notent que des propositions originales sur le sujet pourraient être un véritable atout pour le pilotage en temps réel des systèmes d'urgence (Aringhieri et al., 2017).

1.3.3 Les problèmes opérationnels

Les processus de gestion des appels

Un centre d'appels peut être considéré comme un enchaînement de files d'attente et de serveurs. Les SAMU reçoivent plusieurs types d'appels, par des canaux différents. Ce qui donne lieu à un traitement de l'appel qui peut varier en fonction des caractéristiques de ce dernier. Ces spécificités se doivent d'être modélisées en matière de processus. Cela permet alors de prendre du recul sur la gestion des appels au sein de l'organisation et d'envisager des améliorations. La gestion des appels nous fait nous questionner sur le rôle des ressources et notamment celui des ARM qui décrochent l'appel. Doivent-ils être purement généralistes et être capables de traiter tous les appels ? Ou encore faut-il mieux qu'ils soient spécialisés afin de traiter plus rapidement certains types d'appels ? Enfin, est-il possible d'envisager un compromis entre spécialisation et généralisation pour faire en sorte que les appels soient traités le plus rapidement possible ?

La partie amont concernant l'organisation et le management des processus des centres d'appels présente quelques travaux intéressants. Aboueljinane travaille ainsi sur le SAMU du Val-de-Marne et analyse l'impact de l'augmentation du nombre de ressources pour le décroché des appels. (Aboueljinane, 2014) Elle montre que l'ajout de ressources n'a pas forcément un impact significatif sur les indicateurs de performance. Son analyse montre aussi qu'un gain de quelques secondes entre le décroché de l'appel et l'envoi de moyens peut avoir un effet positif sur le temps de prise en charge total des appels et sur le taux de couverture des interventions en moins de 20 minutes. Toutefois, dans ces mêmes travaux, on ne trouve pas de proposition pour atteindre cette diminution du temps de gestion des appels.

Lewis et al. s'intéressent aux problèmes de management des ressources dans un centre d'appels d'urgence (Lewis et al., 2007). Comme le coût des agents correspond à 70% des coûts d'un centre d'appels, c'est un sujet d'intérêt majeur pour les centres d'appels. Plusieurs facteurs rendent ce terrain propice à la recherche tels que la complexité mathématique et opérationnelle, les files d'attente multiples, le caractère aléatoire de l'arrivée des appels ainsi que le caractère aléatoire de la durée de ceux-ci, mais aussi le côté multicompetences des agents et la diversité grandissante des types d'appels. Les auteurs utilisent la simulation à base d'agents pour modéliser le centre d'appels et gérer les files d'attente. Le modèle bien que défini dans cet article n'est pas testé et les résultats ne sont pas présentés.

Van Buuren et al. testent l'impact de la variation du nombre d'opérateurs sur le temps d'attente avant décrochés (M. v. Buuren et al., 2015). Les auteurs ont divisé

les appels en trois catégories, forte, moyenne et basse priorité. La simulation tourne sur une journée qui reçoit 2000 appels. L'étude montre qu'un centre d'appel avec 3 ARM peut décrocher jusqu'à 1600 requêtes par jour, 2 ARM peuvent prendre jusqu'à 650 requêtes et 1 ARM peut prendre 100 requêtes. Ils notent que leur étude pourrait être améliorée en prenant en compte le lien entre temps de prise en charge de l'appel et charge de travail. En réalité, ils ont constaté que le temps de prise en charge d'un appel est plus faible lorsque la charge de travail est élevée et que cela pourrait avoir une influence sur les performances des centres d'appels. De plus, il suggère qu'il serait intéressant de comparer leur étude avec des centres d'appels où les ressources font à la fois le décroché de l'appel et l'envoi de moyen afin de comparer les performances.

Van Buuren et al. comparent trois modèles de centre d'appels avec des classes d'ARM différents (M. v. Buuren et al., 2017). Ils différencient les ressources qui prennent les appels de celles qui s'occupent d'envoyer les moyens. Ils testent trois scénarios (1) les ressources ont des tâches séparées (2) les ressources sont généralistes (3) certaines ressources sont spécialisées, d'autres sont généralistes. Ils font l'hypothèse que le temps de prise en charge est indépendant de la charge de travail. Ce qui pourrait être amélioré dans un nouveau modèle. Les auteurs, pour des recherches futures, proposent de considérer les centres d'appels d'urgence comme des centres d'appels multiressources avec certaines assignées à des appels de gravité basse et d'autres pour des appels d'urgence vitale. De plus, ils proposent de tester ce modèle sur d'autres centres d'appels qui ne sont pas dédiés à l'urgence pour voir si les résultats peuvent être généralisés.

Montassier et al. proposent enfin une étude quantitative sur les horaires de travail des ARM en France sur le SAMU de Nantes (Montassier et al., 2015). Leurs données s'étendent sur une année entière sur quatre créneaux de travail dans la journée (matin, journée, soir et nuit profonde). Ils s'intéressent à l'impact de la durée des créneaux de travail sur la durée des appels ainsi qu'à la rotation des équipes sur ces horaires. Ils observent que la durée d'appel tend à augmenter (jusqu'à 11,9% d'augmentation) durant les créneaux de nuits et les créneaux de plus de 8h. Ces créneaux de nuit tendent à augmenter la moyenne de temps de communication de 14 secondes. Le fait d'avoir des créneaux de 8h tend à augmenter le temps de communication de 6,4%. Considérant cette augmentation, les auteurs suggèrent que des améliorations pourraient être faites sur les durées de créneaux de travail afin de limiter cette augmentation du temps de traitement.

Les outils d'aides à la décision pour les problématiques cliniques

De nombreux articles dans la littérature sont issus de revues médicales et traitent l'amélioration de la phase de diagnostic, notamment en utilisant la linguistique.

Penn et al. montrent que le choix des mots au décroché de l'appel peut faire gagner jusqu'à 4 secondes sur le temps moyen de traitement de l'appel. Les premiers résultats obtenus les incitent à continuer les recherches sur l'analyse des conversations dans les centres d'appels d'urgence (Penn et al., 2017).

Riou et al. montrent qu'avec l'étude des bandes enregistrées des prises d'appels il est possible d'étudier l'impact des mots choisis sur le temps de diagnostic et la durée

de l'appel. Un autre sujet abordé concernant le diagnostic est « Comment éviter le sur-triage et le sous-triage en utilisant un diagnostic basé sur des critères ? ». Encore aujourd'hui le diagnostic effectué dépend pour beaucoup du médecin qui l'effectue. Toutefois des recherches sont menées pour extraire les critères spécifiques à certaines pathologies afin de créer des procédures de diagnostic (Riou et al., 2017) .

Dami et al. soulignent qu'il n'existe pas de métrique universelle pour évaluer la pertinence de l'envoi de moyen ce qui rend difficile l'évaluation des systèmes d'urgences préhospitalières. Ils mettent en place un système d'évaluation de l'envoi de moyens et évaluent la pertinence de ces choix sur une période d'un mois ayant donné lieu à 29000 envois primaires. Selon leurs critères les taux de sur triage (le fait de trier des appels comme urgent alors qu'ils ne le sont pas) et de sous triage (le fait de trier des appels comme non-urgent alors qu'ils le sont) sont respectivement de 78 et 4,6% (Dami et al., 2015).

Snooks et al. cherchent à identifier les critères permettant d'identifier les chutes chez les personnes âgées. Ils mènent une étude qualitative sur des données passées. Le but est de définir des codes d'envoi de moyen pour identifier rapidement les chutes et envoyer des moyens adaptés (Snooks et al., 2011).

Blomberg et al. proposent une nouvelle méthode utilisant du machine learning afin d'identifier les cas d'arrêts cardiaques en se basant sur les bandes enregistrées. Leurs résultats montrent que l'algorithme permet d'obtenir d'avoir une sensibilité supérieur à celle d'un ARM, toutefois les ARM obtiennent de meilleurs résultats en matière de spécificité concernant l'identification des arrêts cardiaques. Ce genre d'outil d'aide à la décision basé sur de l'Intelligence Artificielle pourrait jouer, très rapidement, un rôle très important dans l'amélioration des performances des centres d'appels (Blomberg et al., 2019).

Le support au diagnostic est un sujet qui fait l'objet de nombreux travaux cliniques. La plupart des publications identifiées ont été faites dans des journaux médicaux. Le constat est que si les centres d'appels ont des fonctionnements différents, les protocoles de diagnostic sont également différents. Bohm et al. dans leur étude, montre qu'il demeure toutefois une absence de niveau de preuve de l'exactitude des systèmes de triage (Bohm et al., 2018).

Les problèmes de routing et de relocalisation

Une fois qu'on a identifié le nombre de bases SMUR, qu'on a décidé où les placer, et que l'on a décidé du nombre d'équipes dans chacune des bases. Il reste encore des problèmes à résoudre pour optimiser la gestion (ou le pilotage) d'une intervention donnée. En effet, en fonction du lieu de l'intervention, il est question de déterminer la base la plus appropriée pour faire partir l'ambulance. Quels chemins doit emprunter cette ambulance ? Une fois que son intervention est terminée, est-ce qu'elle rentre directement à son entrepôt ? Est-ce qu'elle doit rentrer au même entrepôt ? Et si une nouvelle intervention intervient avant qu'elle ne rentre, est-ce que je modifie l'itinéraire de cette ambulance ? Est-ce que j'en envoie une nouvelle ? C'est sur ce genre de questions que des scientifiques tels que (Aboueljinane, 2014 ; M. v. Buuren

et al., 2018; Koch et al., 2003) travaillent. L'objectif de toutes ces questions est de réduire le temps de prise en charge des interventions ainsi que le temps pour se rendre sur le lieu de l'intervention. Toutefois un autre objectif est de réduire les coûts, car si on peut optimiser les trajets, on réduit le kilométrage parcouru et par conséquent les frais d'essence, d'entretien, etc. (Aboueljinane et al., 2014) (M. v. Buuren et al., 2015).

Lorsqu'on parle de problème de *dispatching* cela fait référence à l'affectation d'une équipe à la gestion d'un appel. Cette notion concerne essentiellement la partie aval lorsque la décision d'envoyer des effecteurs est prise. L'enjeu ici est de choisir l'équipe qui pourra arriver le plus rapidement sur place ou qui parcourra le moins de kilomètres. Plusieurs règles de déploiement ont alors été testées. Dans leur projet, (Koch et al., 2003) étudient les règles de déploiement de la Croix-Rouge autrichienne. Ils testent une première règle où toutes les équipes mobiles peuvent intervenir sur l'ensemble du territoire en utilisant une heuristique proche optimale pour l'affectation équipe-intervention. Ils utilisent la règle de l'équipe disponible la plus proche. Une deuxième règle de déploiement teste une réponse plus régionalisée où les équipes sont affectées à des zones géographiques pré-définies et ne peuvent intervenir que si l'appel se trouve dans cette zone. Les résultats montrent que la première règle plus globale donne de meilleurs résultats en termes d'attente des patients, et de kilométrage total parcouru.

Une sous catégorie du problème de localisation correspond à la relocalisation des véhicules d'urgences. Cela consiste à redéployer les véhicules non utilisés lorsque les autres équipes sont en intervention. La particularité de ce problème est sa dimension dynamique, car il faut connaître l'emplacement des équipes de secours en temps réel. De ce fait, on trouve deux grandes manières d'aborder ces problèmes : les problèmes mixtes à plusieurs périodes ou les problèmes véritablement dynamiques. Le coeur de ce sujet est de se concentrer sur la reconfiguration de la flotte de véhicules disponibles pour assurer un service optimal en tout point du territoire. Bien que ce soit un problème dynamique en temps réel, les chercheurs l'ont approché dans un premier temps en proposant des reconfigurations multi-périodiques. Cela permet de fournir aux décideurs l'emplacement des véhicules pour chaque période à l'aide d'un tableau de conformité. L'inconvénient de cette méthode est que les véhicules doivent beaucoup se déplacer afin de maintenir un niveau de conformité acceptable, ce qui peut rajouter des kilomètres parcourus sans pour autant améliorer la qualité de service pour les appels (Bandara et al., 2012). De nombreux travaux sont disponibles à ce sujet, on peut citer notamment ceux de Gendreau et al. et leur modèle multi-période pour la relocalisation des ambulances de Montréal (Gendreau et al., 2001). Leur méthode consiste à résoudre un programme d'optimisation mixte en nombres entiers en temps réel. Leurs résultats montrent que si les appels d'urgence sont très rapprochés, le temps de calcul pour le redéploiement peut être trop important ce qui compromet l'efficacité de la relocalisation. De plus, il faut que le redéploiement des équipes soit possible, car il peut y avoir des cas où la mutualisation des ressources pour les transports pré et intra hospitalier impose aux équipes de retourner à leur base après une intervention. Ce modèle a été amélioré par d'autres auteurs pour prendre en compte plus de paramètres comme l'incertitude temporelle liée au temps de trajets des ambulances (Schmid et al., 2010). Belanger et al proposent une revue

de littérature pour évaluer les différentes méthodes de redéploiement (Bélangier et al., 2016). Le constat aujourd'hui est qu'il existe beaucoup de méthodes de redéploiement, mais que nous ne sommes pas en mesure d'évaluer laquelle serait la meilleure en pratique. De plus, ce sont des problèmes qui nécessitent de bonnes capacités de calcul, les méthodes doivent donc continuer d'évoluer pour être en mesure de traiter ces problèmes dans des temps raisonnables.

Choix de l'établissement d'accueil

Le choix de l'établissement d'accueil est un problème crucial, car le système de soin d'urgence va contribuer à l'engorgement des urgences physiques au sein du système hospitalier. Ainsi il faut être en mesure de prendre en compte cet effet pour guider le patient vers l'établissement de santé qui lui apportera le meilleur soin adapté à son besoin réel et celui qui pénalisera le moins le système de santé dans son ensemble. Une règle assez commune dans la littérature scientifique internationale consiste à envoyer le patient dans l'hôpital le plus proche du lieu de l'intervention (M. v. Buuren et al., 2012 ; Fitzsimmons, 1971). Néanmoins, la proximité n'est pas le seul critère qu'il faille prendre en compte. Dans les faits les patients ne sont pas toujours transportés dans le service d'urgence le plus proche : il faut prendre en compte le besoin de soin réel du patient avant la proximité de l'établissement le plus proche (Ingolfsson et al., 2003). D'autres critères comme la disponibilité des lits, les spécialités et installations présente dans l'établissement de santé ou encore l'engorgement du service d'urgence de l'établissement cible sont parfois considérés. De plus, la prise en compte du trafic en temps réel pourrait avoir une incidence sur le choix de l'établissement d'accueil. C'est pour cela que d'autres auteurs ont fait le choix de se baser sur des répartitions empiriques pour leurs études (Ingolfsson et al., 2003). Les situations de crise de grande ampleur modifient les paramètres pour le choix de l'établissement, que ce soit lors d'attentats, de catastrophes naturelles à grande échelle (Wang et al., 2007), ou encore de pandémie comme nous avons pu le voir récemment avec des transferts inter-régionaux. Il est alors intéressant de travailler avec des listes d'établissements potentiels. Une autre étude intéressante est celle de (Miller et al., 2020) qui analyse l'effet de la fermeture des hôpitaux de proximité sur le temps de réponse, le temps d'accès au traitement, et le temps de service des systèmes d'urgences. Leur étude se base sur des données nationales récoltées entre 2010 et 2016 aux Etats-Unis issus du *National EMS Information System, Area Health Resource, and Center for Medicare & Medicaid Provider of Service files*. Leurs résultats montrent que la fermeture des hôpitaux locaux augmente le temps moyen de transport de 2,6 minutes et le temps de réponse moyen augmente de 7,2 minutes. Cela montre à quel point le choix de l'établissement d'accueil peut avoir une incidence sur le temps de prise en charge du patient.

La gestion opérationnelle de l'appel au niveau du centre d'appels est un sujet qui est assez peu étudié dans la littérature. Les principaux défis à relever concernent la gestion des ressources humaines du centre d'appels. L'influence des horaires de travail ainsi que de la segmentation (spécialisation) ou la polyvalence des ressources sont des sujets qui doivent encore être investigués.

L'étape de diagnostic de l'appel pourrait être aussi améliorée en travaillant sur des procédures de diagnostic téléphonique plus ciblées ou encore en développant des outils d'aide à la décision à destination du personnel médical.

Concernant la partie aval, ce sont les problèmes de déploiement et de relocalisation en temps réel qui présentent un intérêt. Les difficultés résident dans la prise en compte des facteurs extérieurs tels que le trafic ou la météo ainsi que la gestion dynamique en temps réel des véhicules.

Enfin, le choix de l'établissement d'accueil pour le patient nécessite aussi de prendre en compte les critères extérieurs qui peuvent rallonger le trajet entre le patient et l'établissement de santé. Une fois encore, il serait intéressant de développer des outils d'aide à la décision basés sur différents critères permettant de prendre en compte l'état et l'engorgement des établissements de santé ainsi que le choix du patient.

1.3.4 Synthèse : Un déséquilibre des études entre la partie amont et aval.

Le dénominateur commun de toutes ces études est d'améliorer les performances du système de soin préhospitalier. Cela consiste en une double approche d'une part qualitative avec une meilleure analyse du besoin de soin et des décisions adaptées d'autre part quantitative par la gestion des délais. Les délais tentent d'être réduits en faisant en sorte que les solutions proposées réduisent le temps entre le moment où un appel d'urgence est lancé et la prise en charge du patient dans un établissement de santé. On constate un nombre significatif d'études qui travaillent sur la réduction des temps de trajets dans les interventions.

Tableau 1.1 – Synthèse des défis à relever

| | Type de problème | Challenge |
|--------------|-------------------------------|---|
| Stratégique | Dimensionnement | Définir le nombre de ressources nécessaires |
| | Localisation | Placer les équipes de manière à optimiser le maillage |
| | Gestion des nouveaux supports | Comment intégrer les nouvelles technologies ? |
| Tactique | Prévisions des appels | Prévoir l'arrivée temporelle et spatiale des appels |
| | Planning | Planifier les équipes de manière à répondre à la demande |
| Opérationnel | Gestion des appels | Proposer des processus de gestion pour minimiser l'attente et maximiser le temps d'occupation |
| | Aide à la décision (clinique) | Développer des outils pour aider le diagnostic des praticiens |
| | Déploiement et relocalisation | Gérer les équipes d'effecteurs en temps réel pour diminuer les temps de prise en charge |
| | Choix établissement accueil | Choisir l'établissement d'accueil en prenant en compte les contraintes extérieures et le choix du patient |

Un premier constat est que la littérature est riche d'une grande quantité de travaux sur la partie amont pour la résolution de problèmes d'affectation, de tournées de véhicules, de relocalisation. La résolution de ces problèmes permet de réduire le temps d'intervention des équipes, ainsi que le nombre de kilomètres parcourus. Ces problèmes sont principalement étudiés depuis la fin des années 1980 (Fujiwara et al., 1987). En comparaison, les problèmes concernant la partie aval est plus récente avec une volumétrie plus modeste. Pour illustrer ce propos, nous avons effectué deux requêtes sur la plateforme de bases de données bibliographiques et bibliométriques Web of Science⁹. Nous avons fait un tri sur les thématiques de Gestion, d'Informatique

9. <https://www.webofknowledge.com>

et de Génie industriel. Nous avons analysé le nombre de publications par années sur les dernières décennies. Les figures 1.13 et 1.14 montrent les résultats de cette requête. On constate que les sujets concernant la partie aval sont plus nombreux (coefficient de 1 pour 10) et qu'il y a une augmentation des études avec le temps. Du côté des centres d'appels, les articles sont moins nombreux et plus récents avec des publications surtout après les années 2000. Bien que ceci ne constitue pas une revue de littérature systématique au sens strict du terme, cela illustre bien l'écart qui existe entre les deux parties de la chaîne de l'urgence.

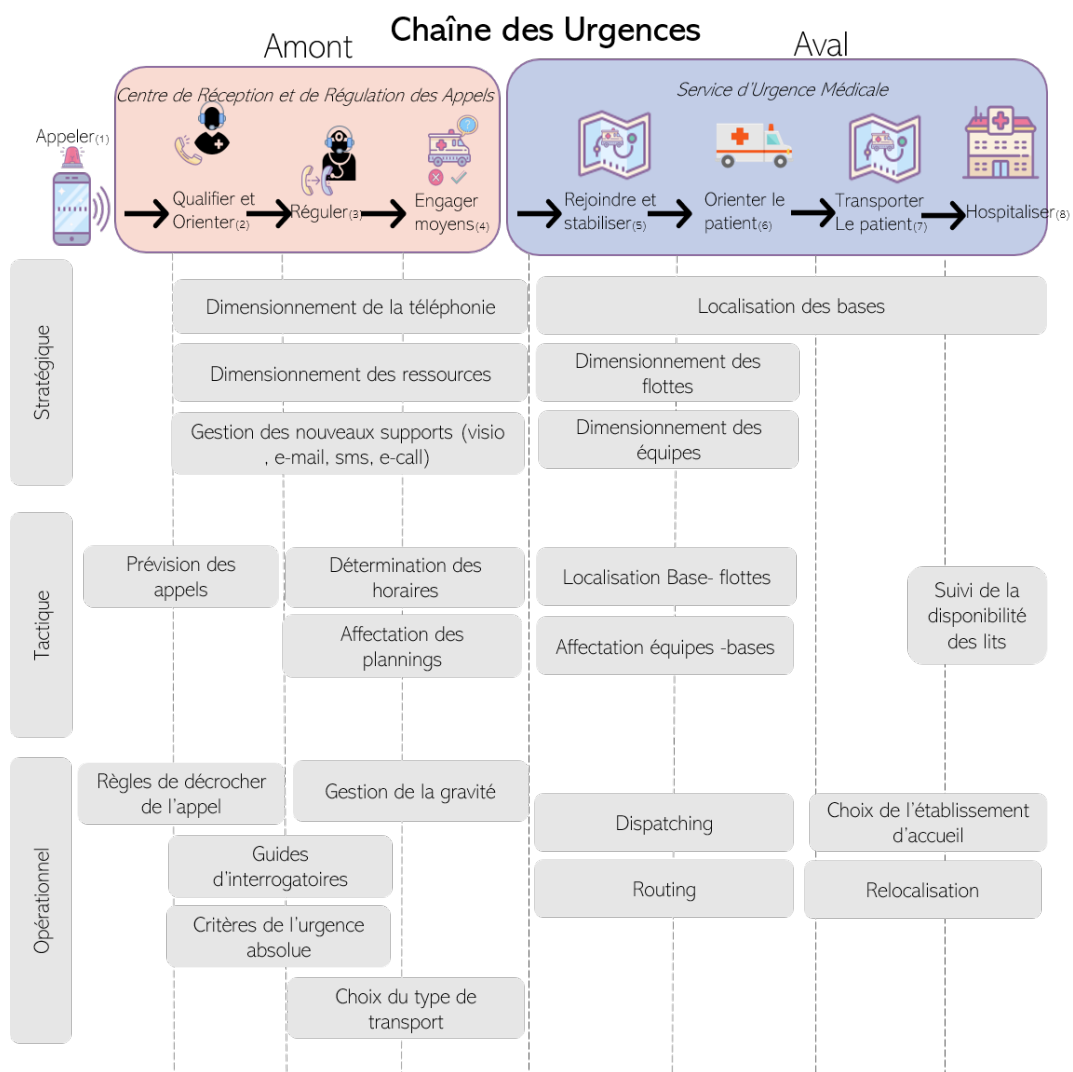


Figure 1.12 – Positionnement des types de sujets en fonction de l'activité de la chaîne des urgences

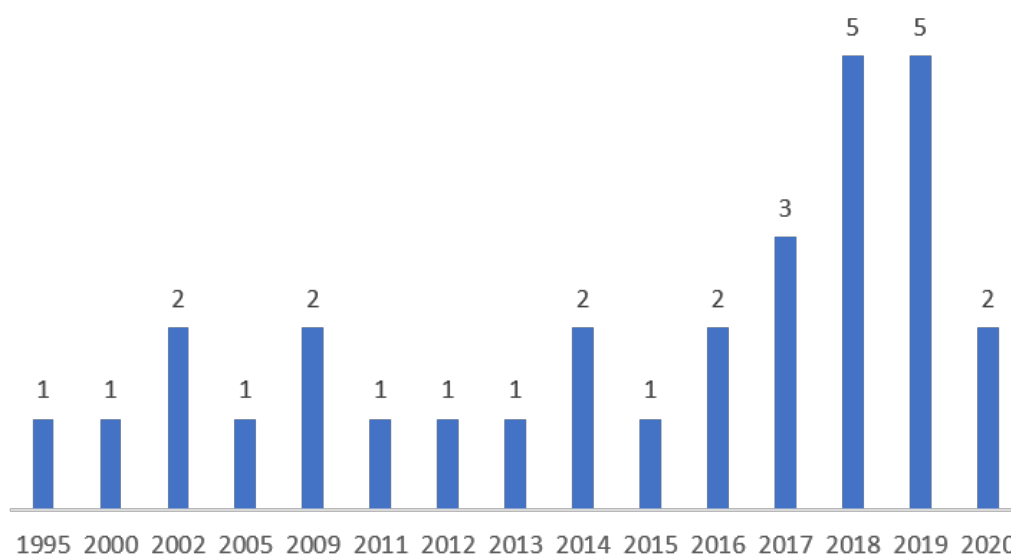


Figure 1.13 – Résultats de recherche du 02/06/2020 sur Web of Science avec la requête :(("Emergency Medical Service*" OR EMS) AND((Call OR Contact OR Communication) AND Center))

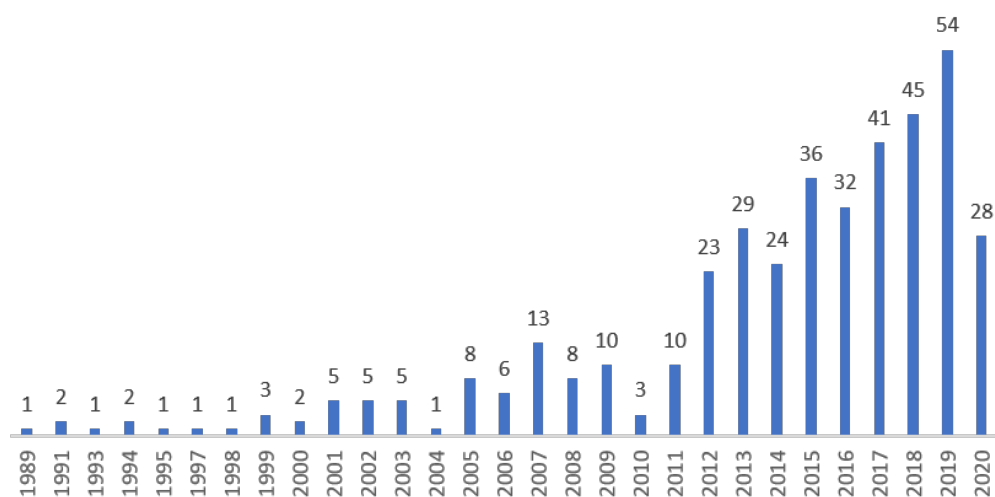


Figure 1.14 – Résultats de recherche du 02/06/2020 sur Web of Science avec la requête : "Emergency Medical Service*" OR EMS) AND (Ambulance* OR transport*)