

La crise du système de santé français

Comment désengorger les urgences ?

Par Benjamin Legros, Frédérique Vidal et Sean Scull

Septembre 2025



SKEMA PUBLIKA

SKEMA PUBLIKA est un think tank indépendant qui a pour objectif de produire une pensée internationale, accessible et loin des codes formatés pour alimenter le débat public et mieux éclairer en amont les décideurs nationaux et internationaux.

Adossé à SKEMA Business School, le think tank aborde des sujets politiques et sociétaux ayant trait aux politiques publiques, sur lesquels SKEMA a une légitimité à s'exprimer. Il les aborde sous l'angle des signes précurseurs, anticipe et formule des recommandations pour « l'après ». Il adopte une approche multidisciplinaire et hybride du traitement de l'information, associant intelligences humaine et numérique. Il s'appuie sur la dimension internationale et transculturelle de SKEMA, présente sur cinq continents et riche de milliers d'étudiants et de centaines de chercheurs.

TABLE DES MATIÈRES

Introduction	4
Fondements théoriques des files d'attente dans les services d'urgence.....	7
La théorie des files d'attente.....	7
particularité des services d'urgence, l'inapplicabilité des formules classiques	7
La théorie des jeux pour comprendre le comportement des patients	10
La sur-congestion dû à l'absence de considération des externalités négatives.....	10
L'aversion au risque d'attente des patients : un amplificateur de congestion	10
Enjeu critique de la priorisation des patients	13
Formation du personnel soignant et non soignant et son rôle dans l'optimisation des flux.....	15
Identification des rôles clés du personnel soignant et non soignant.....	15
Programmes de formations ciblées	16
La place des centres d'appels dans la gestion des urgences.....	18
Rôle actuel des centres d'appels.....	18
Amélioration de la coordination et du partage d'informations	19
Défis et perspectives	21
applicabilité des modèles mathématiques	21
Intégration des facteurs humains et organisationnels.....	21
Les perspectives futures.....	22
Conclusion	23
AUTEURS	24

INTRODUCTION

« Comme avant la pandémie, les soins urgents sont redevenus le réceptacle de la plupart des débordements du système de santé. Dernier dispositif de recours, les structures d'urgence sont elles-mêmes mises en difficulté car sollicitées, outre les cas de réelle urgence, par les patients qui ne parviennent pas à obtenir de rendez-vous rapidement chez un médecin ou par des patients atteints de maladies chroniques non suivis par un médecin-traitant¹. »

Ces propos sont issus du rapport « *L'accueil et le traitement des urgences à l'hôpital* » publié par la Cour des comptes en novembre 2024. Le constat est glaçant, il y est décrit un système de santé à bout de souffle à cause de services saturés qui rencontrent une difficulté à gérer une demande toujours plus importante. Pour sortir l'hôpital public de cette impasse, le rapport de la Cour des comptes souligne une révision indispensable du parcours des patients avant leur prise en charge par les urgences. En effet, les services accueillant les patients pris en charge pour soins non programmés sont des systèmes complexes confrontés à une demande imprévisible et à des ressources limitées, entraînant souvent de l'attente. Ces situations entraînent des conséquences directes sur la qualité des soins, la sécurité des patients et l'efficacité opérationnelle.

Ce *policy paper* SKEMA Publika s'appuie sur les travaux de Benjamin Legros pour montrer comment les outils et les modèles mathématiques peuvent être utilisés pour analyser, comprendre et optimiser la gestion des files d'attente dans les services d'urgence, en tenant compte de facteurs humains et organisationnels. Il explore l'application des théories des files d'attente et d'autres approches mathématiques pour aborder les défis des services d'urgence, en modélisant les processus d'arrivée de patients, analysant l'aversion aux risques, proposant des solutions de priorisation des patients, et en mettant l'accent sur la formation du personnel et le rôle des centres d'appels.

Le sujet traité dans ce *policy paper* est d'actualité au vu des difficultés budgétaires auxquelles l'Etat français et plus particulièrement l'hôpital public doivent faire face. Les statistiques de l'INSEE montrent qu'en 2022 la Sécurité sociale finance à hauteur de 80 % les dépenses de santé des Français². Cela correspond à 11,9% du PIB qui est consacré aux soins de santé³. Ainsi, la santé est une des principales dépenses de l'Etat et en plus elle pèse sur le déficit public. Selon les estimations, en 2024 le déficit des hôpitaux était de 3,4 milliards d'euros selon la Fédération hospitalière de France⁴. Ce *policy paper* s'inscrit aussi dans un contexte politique où le gouvernement cherche à réduire la dette à 3 305,3 milliards d'euros (2024). Ce dernier prépare un plan de redressement budgétaire visant 40 milliards d'euros d'économies d'ici 2026. Cette analyse se veut source d'inspiration dans cet effort de réduction des dépenses en apportant des solutions concrètes en termes de politiques publiques pour sortir de l'impasse de la congestion dans les urgences et de fil en aiguille contribuer à améliorer le système de santé français. Notamment, en allant à l'encontre de l'idée selon laquelle pour pallier le problème de la congestion, il est uniquement nécessaire d'allouer plus de ressources matérielles, financières et humaines. Au contraire, nous présentons une vision alternative selon laquelle il est moins coûteux de repenser le routage des tâches et le rôle des agents, d'où l'utilité d'associer les modèles mathématiques et les nouvelles technologies du numérique à cette réflexion. Ainsi, la problématique que nous posons est la suivante :

¹ *L'accueil et le traitement des urgences à l'hôpital*. (2024). Cour des Comptes. <https://www.ccomptes.fr/fr/publications/laccueil-et-le-traitement-des-urgences-lhopital>

² *Dépenses de santé – France, portrait social* | Insee. (2023). <https://www.insee.fr/fr/statistiques/7666887?sommaire=7666953>

³ La rédaction. (2022). Dépenses de santé : elles ralentissent en 2022 avec le recul du Covid-19. Vie Publique. <https://www.vie-publique.fr/en-bref/291168-depenses-de-sante-elles-ralentissent-en-2022-avec-le-recul-du-covid-19>

⁴ Avec AFP, F. (2025, 22 mai). Les hôpitaux ont connu une dégradation « sans précédent » de leurs finances en 2023. *Franceinfo*. https://www.franceinfo.fr/sante/hopital/les-hopitaux-ont-connu-une-degradation-sans-precedent-de-leurs-finances-en-2023_7265832.html

- *Dans un contexte budgétaire difficile comment optimiser la gestion du flux des patients dans les urgences ?*

Dans une première partie nous allons explorer les fondements théoriques des files d'attente dans les services d'urgence. Dans une deuxième partie nous verrons en quoi la théorie des jeux peut aider à comprendre le comportement des patients. Dans une troisième partie nous soulignerons l'enjeu critique de la priorisation des patients. Dans une quatrième partie nous analyserons en quoi la formation du personnel peut jouer un rôle dans l'optimisation des flux. Dans une cinquième partie nous verrons comment les centres d'appels pourraient contribuer à mieux gérer le flux des patients aux urgences. Enfin dans une sixième partie nous poserons les défis et perspectives qui découlent du routage des tâches dans les services d'urgences. En outre, précisons que dans cette étude, des comparaisons seront aussi faites avec les systèmes d'urgence d'autres pays afin de s'inspirer des politiques publiques mises en place et de produire des recommandations concrètes visant à améliorer l'accès aux soins dans les urgences en France.

CHAPITRE I

FONDEMENTS THEORIQUES DES FILES D'ATTENTE DANS LES SERVICES D'URGENCE

FONDEMENTS THEORIQUES DES FILES D'ATTENTE DANS LES SERVICES D'URGENCE

Les services accueillant des patients nécessitant des soins d'urgence sont soumis à une congestion importante. Selon une étude de la Direction de la recherche, des études, de l'évaluation et des statistiques depuis 2013 le temps d'attente aux urgences a augmenté de 45 min. En 2023, le temps d'attente était de 3 heures et 45 minutes⁵. Par conséquent cette expérience pénible d'attente a conduit de nombreux services à réfléchir à ses causes et éventuellement à trouver des solutions pour la réduire.

LA THEORIE DES FILES D'ATTENTE

Pour comprendre l'attente dans les urgences il est nécessaire de se pencher sur la théorie des files d'attente. Elle est née en 1909 des travaux de l'ingénieur danois A.K. Erlang qui a modélisé mathématiquement « l'attente », comme étant le fait que des usagers ne soient pas servis immédiatement. Ces travaux se sont inspirés du problème de la congestion dans le trafic téléphonique⁶.

Partons d'une question simple : pourquoi attend-on ?

Imaginons la caisse d'un supermarché avec un seul agent : chaque encaissement dure cinq minutes et un client se présente toutes les six minutes. Dans ce cas, aucun client n'attend ; l'agent est libre à chaque nouvelle arrivée. Si, en revanche, un client arrive toutes les quatre minutes, la dynamique change : le premier client ne patiente pas, le deuxième attend une minute, le troisième deux minutes, le quatrième trois minutes, et ainsi de suite. La file s'allonge alors indéfiniment et le système devient instable, sans jamais atteindre un état d'équilibre.

À ce stade, on pourrait croire qu'un système réel ne comporte que des files vides ou infinies... ce qui est loin d'être vrai ! Il manque un ingrédient essentiel : la variabilité. Dans la réalité, les temps de service et les temps inter-arrivées sont aléatoires ; il peut y avoir, par hasard, des arrivées groupées ou des services plus longs. On observe donc une alternance de périodes sans attente et de périodes de congestion, si bien que, même lorsque le temps de service moyen est inférieur au temps inter-arrivées moyen, certains usagers attendent quand même. La théorie des files d'attente permet de modéliser cette variabilité et de calculer, entre autres, le temps moyen d'attente en fonction du temps de service moyen, du temps moyen entre deux arrivées ou encore du nombre d'agents disponibles. Ces formules éclairent les facteurs d'attente et guident la décision. Pour y parvenir, on adopte des hypothèses (souvent simplificatrices) : les arrivées suivent un processus de Poisson, les temps de service sont exponentiels et la discipline de service est « premier arrivé, premier servi ». Dans ce cadre, une chaîne de Markov peut être analysée, donnant des expressions explicites pour le temps d'attente moyen, la distribution des temps d'attente, ou le taux d'occupation des agents. Ces résultats sont essentiels pour décider, par exemple, du routage des patients ou des capacités à installer comme le nombre d'agents, le nombre de salles d'opération⁷.

PARTICULARITE DES SERVICES D'URGENCE, L'INAPPLICABILITE DES FORMULES CLASSIQUES

Dans un service d'urgence hospitalier, les modèles stylisés de files d'attente classiques se heurtent rapidement à la complexité du réel. D'une part, les temps d'examen, de stabilisation, d'imagerie ou d'attente d'un lit d'aval sont longs et très hétérogènes. Le système fonctionne donc le plus souvent en régime transitoire, loin de l'état stationnaire présumé dans les formules de la théorie. D'autre part, le flux d'arrivées n'est pas constant : il suit

⁵ Kachaner, A. (2025, 19 mars). Le temps d'attente aux urgences augmente en France. France Inter. <https://www.radiofrance.fr/franceinter/podcasts/l-info-de-france-inter/l-info-de-france-inter-3029887>

⁶ A. K. Erlang, « The theory of probability and telephone conversations », *Nyt Tidsskrift for Matematik*, b, vol. 20, 1909

A. Erlang, « Solution of some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges », *Elektrotekniker*, vol. 13, 1917

⁷ Cachon, G. and Terwiesch, C. (2019). Operations Management. McGraw Hill Education, New-York, 2nd edition.

des profils horaires, hebdomadaires et saisonniers marqués, d'où une nécessité de modéliser ces flux et de comprendre le comportement des patients. Par exemple on peut noter une forte affluence en soirée, le lundi ou après un épisode grippal puis avec un creux au petit matin, etc. Pour capturer ces variations liées au caractère imprévisible du service d'urgence, il faut recourir à des processus de « poisson non homogènes », à des modèles cycliques ou à des séries temporelles plus élaborées.

La distribution des temps de service s'écarte, elle aussi, de l'exponentielle : on observe souvent des distributions log-normales ou de type phase. S'ajoute la logique de triage : chaque malade est classé selon un niveau de gravité, de sorte que la file d'attente fonctionne à priorités multiples où les urgences vitales préemptent les cas moins sévères — bien loin du simple « premier arrivé, premier servi ». Par ailleurs, une fraction non négligeable de patients quitte la salle d'attente avant d'être prise en charge (*left without being seen*), parfois par renoncement, parfois pour revenir dans les 72 heures lorsqu'une complication survient ; ces comportements de *balking*, *reneging* et *retrial* alourdissent la charge effective et compliquent la planification de capacité.

Ainsi, la congestion n'est pas seulement due au service médical initial : les goulots d'étranglement partagés (laboratoire, radiologie, disponibilité des lits d'aval) provoquent du *boarding*, c'est-à-dire le maintien de patients stabilisés dans des box d'urgence faute de lit d'hospitalisation. On se retrouve alors avec des réseaux de files à capacité finie, des phénomènes de blocage inter-services et un risque clinique dynamique.

Pour traiter cette complexité, la recherche articule des extensions analytiques (processus non stationnaires, distributions de service générales, files à priorités imbriquées, réseaux bloquants, processus de retour) et des approches numériques (simulation par événements discrets, voire jumeaux numériques (*digital twins*)) qui reproduisent minute par minute les parcours patients, le planning du personnel et la disponibilité des ressources. Ces outils éclairent les décisions au-delà de l'intuition : ils quantifient, par exemple, l'effet d'un infirmier supplémentaire, l'impact d'un scanner dédié ou le bénéfice d'une unité d'observation rapide sur les délais d'attente, la saturation du système actuelle et future, et les résultats des politiques de soins sur les résultats de santé.

CHAPITRE II

LA THEORIE DES JEUX POUR COMPRENDRE LE COMPORTEMENT DES PATIENTS

LA THEORIE DES JEUX POUR COMPRENDRE LE COMPORTEMENT DES PATIENTS

Un facteur majeur de la saturation des services d'urgence est l'absence de régulation des flux d'admission. Les patients choisissent librement de se rendre aux urgences plutôt que de fixer un rendez-vous chez leur médecin. Comme ces services sont gratuits, facilement accessibles et dotés de praticiens hautement qualifiés, ils attirent fortement la population, y compris les personnes dont l'état de santé ne relève pas de l'urgence vitale.

LA SUR-CONGESTION DU A L'ABSENCE DE CONSIDERATION DES EXTERNALITES NEGATIVES

La théorie des jeux apporte un éclairage précieux sur la décision des patients qui choisissent malgré tout de se rendre dans un service d'urgence saturé. Le modèle fondateur de Paul Naor (1969)⁸ montre que chaque patient cherche à maximiser une fonction d'utilité décroissante avec le temps d'attente : plus la file est longue, moins elle devrait être attractive. Pourtant, on observe en pratique une congestion persistante et l'absence quasi totale d'autorégulation. Comme l'ont souligné Hassin et Haviv (2003)⁹, ce paradoxe s'explique par la prise en compte incomplète des externalités : un responsable de service admet un patient en considérant à la fois l'attente qu'il subira et celle qu'il infligera aux autres, tandis qu'un individu ne juge que son intérêt propre. Le résultat est un équilibre où la file demeure trop longue et les ressources sont surexploitées.

Pour réduire cette surcharge, plusieurs leviers peuvent être combinés. D'abord, un filtrage à l'entrée peut limiter l'accès aux cas les moins urgents, même si la différenciation entre état réel et état perçu du patient demeure délicate. Ensuite, l'instauration d'une tarification modérée—ou symbolique—à l'admission décourage les passages non essentiels : les patients réellement urgents accepteraient ce coût, tandis que les autres seraient incités à consulter un médecin de ville, au prix toutefois de débats éthiques sur la gratuité des soins. Enfin, miser sur la pédagogie et la sensibilisation aux externalités négatives—comme le proposent Legros et van Leeuwen (2025)¹⁰ avec le concept de « Wait Externality Awareness »—permet de responsabiliser les usagers, même si l'impact de telles campagnes reste incertain et coûteux.

Il n'existe donc pas de solution unique à la congestion des urgences ; cependant, l'articulation pragmatique de ces trois approches—filtrage intelligent, tarification incitative et information du public—peut atténuer la pression sur les services tout en préservant l'accès aux soins pour les situations vraiment critiques.

L'AVERSION AU RISQUE D'ATTENTE DES PATIENTS : UN AMPLIFICATEUR DE CONGESTION

Un article empirique récent montre que les patients ne sont pas seulement sensibles au temps d'attente moyen, mais à l'ensemble de sa distribution¹¹. Les auteurs mettent en évidence une aversion au risque : un patient compare non seulement l'attente moyenne, mais aussi la probabilité d'attendre « plus longtemps que... » au bénéfice espéré du service.

Cette aversion renforce la congestion des files, car elle annule l'effet bénéfique qu'une augmentation de capacité pourrait exercer. Legros, van Leeuwen, de Véricourt et Fransoo (2024)¹² démontrent ainsi qu'accroître la capacité (nombre d'agents, salles d'opération, etc.) peut paradoxalement allonger l'attente à l'équilibre.

⁸ Naor, P. (1969). The regulation of queue size by levying tolls. *Econometrica: journal of the Econometric Society*, 15-24.

⁹ Hassin, R., & Haviv, M. (2003). *To queue or not to queue: Equilibrium behavior in queueing systems* (Vol. 59). Springer Science & Business Media.

¹⁰ Legros, B., & van Leeuwen, J. (2025). Wait Externality Awareness for Socially Responsible Queueing. *Available at SSRN 5182011*.

¹¹ Kagan, E., Hyndman, K. B., & Davis, A. M. (2024). Beyond Averages: How Do Customers Respond to Wait Time Distributions?. *Available at SSRN 4899240*.

¹² Legros, B., van Leeuwen, J., de Véricourt, F., & Fransoo, J. C. (2024). Risk Aversion Undermines Benefits of Capacity Expansion in Services. *Available at SSRN 4977597*.

Lorsque la capacité est accrue, le temps d'attente moyen et sa variabilité diminuent, rendant le service plus attrayant pour les patients avertis au risque ; cet afflux supplémentaire de patients finit par dépasser la nouvelle capacité et les délais redeviennent élevés, et même plus élevés que ceux avant l'augmentation de capacité. Ce phénomène est connu dans le domaine de l'urbanisme. On appelle cela l'« effet autoroute élargie » : comme une voie rapide agrandie qui, en attirant plus de conducteurs, recrée l'embouteillage qu'elle devait résorber. Pour éviter qu'un tel effet rebond n'aggrave la congestion, plusieurs pistes se dégagent : évaluer précisément l'aversion des usagers au risque d'attente, puisque c'est elle qui conditionne la réaction de la demande ; favoriser un réseau de petites unités plutôt que la centralisation dans de grands établissements, afin de limiter l'attractivité excessive liée à la mutualisation ; enfin, afficher de façon fiable les délais d'attente grâce à des modèles prédictifs solides—et, à terme, recourir à l'intelligence artificielle pour un pré-triage automatisé (NLP, *deep learning*) qui oriente les patients en amont, lisse le flux et améliore la précision des estimations et réduit ainsi l'impact de l'aversion aux risques.

CHAPITRE III

ENJEU CRITIQUE DE LA PRIORISATION DES PATIENTS

ENJEU CRITIQUE DE LA PRIORISATION DES PATIENTS

Un élément essentiel de la gestion de la congestion et de la gravité des patients est l'ordonnancement des tâches, réalisé au moyen de politiques de priorisation. La gravité détermine essentiellement l'ordre dans lequel les patients sont pris en charge : on attribue des couleurs (bleu, vert, rouge) correspondant à des niveaux d'urgence, puis on applique, à l'intérieur de chaque classe, une règle « premier arrivé, premier servi » — assortie, bien sûr, des exceptions nécessaires.

Cependant, classer les patients par gravité n'est pas aussi simple qu'il y paraît. Lorsque le service se décongestionne, certains actes jugés moins urgents deviennent plus pressants ; inversement, en période de forte affluence, on peut se permettre de faire attendre des cas modestement graves. De plus, l'urgence ne peut être décrétée sur un diagnostic seul à l'arrivée, mais est aussi la résultante des comorbidités, de l'état aigu du patient, et de son temps d'attente. Il existe donc un lien étroit entre la politique de priorisation (ou de routage) et l'état instantané du système.

Un outil puissant pour définir des politiques vraiment optimales est la programmation dynamique, fondée sur les Processus de Décision Markoviens (MDP). Elle décompose un problème séquentiel en sous-problèmes successifs : à chaque état du système correspond une « valeur » qui traduit le coût futur attendu (actes reportés, abandons, mortalité, etc.) si l'on suit telle ou telle décision. L'équation de Bellman, cœur de la méthode, permet ainsi de remonter de l'instant final vers l'instant présent et d'identifier, à chaque étape, la décision qui minimise le coût total. En pratique, on obtient des règles conditionnelles, souvent très simples : par exemple, « si le nombre de patients de classe rouge dépasse N , déprioriser les nouveaux cas verts jusqu'à ce que la file se résorbe ».

Le livre de Puterman (2014)¹³ décrit en détail l'application de cet outil et montre que les politiques optimales présentent une structure de seuils : au-delà d'un certain nombre de patients en file, il devient rationnel de retarder les cas moins critiques pour éviter une attente excessive aux cas graves déjà présents.

Ces politiques restent toutefois difficiles à déployer : elles exigent un système d'information capable de suivre en temps réel l'état du service et de calculer la décision recommandée. Par ailleurs, à la différence d'autres files (p. ex. centres d'appels), il faut reprioriser les patients lorsque leur attente individuelle s'allonge ; on raisonne donc sur l'attente propre à chaque patient, non plus seulement sur le volume global. Il s'agit de modèle de priorisation dynamique. Trouver l'optimum devient plus complexe, mais Legros, Jouini et Koole (2018)¹⁴ proposent un algorithme d'uniformisation efficace qui calcule ces politiques « basées sur l'attente ».

Ces règles surpassent celles fondées uniquement sur le nombre de patients, car elles tiennent compte à la fois de la gravité et du temps déjà subi par chacun. Elles restent néanmoins perfectibles : d'autres facteurs — équité, âge, sexe, risques de complications — doivent aussi guider le routage. Une analyse empirique, appuyée sur le *machine learning*, peut enrichir les politiques issues de la programmation dynamique et mieux capter la diversité des situations cliniques.

À terme, l'objectif est de fournir un outil d'aide à la décision qui, même s'il n'est pas optimal au sens strict, améliore la gestion des flux, réduit l'attente individuelle et respecte la spécificité de chaque patient. Cette coopération homme-machine ne pourra toutefois réussir que si les soignants se voient confier de réelles responsabilités et sont valorisés pour la qualité de leurs décisions.

¹³ Puterman, M. L. (2014). *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

¹⁴ Legros, B., Jouini, O., & Koole, G. (2018). A uniformization approach for the dynamic control of queueing systems with abandonments. *Operations Research*, 66(1), 200-209.

CHAPITRE IV

FORMATION DU PERSONNEL ET SON RÔLE DANS L'OPTIMISATION DES FLUX

FORMATION DU PERSONNEL SOIGNANT ET NON SOIGNANT ET SON ROLE DANS L'OPTIMISATION DES FLUX

IDENTIFICATION DES ROLES CLES DU PERSONNEL SOIGNANT ET NON SOIGNANT

Selon un rapport de la Cour des comptes, en 2023, la part du personnel non soignant dans les hôpitaux publics était de 29%¹⁵. En d'autres termes, c'est presque un tiers du personnel qui ne voit jamais un patient et dont le rôle est de diriger les tâches non médicales en lien avec l'activité de l'hôpital. Ces fonctions non soignantes recouvrent des métiers et des secteurs d'activité très différents essentiels à l'appui aux fonctions soignantes. Une partie du personnel gère les tâches administratives en lien avec l'orientation du parcours du patient (accueil, admission, secrétariat médical), et une autre partie s'occupe de la gestion de l'hôpital (gestion de la paie, comptabilité, etc.). Puis, une autre partie du personnel administratif gère la logistique, qui comprend les prestations hôtelières et les approvisionnements pour la prise en charge des patients. Enfin, une autre partie est composée de techniciens et d'ouvriers qui s'occupent de l'entretien des locaux. Or le rapport de la Cour des comptes établit que la coordination des fonctions soignantes et non soignantes est une condition de la fluidité du parcours du patient. Une piste d'amélioration serait d'augmenter la polyvalence et l'efficacité du personnel. Spécifiquement, il faudrait diversifier les profils des non soignants et former les soignants aux politiques de gestion des flux. Cela permettrait de réduire les erreurs et les goulots d'étranglement aux urgences. Un rapport du TRISAN souligne le fait qu'en Allemagne pour désengorger les urgences il a été mis en place un système de cabinets de garde (*Bereitschaftsdienstzentrale – BDZ; Notfallpraxen*) où exercent des médecins libéraux. Ces cabinets se situent dans les hôpitaux même ou à proximité de l'hôpital afin de pouvoir y envoyer les patients venus aux urgences pour une demande de soins non programmée et non urgente¹⁶. Au Royaume-Uni les patients non urgents sont encouragés à utiliser d'autres services d'urgences comme les "*NHS Walk-in Centres*", ce sont des centres de traitement d'urgence qui traitent des problèmes médicaux moins graves et sans rendez-vous¹⁷.

Ce n'est pas une nouvelle inédite mais l'hôpital public connaît une crise de l'emploi qui s'explique en partie par une mauvaise rémunération et des conditions de travail réputées difficiles. Selon France Travail, 99 % des établissements déclarent rencontrer des difficultés de recrutement¹⁸. Paradoxalement, la solution au problème de surmenage à l'hôpital ne réside pas toujours dans l'augmentation des effectifs. Une solution pour améliorer l'efficacité à ressource constante est de rendre plus performante la prise de rendez-vous à travers une optimisation des horaires et d'anticiper les aléas des médecins ainsi que du personnel médical (*no show* et retards). Aujourd'hui on pose le constat d'un système de prise de rendez-vous inefficace du patient car rien n'est fait pour anticiper ces aléas¹⁹. La recherche démontre que les clients ne sont pas nécessairement ponctuels et peuvent arriver plus tôt ou plus tard que l'heure de leur rendez-vous, voire ne pas se présenter du tout. Ce qui

¹⁵ Cour des comptes. (2025). Le personnel non soignant à l'hôpital public : repenser les fonctions support. <https://www.ccomptes.fr/sites/default/files/2025-05/20250526-RALFSS-2025-personnel-non-soignant-hopital-public.pdf>

¹⁶ TRISAN. (2025). Les soins médicaux urgents en Allemagne. https://www.trisan.org/fileadmin/user_upload/Rettungsdienste-VF_final.pdf

¹⁷ Schofield, A. (2025). Urgent treatment centres and walk-in centres. NHS Cheshire And Merseyside. <https://www.cheshireandmerseyside.nhs.uk/your-health/helping-you-stay-well/urgent-treatment-centres-and-walk-in-centres/>

¹⁸ « Les établissements publics de santé vont devoir engager des recrutements importants et améliorer les conditions de travail » - France Travail. (2023). France Travail. <https://www.francetravail.org/accueil/actualites/2023/les-etablissements-publics-de-sante-vont-devoir-engager-des-recrutements-importants-et-ameliorer-les-conditions-de-travail.html?type=article>

¹⁹ Legros, B. (2021). Agents' self-routing for blended operations to balance inbound and outbound services. *Production and Operations Management*, 30(10), 3599-3614.

fait que les heures d'arrivée par rapport aux rendez-vous programmés sont aléatoires. Or, organiser le routage des patients sur le temps réel d'attente et non sur le nombre d'attente serait une solution alternative²⁰.

Toutefois, pour éviter un taux de départ trop important qui réduit l'efficacité de l'hôpital, une autre approche serait de prendre en compte les risques du personnel soignant et non soignant de faire des heures supplémentaires en optimisant les plannings et les affectations. Considérer cette contrainte dans les prises de décisions permettrait de réduire les coûts liés au stress au travail et au *burnout*.

PROGRAMMES DE FORMATIONS CIBLEES

En termes de programmes de formations d'amélioration de la gestion des flux de patients pour le personnel non soignant, plusieurs pistes peuvent être explorées.

Une première piste serait de développer, à travers un système de formation en continue, leurs compétences médicales d'urgence pour que ces derniers puissent épauler leurs collègues du personnel soignant lors de périodes de congestion. En effet, cela permettrait au personnel soignant et non-soignant de disposer de compétences communes et donc d'améliorer la capacité de coopération en période de tension. Une seconde piste serait d'améliorer leur gestion du stress pour garantir une prise de décision rationnelle pour éviter les *burnout*. Concrètement, on pourrait imaginer des ateliers de simulation de scénarios qui amélioreraient leur résistance au stress, mais aussi leur réactivité lors de flux importants de patients. Enfin, dans une société où la communication et l'informatique sont au centre de tout, il est impératif de former le personnel non soignant aux méthodes de communication et à l'utilisation optimale des systèmes informatiques. En ce sens, les centres d'appels peuvent jouer un rôle clé, et c'est ce que nous allons explorer ci-dessous.

²⁰ Jouini, O., Benjaafar, S., Lu, B., Li, S., & Legros, B. (2022). Appointment-driven queueing systems with non-punctual customers. *Queueing Systems*, 101(1), 1-56.

CHAPITRE V

LA PLACE DES CENTRES D'APPELS DANS LA GESTION DES URGENCES

LA PLACE DES CENTRES D'APPELS DANS LA GESTION DES URGENCES

ROLE ACTUEL DES CENTRES D'APPELS

Les trois principaux centres d'appels d'urgence en France sont le 15, qui est le service d'aide médicale urgente (SAMU), le 18, qui est celui des pompiers, et le 112, qui est le numéro de téléphone réservé aux appels d'urgence dans l'ensemble de l'Union européenne. Leur rôle est d'agir comme une première ligne de contact avec le patient en menant une évaluation primaire de la situation et en apportant une assistance immédiate aux personnes en détresse. En outre, leur fonction est également de désengorger les services d'urgences physiques en orientant le patient vers les services de soins adaptés.

Ainsi, revoir le rôle des centres d'appels d'urgence pourrait incarner une solution dans l'optimisation de la gestion des flux de patients à l'hôpital. Cela pourrait par exemple se traduire par une optimisation de la capacité des centres d'appels pour réduire les temps d'attente téléphoniques et améliorer l'orientation en privilégiant une approche individualisée plutôt que standardisée. En ce sens, une réforme à privilégier serait de faire des politiques de réservation de ressources pour garder de la disponibilité pour les futures demandes en transférant toutes les demandes identifiées non urgentes vers d'autres canaux ou centres.

L'impact de la télémedecine et des consultations à distance sur le flux aux urgences n'est pas à négliger et pourrait être une piste à approfondir. Depuis la loi du 21 juillet 2009 portant sur la réforme de l'hôpital et relative aux patients, à la santé et aux territoires, la télémedecine est définie et réglementée en France. Par la télémedecine, le patient peut mener une téléconsultation (une consultation à distance entre un médecin et un patient), une téléexpertise (un échange entre au moins deux médecins qui arrêtent ensemble, avec le consentement du patient, un diagnostic ou une stratégie thérapeutique), la télésurveillance médicale (un médecin interprète à distance les données cliniques ou biologiques recueillies), la téléassistance (un médecin assiste à distance l'un de ses confrères pendant un acte médical ou chirurgical) et la régulation (la réponse médicale apportée par les centres d'appel d'urgence). D'après la Direction de la recherche, des études, de l'évaluation et des statistiques (Drees) le nombre de téléconsultations réalisées par des médecins généralistes libéraux est évalué à 9,4 millions en 2021²¹. Une généralisation de cette pratique permettrait de faire des économies grâce à une restructuration des soins, une mise en commun des compétences médicales et surtout elle permettrait de désengorger les hôpitaux publics.

A titre de comparaison la Suède a été le pays européen précurseur dans l'adoption de la télémedecine. Au royaume scandinave la télémedecine est remboursée depuis 2016²². Un rapport de l'agence nationale de santé digitale suédoise stipule que cette pratique a été facilitée par une mise en place généralisée des outils numériques dans le secteur de la santé. Notamment par la démocratisation de l'identité numérique (*e-legitimation*), du carnet de santé en ligne mais aussi avec la mise en place de bornes d'enregistrement et de paiement pour désengorger les urgences²³. Dans le *policy paper* « Faire de la santé une priorité européenne » l'Institut Jacques Delors souligne le fait que la numérisation du domaine de la santé est une priorité au niveau de Bruxelles. Un Espace européen des données de santé (*EHDS-European Health Data Space*) est en ce moment même en train

²¹ Vie Publique. (2025). La télémedecine, une solution pour faciliter l'accès aux soins ? <https://www.vie-publique.fr/eclairage/18473-la-telemedecine-une-solution-pour-faciliter-lacces-aux-soins>

²² De Foucaud, I. (2019, 10 septembre). Comment la télémedecine s'est installée dans le paysage médical en Suède. *Challenges*. https://www.challenges.fr/entreprise/sante-et-pharmacie/comment-la-telemedecine-s-est-installee-dans-le-paysage-medical-en-suede_673765

²³ E-hälsomyndigheten. (2021). Uppföljning Vision e-hälsa 2025. https://www.ehalsomyndigheten.se/globalassets/ehm/3_om-oss/rapporter/uppfoljning-vision-e-halsa-2025-rapport-avseende-2021.pdf

d'être mis en place. L'idée étant que les patients bénéficieront d'un accès simple et immédiat à leurs données de santé électroniques dans les pays de l'UE²⁴.

AMELIORATION DE LA COORDINATION ET DU PARTAGE D'INFORMATIONS

Comme nous venons de l'aborder plus haut, le besoin de partager les données de santé fait partie de l'ère du temps dans la mesure où nous vivons dans une société mondialisée où les flux transfrontaliers de personnes sont devenus une réalité commune. Cette plus grande mobilité des populations implique donc une plus importante coopération entre les acteurs de la santé au niveau national mais surtout au niveau européen (spécifiquement, une plus grande coopération entre les acteurs préhospitaliers et hospitaliers, mais également entre acteurs préhospitaliers).

Dans le *policy paper* « *Pour une Europe de la santé* », l'Institut Jacques Delors pose le constat d'un échange de données de santé à des fins de soins et de recherche trop limité alors que le numérique est un formidable outil pour améliorer le parcours de soins des patients²⁵. Cette réalité implique un besoin d'amélioration de la coordination et du partage d'information entre les services de santé. En ce sens, il serait pertinent de faciliter les protocoles de transfert d'informations entre le centre d'appel et le service d'urgence. La mise en place de systèmes d'information intégrés communs aux services pour stocker, traiter et assurer la diffusion des données représenterait un précieux gain de temps, ce qui éviterait d'engorger les urgences. Une autre technologie qui pourrait contribuer à fluidifier l'afflux aux urgences est l'IA, qui est un domaine de recherche en pleine expansion et promis à un grand avenir dans la santé. Comme le rapporte le site officiel du Gouvernement, l'IA est déjà employée pour enrichir les soins en santé mentale à travers des dispositifs comme les objets connectés, les chatbots et les agents conversationnels²⁶. A l'instar de ce qui est fait dans le domaine de la santé mentale, on pourrait utiliser une approche multicanale pour éviter l'engorgement des urgences. Cela en généralisant l'utilisation de chats instantanés²⁷. En effet, la mise en place de chat instantané avec une IA ou un agent, permettrait au potentiel patient d'échanger directement en ligne et donc d'éviter de se déplacer jusqu'à l'hôpital pour une urgence qui n'en est peut-être pas une.

²⁴ *Faire de la santé une priorité européenne* - Institut Jacques Delors. (2021). Institut Jacques Delors. <https://institutdelors.eu/publications/faire-de-la-sante-une-priorite-europeenne/>

²⁵ *Pour une Europe de la santé* - Institut Jacques Delors. (2021). Institut Jacques Delors. <https://institutdelors.eu/publications/pour-une-europe-de-la-sante/>

²⁶ Info.gouv.fr.(2025). L'intelligence artificielle au service de la santé mentale. <https://www.info.gouv.fr/actualite/lintelligence-artificielle-au-service-de-la-sante-mentale#:~:text=L'IA%20et%20les%20outils%20en%20sant%C3%A9%20mentale,-L'IA%20est&text=Ces%20outils%20recueillent%20des%20donn%C3%A9es,%C3%A9tat%20clinique%20de%20leurs%20patients>

²⁷ Legros, B., & Jouini, O. (2019). On the scheduling of operations in a chat contact center. *European Journal of Operational Research*, 274(1), 303-316.

CHAPITRE VI

DEFIS ET PERSPECTIVES

DEFIS ET PERSPECTIVES

APPLICABILITE DES MODELES MATHEMATIQUES

Nicolas Copernic affirmait « *les mathématiques ne sont écrites que pour les mathématiciens*²⁸ ». Or, c'est justement une des faiblesses de cette discipline, car à un certain niveau, pour la majorité de la population, c'est une science difficilement compréhensible.

Pour être performant et changer les décisions, les solutions issues du modèle mathématique doivent être expliquées et comprises en dehors du cadre posé par le modèle stylisé étudié. Même si l'incompréhension des mathématiques limite leur applicabilité, elles ne doivent pas rester comme des solutions magiques ou des surprises qui marchent sans savoir pourquoi.

De plus il faut garder à l'esprit que les mathématiques ne sont possibles que dans un frame simplificateur de la réalité. En d'autres mots, les modèles ne sont pas strictement conformes à la réalité et ignorent de nombreux aspects. Ensuite, les modèles mathématiques nécessitent des données fiables et complètes, ce qui nécessite un système de capture des données performant. Il suffit qu'une valeur soit erronée lors d'une première prédiction et l'estimation erronée des paramètres peut conduire à des décisions inadaptées. Finalement, la faiblesse des modèles mathématiques réside dans la difficulté à capturer certains aspects de la complexité humaine et les événements rares.

INTEGRATION DES FACTEURS HUMAINS ET ORGANISATIONNELS

Pour que les modèles mathématiques reflètent au mieux la réalité, il est nécessaire d'y intégrer les facteurs humains et organisationnels et dans notre cas il s'agit de la conception du modèle visant à réduire la congestion aux urgences. Dans un premier temps, il est nécessaire d'intégrer les externalités négatives induites par le comportement du personnel hospitalier contribuant à une plus grande congestion. Dans le cas de l'hôpital, il s'agit d'intégrer la culture d'entreprise qui a un impact sur le comportement du personnel et sur la dynamique de la structure. Une autre nécessité est d'intégrer les potentielles résistances au changement des patients comme du personnel qui pourrait être hostile à toute idée d'évolution de leurs méthodes de travail. En outre, la motivation du personnel soignant est à prendre en compte et cela tout particulièrement en lien avec la charge de travail. Une étude publiée par la direction de la recherche, des études, de l'évaluation et des statistiques souligne une dégradation des conditions de travail à l'hôpital entre 2003 et 2019. Elle y reste plus élevée que dans les autres secteurs d'activité avec des horaires atypiques, un rythme intense ainsi que des exigences émotionnelles et physiques²⁹. Ces conditions de travail difficiles induisent un risque de mal-être du personnel soignant et donc c'est toute l'attractivité du métier qui est remise en cause. Une autre étude de la Direction de la recherche, de l'évaluation, des études et des statistiques publiées en 2023 rapporte une prévalence accrue de la dépression et de l'anxiété due aux conditions de travail. Selon l'étude à l'hôpital, en 2021, 41 % des personnes ont des symptômes de dépression légère à sévère, contre 33 % dans l'ensemble des personnes en emploi³⁰.

Dans un second temps, il est pertinent d'être conscient des externalités négatives qui découlent du fait qu'un patient rejoint un système qui, dans notre cas, sont les urgences. Pour sensibiliser sur ce sujet, deux solutions

²⁸ Nicolas Copernic-Les mathématiques ne sont écrites que pour les. (2025). Evéne.fr. <http://eveve.lefigaro.fr/citation/mathematiques-ecrites-mathematiciens-69259.php>

²⁹ Vie Publique. (2021). Hôpital : des conditions de travail de plus en plus difficiles. <https://www.vie-publique.fr/en-bref/282444-hopital-des-conditions-de-travail-de-plus-en-plus-difficiles>

³⁰ À l'hôpital, une prévalence accrue de la dépression et de l'anxiété due notamment aux conditions de travail / Direction de la recherche, des études, de l'évaluation et des statistiques. (2023). <https://drees.solidarites-sante.gouv.fr/publications-communique-de-presse/etudes-et-resultats/lhopital-une-prevalence-accrue-de-la>

existent, les incitations contraignantes et non-contraignantes³¹. La première stratégie fait référence au contrôle des admissions ou à la mise en place d'une stratégie de tarification (comme mentionné plus haut) pour réduire la congestion. L'impasse de cette stratégie dans le cadre des urgences est qu'il s'agit d'un service public. Par conséquent, cette approche pourrait être difficilement justifiable d'un point de vue éthique. La seconde stratégie, l'alternative non-contraignante, vise également à faire comprendre au patient l'externalité négative engendrée par une décision de rejoindre la file d'attente, qui dans l'hypothèse d'une congestion engendre une surexploitation des ressources hospitalières. Une approche alternative serait de sensibiliser le patient lors du temps d'attente par des campagnes d'affichage à l'entrée des urgences, en reprenant le modèle de ce qui est fait lors des campagnes de sécurité routière sur le bord des routes. On pourrait aller plus loin et sensibiliser au-delà du lieu même de l'hôpital par le biais de campagnes médias, que ce soit sur internet, sur les réseaux sociaux ou encore à la télévision. Une autre solution serait également de mettre en place un dispositif de réservation en ligne où le patient est sensibilisé sur le temps d'attente additionnel généré par son arrivée. L'avantage de cette stratégie non-contraignante réside dans le fait qu'il ne s'agit pas d'une incitation économique ou d'une stratégie contraignante influant sur la liberté de choix. Aux Etats-Unis, certains groupes de santé permettent aux patients de prendre rendez-vous en ligne pour les urgences non vitales. Permettant ainsi aux patients d'attendre chez eux plutôt que dans une salle d'attente, ce qui contribue à réduire les files d'attente³². Une stratégie similaire a été mise en place au Québec. Dans l'objectif de dissuader les Québécois de se rendre aux urgences, qui souffrent de problèmes de débordement, le ministère de la Santé présente un site Internet où les temps d'attente des hôpitaux sont mis à jour³³.

LES PERSPECTIVES FUTURES

Ci-dessous nous avons synthétisé les potentielles pistes d'actions à mettre en place pour fluidifier le service des urgences en France :

1. Par l'utilisation des nouvelles technologies du numérique :

- **Favoriser la coopération homme-machine** en utilisant l'IA pour un pré-triage automatique qui oriente les patients en amont, lisse le flux et améliore la précision des estimations et réduit ainsi l'impact de l'aversion aux risques.
- **Etendre la pratique de la télémedecine et des solutions numériques dans le domaine de la santé** par la mise en place de bornes d'enregistrement et de paiement.
- **Utiliser des jumeaux numériques pour tester les changements d'organisation** et de stratégie de manière dynamique et voir l'impact hypothétique sur les performances de l'hôpital.

2. Par la formation du personnel hospitalier :

- **Augmenter la polyvalence et l'efficacité du personnel soignant et non-soignant en formant ces derniers** dans leur capacité d'optimiser le temps de séjour et la fluidité des patients.
- **Développer un système de formation en continue sur les compétences médicales d'urgence** afin d'épauler le personnel soignant en période de congestion. Tous les personnels auraient des compétences communes qui faciliteraient la coopération.
- **Améliorer la gestion du stress par des formations avec des simulations de scénario.** Le développement de plateformes de simulation interactives pour la formation et la planification pourrait être privilégié.

3. Par la réorganisation :

³¹ Legros, Benjamin and van Leeuwen, Johan, Wait Externality Awareness for Socially Responsible Queueing (March 17, 2025). Available at SSRN: <https://ssrn.com/abstract=5182011> or <http://dx.doi.org/10.2139/ssrn.5182011>

³² Aux Etats-Unis, les patients peuvent prendre rendez-vous pour aller aux urgences. (2022, 13 décembre). What's Up Doc. <https://www.whatsupdoc-lemag.fr/article/aux-etats-unis-les-patients-peuvent-prendre-rendez-vous-pour-aller-aux-urgences>

³³ Info, R. (2023, 31 janvier). Urgences : un nouveau site répertorie les temps d'attente au Québec. *Radio-Canada*. <https://ici.radio-canada.ca/nouvelle/1952273/sante-dube-hopital-soins-urgence-temps-attente-site>

- **Mettre en place un système de filtrage des patients à l'entrée des urgences.** Néanmoins, nous soulignons que la limite de cette approche est de discerner la différence entre état réel et état perçu du patient.
 - **Favoriser la construction d'hôpitaux de proximité avec des soins basiques en supplément d'hôpitaux de recours disposant de plateaux techniques et de prises en charge.**
 - **Prendre en compte dans le processus décisionnel le risque de départs liés au stress et au *burnout*.** Pour cela optimiser les plannings et les affectations pour éviter aux soignants les heures supplémentaires non panifiées.
 - **Faciliter les protocoles de transfert d'informations entre le centre d'appel et le service d'urgence** par un système d'information intégré.
- 4. Par les incitations contraignantes et non contraignantes :**
- **Instaurer des incitations contraignantes comme une tarification symbolique** à l'admission pour décourager les passages non essentiels.
 - Instaurer des incitations non-contraignantes pour faire comprendre au patient l'externalité négative engendrée par la décision de rejoindre une file d'attente qui peut contribuer à la surexploitation des ressources hospitalières. Une première solution serait de **sensibiliser le patient par des campagnes d'affichage à l'entrée des urgences**. Une seconde de **sensibiliser par des campagnes dans les médias**. Une troisième de **mettre en place un dispositif de réservation en ligne** où le patient est sensibilisé sur le temps d'attente additionnel généré par son arrivée.

CONCLUSION

En conclusion, ce *policy paper* s'est efforcé d'analyser la manière dont les approches mathématiques peuvent éclairer et améliorer la gestion des files d'attente dans les services d'urgence, en intégrant les dimensions de risque, de priorisation, de personnel, du numérique, de réorganisation et de centres d'appels. Nous avons proposé une série de recommandations pour les gestionnaires hospitaliers et les décideurs politiques en soulignant l'importance de privilégier une approche interdisciplinaire combinant l'expertise médicale, opérationnelle et mathématique pour relever les défis futurs des services d'urgence.

AUTEURS

Benjamin Legros, est professeur d'Operations Management à SKEMA Business School (Centre for Analytics & Management Science, campus Grand Paris). Ses recherches portent sur l'optimisation des décisions en temps réel dans des systèmes stochastiques (files d'attente, MDP) avec applications aux centres de contacts, aux transports et à la santé. Il est titulaire d'un doctorat en génie industriel de CentraleSupélec (2013) et d'une HDR (2024). Ses travaux sont publiés notamment dans Production and Operations Management, Operations Research et Manufacturing & Service Operations Management.

Frédérique Vidal, Directrice du développement de SKEMA Publika et directrice de la stratégie et de l'impact scientifique de SKEMA Business School. Professeur des universités en biologie, Présidente de l'université de Nice Sophia Antipolis entre 2012 et 2017, puis ministre de l'Enseignement Supérieur, de la Recherche et de l'Innovation, dans les gouvernements Philippe et Castex de 2017 à 2022. Elle a été conseillère spéciale auprès du Président de l'EFMD et est aussi actuellement Représentante Permanente de la Principauté de Monaco auprès des Nations Unies Environnement et de la Commission Baleinière.

Sean Scull, chargé de projets think tank, est diplômé de sciences politiques avec une spécialisation en relations internationales de l'université de Göteborg et d'un master en politiques internationales avec une spécialisation en politique anglophone de l'université de Toulon. Sean a vécu et travaillé en Suède et aux États-Unis d'Amérique.

Nous remercions chaleureusement **Mathieu Raux** et **Thomas Botrel** pour les échanges riches que nous avons eus ces dernières années et pour leurs conseils avisés, qui ont grandement éclairé la rédaction de cet article.

Les biographies des auteurs sont disponibles sur publika.skema.edu

Date de publication : septembre 2025

Retrouvez l'ensemble de nos publications sur publika.skema.edu

Contact : publika@skema.edu