CAMION: a catchment area maximization algorithm, with application to oncology accessibility in metropolitan France.

Eric Daoud[1,2], Anne-Sophie Hamy[1,3], Elise Dumas[1,4,5], Lidia Delrieu[1], Beatriz Grandal Rejo[1,6], Christine Le Bihan-Benjamin[7], Sophie Houzard[7], Philippe-Jean Bousquet[8,9], Judicaël Hotton[10], Aude-Marie Savoye[11], Christelle Jouannaud[11], Chloé-Agathe Azencott[4,5,12], Marc Lelarge[2], Fabien Reyal[1,6,10,*]

1. Residual Tumor & Response to Treatment Laboratory, RT2Lab, INSERM, U932 Immunity and Cancer, Institut Curie, Université Paris, 75005 Paris, France
2. INRIA, DI/ENS, PSL Research University, Paris, France
3. Department of Medical Oncology, Institut Curie, Paris, France
4. INSERM, U900, 75005 Paris, France
5. MINES ParisTech, PSL Research University, CBIO-Centre for Computational Biology, 75006 Paris, France
6. Department of Surgery, Institut Curie, Paris, France
7. Health Data and Assessment Department, Survey Data Science and Assessment Division, National Cancer Institute, 52 avenue André Morizet 92100 Boulogne-Billancourt, France
8. Aix Marseille Univ, Inserm, IRD, SESSTIM, Equipe Labellisée Ligue Contre le Cancer, Marseille, France
9. Survey Data Science and Assessment Division, National Cancer Institute, 52 avenue André Morizet 92100 Boulogne-Billancourt, France
10. Department of Surgery, Institut Jean Godinot, Reims, France
11. Department of Medical Oncology, Institut Jean Godinot, Reims, France
12. Institut Curie, PSL Research University, 75005 Paris, France

Corresponding Author: Fabien Reyal, Institut Curie, 26 rue d'Ulm, Paris 75005, France. Phone: +33(0)142346339; E-mail: fabien.reyal@curie.fr.

Abstract

Background: Access to health services plays a key role in cancer survival. Uneven distributions of populations and health facilities lead to geographical disparities. Location-allocation algorithms can address these disparities by finding new locations and capacities for health facilities. However, in oncology, opening new hospitals or moving them is difficult in practice, and should be handled carefully.

Methods: We propose a method to measure the spatial accessibility to oncology care and identify the hospitals to grow to reduce disparities. We first ran a clustering algorithm to automatically label the hospitals in terms of oncology specialization. Then, we computed an accessibility score to these hospitals for every population location. Finally, we introduced *CAMION*, an optimization algorithm based on Linear Programming that reduces disparities in oncology accessibility by identifying health facilities that should increase their capacities.

Results: We demonstrate our algorithm in metropolitan France. The clustering step let us identify different oncology specialization levels for hospitals. Most of the population in metropolitan France lived in good accessibility areas, especially in large cities. Lower accessibility zones are often rural or suburban municipalities. The optimization algorithm effectively manages to identify hospitals to grow, based on current oncology specialization and accessibility scores.

Discussion: There is a tradeoff to be found by patients, between care center proximity and care center expertise, which is less likely to happen for patients living in good accessibility areas. The accessibility score is deliberately non-specific to cancer type but can be adapted to more precise pathologies. Our method is replicable in any country, given hospitals and population locations data. We developed a web application intended for healthcare professionals to let them to run the optimization algorithm with the desired parameters and visualize the results.

Introduction

Cancer is a leading cause of death worldwide, accounting for nearly 10 million deaths in 2020. While a lot of the ongoing research is focusing on finding new cancer treatments, accessibility to oncology care receives less attention. Yet, several studies have showed that access to health services plays a key role in cancer survival. For instance, geographic residency

status and social environment seem to explain treatment and prognosis disparities for patients

with non-small cell lung cancer (1). In France, increases in travel times to health services

were associated with lower survival rates for patients with a colorectal cancer (2). In New

Zealand, living in deprived areas, far from a cancer center or from primary care was associat-

ed with lower survival chances for patients with colorectal, lung and prostate cancers (3).

Accessibility refers to the relative ease by which services can be reached from a given lo-

cation (4). Accessibility can be defined by spatial factors, determined by where you are; and

non-spatial factors, determined by who you are (5). Spatial accessibility methods assess the

availability of supply locations from demand locations, connected by a travel impedance met-

ric. Supply locations are characterized by their capacity or quantity of available resource.

Similarly, demand locations are characterized by their population. Such methods have been

successfully used to measure access to healthcare, such as primary care (6) or oncology care

(4,7,8) in several countries including France (9–11). In what follows, we restrict accessibility

to spatial accessibility and use both terms interchangeably.

Uneven distributions of population and health-care providers lead to geographic disparity

in accessibility for patients (12). For instance, Weiss et al. (13) showed that 8.9% of the glob-

al population could not reach healthcare within one hour if they have access to motorized

transport. In Germany, Bauer et al. (14) shown that 10% of the population lived in areas with

low accessibility for internal medicine and surgery. Location-allocation algorithms (15) can

optimize the distribution and supply of health providers to reduce accessibility disparities.

These algorithms seek the optimal placement of facilities for a desirable objective under cer-

tain constraints (4). For instance, Luo et al. developed an optimization algorithm to improve

the healthcare planning in rural China by finding the best place and capacity for new health

facilities (16). Tao et al. worked on a spatial optimization model to maximize equity in acces-

sibility to residential care facility in Beijing, China (17). When optimizing health accessibil-

ity, there are two competing goals: equity and efficiency (18,19). Equity may be defined as

equal access to healthcare for everyone (20). An efficient situation is when everything has

been done to help any person without harming anyone else (21). While some argue that effi-

ciency should be addressed in priority (21), others agree that equity is a matter of ethical ob-

ligation, especially in public health (22,23).

The goal of this paper is to apply spatial accessibility methods to oncology care centers

and propose an optimization algorithm to reduce disparities. We demonstrate our results in

metropolitan France. There are many care centers in France, which do not share the same

degree of oncology specialization. Therefore, we first run a clustering algorithm to automati-

cally group the care centers based on their medical statistics and attributes. Using these clus-

ters, we label the care centers in terms of hospital development and oncology specialization.

Then, we compute an oncology accessibility score for every municipality in metropolitan

France. We then introduce *CAMION*, an optimization algorithm based on *Linear Program-*

ming which uses the clusters of care centers and the accessibility scores to suggest, given a

limited budget, where to increase hospital capacity to improve the oncology accessibility.

Finally, our method is packaged into a web application intended to healthcare professionals

so they can run the optimization algorithm with the desired parameters for any region.

Methods

Data collection

Health data is collected from two sources: the *French national administrative database*

(PMSI) and the *French annual health facilities statistics* (*SAE*). *PMSI* data includes discharge

summaries for all inpatients admitted to public and private hospitals in France. The *SAE* da-

tabase is a compulsory and exhaustive administrative survey of all public and private hospi-

tals in France. The survey is sent every year and describes the activities of the hospitals as

well as the list of services and their staff. We restricted the analysis to the year 2018. We in-

clude every hospital in metropolitan France that declared a *Medicine, Surgery or Obstetric*

(MCO) activity in the *SAE* survey, in 2018. We also included the liberal radiotherapy care

centers, with no *MCO* activity. The resulting dataset contains 1,662 care centers.

Geographic and travel data were retrieved from open data platforms. Municipalities and

their census statistics were extracted from the *National Statistics Bureau of France (INSEE)*

website. We used the *OpenRouteService (ORS) API* to compute the driving routes between

hospitals and municipalities, which is necessary for the accessibility score.

Care centers characterization

We selected a list of 24 variables with the help of medical experts to characterize the care

centers. The list of variables and their definitions is available in the supplementary materials.

The variables are either binary when they encode the presence or absence of a service; or

discrete when they encode the number of stays. We only focus on treatments received in hos-

pitals.

Given the large number of care centers, we use a clustering algorithm to automatically

group together similar care centers. More specifically, we first run a *Principal Component*

Analysis (PCA) algorithm on the *SAE* dataset that describes the care centers. The input data

has 24 variables, and we perform the dimensionality reduction with $n = 2$ components. We

tried different number of components, from 2 to 5, but we found 2 gave good and easy to

interpret results. We then run a clustering algorithm on the PCA-reduced dataset to automati-

cally isolate care centers with similar statistics. We tried several algorithms like *K-Means*

(24), DBSCAN (25) and Spectral Clustering (26). In our case, *Spectral Clustering* with 8

clusters gave the most interpretable and better isolated groups. For the number k of clusters,

we tested all values from 2 to 10 and manually interpreted the results with medical experts.

Accessibility score

There are several ways to compute accessibility to healthcare (6). The easiest and most straightforward methods are computed within bordered areas, like provider-to-population ratios in each municipality. While they are very intuitive, these methods do not account for border crossing, or travel impedance, which makes them less accurate. Recently, a new type of method has been developed and is now used in most spatial accessibility papers. This algorithm is called *Two Step Floating Catchment Area (2SFCA)* (27). It is a two-step method that first computes a provider-to-population ratio for each provider location. In the second step, for each population location, an accessibility score is obtained by summing the provider-to-population ratios. For the algorithm to work, a catchment threshold (distance or travel time) must be set. Above this threshold, a provider location is considered unreachable from the population location, and vice versa. The *2SFCA* method does not account for distance decay: a care center is either reachable or not. The *Enhanced Two Step Floating Catchment Area (e2SFCA)* (28) addresses this limitation by applying weights to differentiate travel zones in both steps.

We now explain more formally how to compute *eS2FCA* scores. Consider P_i the population at location i, with $1 \leq i \leq n$ where n is the number of population locations. Similarly, consider S_u the capacity of care center u, with $1 \leq u \leq m$ where m is the number of care centers. Finally, let d_{iu} be the matrix of size $n \times m$ containing the distances between location i and care center u. We consider r sub-catchment zones each associated with a weight W_s, and a distance D_s, with $1 \leq s \leq r$, such that $D_1 < D_2 < \cdots < D_r$ and $W_1 > W_2 > \cdots > W_r$. The resulting r travel intervals are $I_1 = [0, D_1]$, $I_2 = [D_1, D_2], \ldots, I_r = [D_{r-1}-, D_r]$. The accessibility A_i of a population location i is computed in two steps. Step 1: for every care center u, compute its weighted capacity-to-population ratio R_u. Step 2: for every population location, compute A_i as the sum all the weighted R_u of the reachable care centers.

$$R_u = \frac{S_u}{\sum_{s=1}^{r} W_s \sum_{i, d_{iu} \in I_s} P_i}$$

$$A_i = \sum_{s=1}^{r} W_s \sum_{u, d_{iu} \in I_s} R_u$$

The capacity of a care center is balanced by the total population with access to it. A popu-

lation location that solely has access to low capacities or overcrowded care centers will have

a low accessibility score. Similarly, a population location will have low accessibility scores if

the distance to get to the nearby care centers is large.

As we want to compute the accessibility to oncology care centers, we chose S_u to be the

oncology activity of a hospital u. We define oncology activity as the sum of the number of

medical and surgery stays related to cancer, and the number of patients with chemotherapy or

radiotherapy. A care center with no oncology activity will have $R_u = 0$ and a municipality

that solely has access to this care center u will have $A_i = 0$. We use driving duration as travel

impedance metric, and we set the maximum catchment area to a 90-minute drive. In 2018,

only 24,152 patients out of 761,057 (3.2%) had travel duration greater than 90 minutes for

cancer related pathways. This is low enough to consider that care centers are non-reachable

beyond this distance. We divide the catchment area into 3 intervals: $I_1 = (0, 30]$, $I_2 =$

$(30, 60]$ and $I_3 = (60, 90]$. The associated weights are respectively $W_1 = 1, W_2 = 0.042$ and

$W_3 = 0.09$. These sub catchment areas are set based on the cancer pathways travel duration

distributions and validated with medical experts. The weights are the same than the *e2SFCA*

paper (28).

For privacy reasons, municipalities with small populations are grouped in entities called

"geographic codes" in the *PMSI* data. We decided to compute the accessibility score for each

geographic code and municipalities that are grouped in the same code will have the same

accessibility score.

Accessibility optimization

Regarding efficiency optimization, the most popular algorithms are *p-median*, location set covering problem *(LSCP)* and *maximum covering location problem (MCLP)*. The *p-median* algorithm minimizes the weighted sum of distances between users and facilities (29). *LSCP* minimizes the number of facilities needed to cover all demand (30). *MCLP* maximizes the demand covered within a desired distance or time threshold by locating a given number of facilities (31).

To reach equal access to healthcare, quadratic programming has been used to minimize the variance of accessibility scores defined by the *2SFCA* (32). Similarly, a *Particle Swarm Optimization (PSO)* algorithm was developed to minimize the total square difference between the accessibility score of each demand location and the weighted average accessibility score (17). Finally, a two-step optimization algorithm has been developed to address the dual objectives of efficiency and equality, by first choosing where to site new hospitals and then deciding which capacity they should have (16,33).

However, most of the previous algorithms seek locations to open new health facilities. In this work, we are interested in the case where the health facilities are fixed, and the only lever to improve accessibility is to increase their capacities. Given a capacity budget, we want to know which facilities to grow and by how much. We introduce *CAMION*, an accessibility optimization algorithm based on *Floating Catchment Area* and *Linear Programming*. The initial accessibility score was computed with the *Enhanced Two Step Floating Catchment Area (e2SFCA)* (28) but our algorithm can generalize to more *FCA* derivatives.

We model the problem as an optimization task. In our case, we want our optimization algorithm to find new care centers capacities given some constraints, so that the total accessibility is maximum. We apply optimization on a given region only, rather than on the whole metropolitan France. We chose this approach because healthcare planning is handled region-

ally rather than nationally. We show below that our optimization problem is a Linear Pro-

gramming problem.

In its standard form, *Linear Programming* finds a vector x that maximizes $c^T x$ under con-

straints $Ax \leq b$, where A is a matrix and b a vector. Boundaries can be set to x such as $x \geq$

0. Consider x_u the new capacity of a care center u, to be computed by the algorithm. Let Q_u

and W_u be two vectors of size m, defined as follows:

$$Q_u = \sum_{s=1}^{r} W_s \sum_{i,d_{iu} \in I_s} P_i$$

$$W_u = \sum_{s=1}^{r} \sum_{i,d_{iu} \in I_s} W_s$$

205   We can compute the total accessibility as a sum on the m care centers:

$$\sum_i A_i = \sum_i \sum_{s=1}^{r} W_s \sum_{u,d_{iu} \in I_s} \frac{S_u}{Q_u}$$

$$\sum_i A_i = \sum_s \sum_{i,u,d_{iu} \in I_s} W_s \frac{S_u}{Q_u}$$

$$\sum_i A_i = \sum_u \frac{S_u}{Q_u} \sum_s \sum_{i,d_{iu}} W_s$$

$$\sum_i A_i = \sum_u \frac{S_u}{Q_u} W_u$$

206   The last equation can be rewritten in the *Linear Programming* standard form with:

$$c = \frac{W_u}{Q_u}$$

$$x_u = S_u$$

$$b \geq \sum_u x_u$$

$$x_{u_{min}} \leq x_u \leq x_{u_{max}}$$

The user-defined parameters are b, $x_{u_{min}}$ and $x_{u_{max}}$. b is the total capacity to be shared

across all the care centers. $x_{u_{min}}$ and $x_{u_{max}}$ are the capacity boundaries for care center u. If b

is set to the current total capacity, a care center can't be grown unless another one is de-

creased. If $b > \sum_u x_u$, the capacity of care centers can be increased without decreasing other

centers. We know how to solve *Linear Programming* and we used the *SciPy* (34) implemen-

tation of the revised simplex method as explained in (35).

We now detail how we set the user-defined parameters to apply the *Linear Programming*

algorithm to our specific case. The additional capacity was set as +3% of the overall activity

of the region's care centers: $b = 1.03 \times \sum_u x_u$. The choice of the boundaries $x_{u_{min}}$ and $x_{u_{max}}$

is crucial and must be realistic. We studied the hospitals activity on the past four years (2016

to 2019) to retrieve the average growth percentage of a care center. The growth percentage is

computed as follows: $(S_{2019} - S_{2016})/S_{2016}$. Among the care centers that grew and who had an

existing oncology activity, the mean growth percentage was 23%. Hence, we set $x_{u_{max}}$ as

+20% of the care center capacity. Regarding $x_{u_{min}}$, we set the boundary based on the cluster

of the care center. For the three most specialized clusters, we set their $x_{u_{min}}$ equal to their

current activity. We did this to prevent the algorithm from decreasing the most specialized

and well-equipped care centers. Regarding the care centers from the other clusters, $x_{u_{min}} =$

0, so that they could be emptied if need be. Finally, we set $x_{u_{max}} = 0$ if the care center be-

longs to the least specialized cluster. The new capacities are indicative and should be further

investigated to make sure they are relevant. Especially when setting an existing oncology

activity to 0.

We developed a web application that allows the users to run the optimization algorithm in

any region with the parameters they want. The application displays accessibility results and

optimization outcomes on an interactive map with additional plots. The user can browse the

list of care centers by cluster and the list of municipalities with their accessibility scores.

Results

Population and hospitals distribution in metropolitan France

In 2018, the population in France was 66,993 million. Mainland France hosts 64,812 million inhabitants (96.8%), while the remaining 2,181 million (3.2%) live in overseas departments and regions. Metropolitan France is divided into 13 administrative regions and 96 departments. The population density in France is unevenly distributed. In 2020, the overall population density in metropolitan France was 119 inhabitants per square kilometer. *Ile-de-France* region has the highest population density with 1,022 inhabitants per square kilometer. Density in other regions in metropolitan France range between 40 and 187 inhabitants/km^2. Denser areas are located near the coastline and around the largest cities like *Paris, Marseille, Lyon, Strasbourg, Toulouse, or Bordeaux*. The middle of the country is rural, and the population densities are low. While there are a great variety of regions and landscapes, the country is becoming more urbanized. This "*rural exodus*" is largely responsible of what is known as the "*empty diagonal*", a band of very low-density population that stretches from the southwest to the northeast.

We now describe the spatial distribution and specificities of the 1,662 hospitals included in this study. There are different types of hospitals in France: *Centres Hospitaliers* (CH, n=667) and *Centres Hospitaliers Régionaux / Universitaires* (CHR/U, n=142) are state-run hospitals; *Centres de Lutte Contre le Cancer* (CLCC, n=26) and *Participants au Service Public Hospitalier / Etablissement à But Non Lucratif* (PSPH/EBNL, n=142) are both private hospitals of collective interest, though *CLCC* are oncology dedicated; private hospitals (n=606) are privately run and for-profit. The non-*MCO* care centers with radiotherapy activity (n=79) are mostly private practice structures and are referred as *Other*. **Table 1** shows the number of care centers and their oncology activity per hospital type and region. Most of the care centers are public, but a non-neglectable part are private. *CLCC* represent only 1.6% of

the care centers, yet they are responsible for 14.2% of the overall oncology activity. The care

centers are unevenly distributed across the country. For instance, *Corse and Centre-Val-de-*

Loire are the only two regions with no *CLCC* care centers. Moreover, the proportion of on-

cology activity per hospital type varies from a region to another. For instance, in *Nouvelle-*

Aquitaine, 47.1% of the oncology activity is handled by private care centers, whereas in *Pro-*

vence-Alpes-Cote-d'Azur it is 21.4%.

Variable value per region	Hospital type						
N = number of centers	CH	CHR/U	CLCC	Other	PSPH/EBNL	Private	Overall
A = oncology activity (radio., chemo., surgery)	n=667	n=142	n=26	n=79	n=142	n=606	n=1,662
Auvergne-Rhône-Alpes	98 (49,2%)	21 (10,6%)	2 (1%)	7 (3,5%)	13 (6,5%)	58 (29,1%)	199
	34,597 (26,7%)	31,706 (24,5%)	16,966 (13,1%)	6,710 (5,2%)	6,146 (4,7%)	33,297 (25,7%)	129,422
Bourgogne-Franche-Comté	53 (64,6%)	4 (4,9%)	1 (1,2%)	5 (6,1%)	2 (2,4%)	17 (20,7%)	82
	12,238 (27,6%)	10,621 (24%)	5,844 (13,2%)	4,405 (9,9%)	657 (1,5%)	10,571 (23,8%)	44,336
Bretagne	38 (33%)	8 (7%)	1 (0,9%)	6 (5,2%)	11 (9,6%)	51 (44,3%)	115
	15,953 (27%)	11,020 (18,6%)	6,341 (10,7%)	5,553 (9,4%)	2,050 (3,5%)	18,199 (30,8%)	59,116
Centre-Val de Loire	29 (46,8%)	4 (6,5%)	0 (0%)	6 (9,7%)	2 (3,2%)	21 (33,9%)	62
	6,989 (19,6%)	11,524 (32,2%)	0 (0%)	5,137 (14,4%)	32 (0,1%)	12,058 (33,7%)	35,74
Corse	7 (53,8%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	6 (46,2%)	13
	3,486 (66,3%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1,773 (33,7%)	5,259
Grand Est	70 (41,7%)	17 (10,1%)	3 (1,8%)	6 (3,6%)	30 (17,9%)	42 (25%)	168
	17,428 (19,6%)	22,123 (24,9%)	13,176 (14,8%)	6,793 (7,7%)	7,683 (8,7%)	21,553 (24,3%)	88,756
Hauts-de-France	56 (40%)	11 (7,9%)	1 (0,7%)	11 (7,9%)	12 (8,6%)	49 (35%)	140
	21,864 (26%)	15,934 (19%)	6,947 (8,3%)	8,618 (10,3%)	5,242 (6,2%)	25,399 (30,2%)	84,004
Île-de-France	40 (47,6%)	5 (6%)	4 (4,8%)	6 (7,1%)	3 (3,6%)	26 (31%)	84
	7,573 (14,9%)	7,947 (15,7%)	14,210 (28%)	5,419 (10,7%)	0 (0%)	15,627 (30,8%)	50,776
Normandie	70 (44,9%)	10 (6,4%)	1 (0,6%)	7 (4,5%)	12 (7,7%)	56 (35,9%)	156
	37,844 (33%)	26,244 (22,9%)	7,477 (6,5%)	7,157 (6,2%)	2,824 (2,5%)	33,271 (29%)	114,817
Nouvelle-Aquitaine	66 (37,7%)	14 (8%)	2 (1,1%)	7 (4%)	6 (3,4%)	80 (45,7%)	175
	14,735 (12,1%)	20,915 (17,2%)	16,047 (13,2%)	11,572 (9,5%)	1,098 (0,9%)	57,374 (47,1%)	121,741
Occitanie	34 (44,7%)	5 (6,6%)	3 (3,9%)	4 (5,3%)	5 (6,6%)	25 (32,9%)	76
	11,901 (18,9%)	11,374 (18,1%)	12,564 (19,9%)	3,422 (5,4%)	3,916 (6,2%)	19,822 (31,5%)	62,999
Pays de la Loire	53 (34,4%)	10 (6,5%)	3 (1,9%)	4 (2,6%)	15 (9,7%)	69 (44,8%)	154
	14,632 (13,6%)	16,533 (15,4%)	21,924 (20,4%)	6,172 (5,7%)	10,918 (10,2%)	37,176 (34,6%)	107,355
Provence-Alpes-Côte d'Azur	53 (22,3%)	33 (13,9%)	5 (2,1%)	10 (4,2%)	31 (13%)	106 (44,5%)	238
	24,390 (12,6%)	66,406 (34,2%)	34,028 (17,5%)	12,817 (6,6%)	14,981 (7,7%)	41,577 (21,4%)	194,199
Grand Total	667 (40,1%)	142 (8,5%)	26 (1,6%)	79 (4,8%)	142 (8,5%)	606 (36,5%)	1662
	223,630 (20,4%)	252,347 (23%)	155,524 (14,2%)	83,775 (7,6%)	55,547 (5,1%)	32,7697 (29,8%)	1,098,520

Table 1: Number of care centers (N) and overall oncology activity (A) per hospital type

and region. Oncology activity is the sum of the number of patients with radiotherapy or

chemotherapy, and the number of medical or surgery stays related to cancer. *CH* and *CHR/U*

are public hospitals; *CLCC* and *PSPH/EBNL* are private hospitals of collective interest,

though *CLCC* are oncology dedicated; private hospitals are for-profit. *Other* hospitals are

mostly private practice *radiotherapy* structures. The percentages sum to 100% row-wise. In

Nouvelle-Aquitaine region, 47.1% of the oncology activity is handled by private care centers,

whereas in *Provence-Alpes-Cote-d'Azur* region it is 21.4%.

Care centers characterization

While it is obvious that *CLCC* care centers are suited for oncology care, it is difficult to

assess the degree of oncology specialization for other care centers. Our clustering algorithm

assigns the n=1,662 care centers into 8 clusters, sorted by oncology specialization. **Figure 1**

shows the distribution of some of the key health services per cluster. These services are *biol-*

ogy, radiotherapy, chemotherapy, cancer surgery, intensive unit, palliative care, oncology

unit, medication circuit, surgery, and outpatient surgery. The three oncology services are

cancer surgery, radiotherapy, and chemotherapy. We see that care centers from clusters 1

(n=79) and 2 (n=39) all have these 3 services, hence they are the most suited hospitals for

oncology care. Centers from cluster 3 (n=451) have *cancer surgery* and *chemotherapy* but

lack *radiotherapy*. The most part of the n=381 centers from cluster 4 have *cancer surgery*,

but no *radiotherapy* nor *chemotherapy*. Care centers from cluster 5 (n=2) and cluster 6 (n=7)

have *radiotherapy* and *chemotherapy* services, but no *cancer surgery*. Care centers in cluster

7 (n=77) are dedicated to *radiotherapy* and mostly private practice structures. Finally, care

centers 8 (n=626) have none of the 3 oncology services. To sum up, hospitals from clusters 1

and 2 (n=118) are "all-in-one" care centers that provide the most "ideal" oncology care. Cen-

ters from clusters 3 and 4 (n=382) provide oncology care but will have to be coordinated with

additional structures during the pathways. Hospitals within clusters 5, 6 and 7 (n=86) are not

allowed to perform *cancer surgery* but provide *chemotherapy* or *radiotherapy*. The remain-

ing n=626 care centers in cluster 8 are not equipped for oncology care. Hospital types are

unevenly distributed among the clusters. For instance, 76.9% of the *CLCC* care centers are

placed in cluster 1, as they are the most specialized centers. In cluster 7, we find external *ra-*

diotherapy units of some CLCC centers, and private practice structures. The proportion of

private care centers varies as well: cluster 1 has almost no private care center while cluster 2

has 61.5% of private hospitals. Moreover, most of the oncology activity is handled by care

centers from clusters 1 and 3. Also, the overall oncology activity from the n=79 centers in

cluster 1 is almost as large as the activity of the n=451 hospitals from cluster 4.

Figure 1: Distribution of the care centers services and equipment per cluster. Each radar

plot axis shows the percentage of the care centers within the cluster that have the correspond-

ing attribute. In Cluster 1, the care centers have all the listed services. In cluster 8, the centers

have almost none of the services. Care centers from cluster 1 (n=79) and cluster 2 (n=39) are

the most suited for oncology care.

Accessibility score computation

We computed the spatial accessibility score to these care centers for every municipality in

metropolitan France, using the *e2SFCA* algorithm and oncology activity as supply variable.

We compared the accessibility distributions with *e2SFCA* vs. regular *2SFCA*. The accessibil-

ity was lower with *e2SFCA* because of the weight decay. We also studied the influence of the

supply variable in the accessibility score. Accessibility is much higher if we use the number

of *Medical, Surgery and Obstetric* (*MCO*) stays as supply, instead of the oncology activity.

This makes sense since oncology care centers are less common and the overall *MCO* activity

is higher than the oncology activity. The oncology accessibility is unevenly distributed across

the country, as displayed on **Figure 2**. For better readability, we cut the accessibility scores

into 5 quantiles. *Q5* colored in dark green contains the top 20% accessibility municipalities,

and *Q1* in light yellow contains the bottom 20% ones. The lowest accessibility zones are

mostly located in the center of the country and in mountainous regions like the *Alps* or the

Pyrenees. Plot (B) shows that most of the population (51.6%) lives in top 20% accessibility

municipalities, while 6.3 % lives in the bottom 20% quantile. On map (A), care centers are

displayed as squares, colored by cluster index, and sized by oncology activity. We see that

accessibility is highest near the most specialized care centers. Indeed, the proportion of care

centers from specialized clusters decreases in lower accessibility quantiles (C). We then

ranked the departments by median accessibility and showed the top-10 and bottom-10 on plot

(D). Among the top-5 departments, 4 are in *Ile-de-France*. Departments from the bottom-10

are rural or mountainous areas like *Lozère* and *Alpes-de-Haute-Provence*. We notice dispari-

ties within departments as well, as outlined by the large interquartile range in *Hérault* or

Alpes-Maritimes. On the contrary, this spread is very narrow in *Ile-de-France* departments.

Figure 2: Distribution of the accessibility score computed with enhanced two step float-

ing catchment area (e2SFCA), in metropolitan France. Plot (A) shows municipalities col-

ored by accessibility quantile. The care centers are drawn as squares, colored by cluster, and

sized by oncology activity. Plot (B) shows the total population by accessibility quantile. Plot

(C) displays the percentage of care centers by cluster by accessibility quantile. Plot (D) shows

the top 10 and bottom 10 list of the departments, ranked by median accessibility.

Accessibility score should be put into perspective with population density. Overall, the

denser municipalities have a median accessibility around 0.02. Municipalities with low popu-

lation densities have more extreme values. **Figure 3** compares accessibility and population

density for three different regions: *Provence-Alpes-Cote-d'Azur* (A), *Ile-de-France* (B), and

Bourgogne-Franche-Comté (C). Municipalities are displayed as squares, colored by accessi-

bility quantile, and sized by population density. These regions show very different profiles. In

Provence-Alpes-Cote-d'Azur (A), accessibility is essentially low in non-dense municipalities

near the *Alps*. However, in *Bourgogne-Franche-Comté* (C), we see dense municipalities with

poor accessibility scores, representing a large proportion of the region. We also drew similar

maps (D, E and F) where municipalities are colored based on the average travel duration for

patients with cancer in 2018. We see that the average travel time is higher in municipalities

with poor accessibility scores. The surface percentage with low accessibility varies from a

region to another. For instance, in *Bourgogne-Franche-Comté*, 34.5% of the region has a Q1

accessibility, that is 15.6% of the region's population. Sometimes, the Q1 surface can be

large but might contain very few inhabitants. This happens in *Ile-de-France*, where 15% of

the surface is Q1 accessibility, representing less than 1% of the region's population. Finally,

we compared our accessibility score with the department exit ratio, by municipality. Depart-

ment exit ratio is defined as the proportion of cancer patients who visited a care center out-

side from their department of residence and was computed using the *PMSI* database. In *Pro-*

vence-Alpes-Cote-d'Azur, the exit ratio is higher in departments with low accessibility scores

and few oncology specialized care centers, as in *Alpes-de-Haute-Provence* and *Hautes-Alpes*.

While the *Var* department has some oncology centers, exit ratio remains high since larger

care centers are in *Marseille* and *Nice*.

Figure 3: Comparison of population density with accessibility scores and patient

average travel time for cancer pathways. Showing results in three regions: *Provence-*

Alpes-Cote-d'Azur (A, D), *Ile-de-France* (B, E) and *Bourgogne-Franche-Comté* (C, F).

Municipalities are drawn as squares, sized by population density and colored by either

accessibility quantile (A, B, C) or patient average travel time (D, E, F).

We now focus on the region *Provence-Alpes-Cote-d'Azur*. This region is the far south-

eastern on the mainland. The region's population was 5,048 million in 2018. Its prefecture

and largest city is *Marseille*. The region contains six departments. *Bouches-du-Rhone*, *Var*

and *Alpes-Maritimes* are located on the coastline and gather the largest cities like *Marseille*,

Nice, or *Toulon*. *Alpes-de-Haute-Provence, Vaucluse, and Hautes-Alpes* are inland depart-

ments, with a majority of rural and mountainous areas. Results are shown on **Figure 4**. By

comparing maps (A) and (B), we confirm that the accessibility is maximum in denser areas of

the region. Average patients travel time are displayed on map (C) and we drew the major

roads (primary, motorway, and truck) in red. The road system is well developed on the coast,

rallying the larger cities of the region. However, driving from the rural areas in the *Alps* to the

major cities is hard, resulting in higher travel times. The accessibility is unevenly spread

within the departments, especially in *Alpes-Maritimes* where the distribution is multi-modal

(D). There, cities like *Nice* and *Cannes* have large hospitals thus good accessibility, while the

northern areas of the department are mostly mountains. Accessibility is higher in municipali-

ties with dense populations, for all the departments (E). Finally, the average travel time de-

creases when the accessibility score increases. This makes sense since the accessibility score

was computed based on the driving distance between population locations and care centers.

However, it confirms that patients living in poor accessibility zones effectively travel further

to seek oncology care. In *Bouches-du-Rhone*, nearly all the municipalities have an average

travel time lower than 30 minutes, while in *Alpes-de-Haute-Provence*, average travel times

are rarely lower than 60 minutes (F).

Figure 4: Accessibility distribution in Provence-Alpes-Cote-d'Azur region. Map (A)

shows the region accessibility distribution per municipality. Map (B) displays the population

density discretized in 5 bins. The map on plot (C) displays the average travel time for cancer

pathways. Large roads (primary, motorway, and trucks) are drawn in red. Plot (D) shows the

accessibility distribution per department of the region. Plot (E) shows the accessibility distri-

bution by municipality population density and department. Plot (F) compares the accessibility

score from municipalities with the average travel time for cancer pathways.

Accessibility optimization

Since we focused on describing the accessibility situation in *Provence-Alpes-Cote-d'Azur*,

we now present the outcomes of our optimization algorithm in this same region. The algo-

rithm was run with the user-specified parameters stated in the Methods Section: we chose to

increase the overall oncology activity in the region by 3% (+3,221 activity) and capped care

centers to a 20% maximum growth. The median accessibility in the region went from 0.0093

to 0.0103, a 11.1% increase. The results are shown on **Figure 5**. Map (A) displays the acces-

sibility delta ($A_{i_{after}} - A_{i_{before}}$) as well as the care centers eligible to grow. Centers from clus-

ter 8 were hidden since we considered that they couldn't provide any oncology activity. The

algorithm identified a list of 26 care centers where the oncology activity could grow to max-

imize the total accessibility in the region. These centers are either public or private hospitals,

primarily located in the *Avignon* and *Gap* areas. The care centers located in high accessibility

areas near *Marseille* and *Nice* were ignored by the algorithm because improving these zones

is not a priority. The care center that grew the most is *Clinique Sainte Catherine*, in *Avignon*.

Interestingly, this care center was recently bought by the *Unicancer* group, which coordinates

all the cancer centers in France. This hospital's type will change to become a new *CLCC*.

Thus, it is expected to grow in the next years and to be equipped with more oncology services

and staff.

Figure 5: Accessibility delta in Provence-Alpes-Cote-d'Azur (PACA) region after running the optimization algorithm. Map (A) displays the accessibility delta ($A_{i_{after}} - A_{i_{before}}$) by municipality. Plot (B) shows the capacity delta ($S_{u_{after}} - S_{u_{before}}$) distribution. Capacity was defined as the oncology activity: the number of patients with chemotherapy or radiotherapy and the number of medical or surgery stays related to oncology. We show the list of the care centers that grew the most (C) and by how much. For instance, the hospital *Institut Sainte Catherine* in *Avignon*, was assigned a +1,030 capacity, for a total of n=6,179. Additional activity was 3,221. 26 centers grew and 1 decreased. Median accessibility before optimization was 0.0093 and 0.0103 after, corresponding to a 11.1% increase. Accessibility increased around cities like *Avignon* and *Gap*. Care centers near *Nice* were left unchanged by the algorithm.

While we described the results in *Provence-Alpes-Cote-d'Azur* region, we ran the algorithm with similar parameters on every region in metropolitan France. The results are available in the *Supplementary Materials* and on the web application. We observe two types of optimization strategies. For most regions, the algorithm manages to find a couple of areas where the accessibility can be locally improved, like it did in *Provence-Alpes-Cote-d'Azur* near *Gap* and *Avignon*. However, for regions like *Ile-de-France* and *Haut-de-France*, the hospital capacity increase is more uniformly distributed across the region. Most of the time, the algorithm left untouched the large care centers located in dense cities with good accessibilities. This can be explained by the relatively low value of the additional activity parameter: with a very large value of additional activity, every care center will grow. If we keep it low, the algorithm identifies in which areas hospital capacity should be increased in priority.

Discussion

We observe disparities in both care centers and their accessibility. The clustering algo-
rithm successfully groups similar hospitals and lets us identify the care centers best suited for
oncology care. Some variables in the *SAE* survey are declarative and potentially differ from
the reality. We are aware of this bias, but we do not expect major differences that could dis-
tort our clustering results.

Receiving treatment in a care center with *surgery*, *chemotherapy* and *radiotherapy* activi-
ties is easier for the patient and leads to better care pathways. Care centers from cluster 1 will
be the better choice for cancer treatment and correspond to modern oncology care specifica-
tions. However, these centers are a minority and sparsely located, essentially in dense areas
and in large cities. While the inhabitants of large cities and metropolitan areas will have no
problem reaching them, rural areas residents live far away from these centers. This popula-
tion often has better access to care centers from intermediate clusters. Such centers do not
have all the key services and the patients are more likely to visit multiple hospitals during
their care pathways.

Longer drives to reach a more specialized care center could be considered more acceptable
for *surgery*, where the hospital volumetry and surgeon expertise matter. However, for more
frequent interventions like *chemotherapy* and *radiotherapy* especially, patients should priori-
tize short travels. There is a tradeoff to be found by patients, between care center proximity
and care center expertise. This dilemma will be more frequent for patients living in rural are-
as than patients living in dense cities with large care centers nearby.

Specific attention should be given to municipalities with very poor access to oncology
care centers. While we saw that most of the population lives in high accessibility areas,
around 6% of the population lives in the bottom 20% accessibility quantile. Among these
municipalities, some are very rural and mountainous like those in the *Alpes-de-Haute-*

Provence in *Provence-Alpes-Cote-d'Azur* region. Such areas cannot be expected to have a

very good healthcare coverage. By contrast, the case of suburban areas with relatively dense

population and poor accessibility should be addressed more easily. Our optimization algo-

rithm can help driving public health policies, as it effectively identifies areas where accessi-

bility could grow, by allocating additional oncology activity to a restricted number of care

centers. The proposed growth factors are indicative and do not have to be effective within a

year, as it represents a considerable effort for care centers to increase their activity.

Our oncology accessibility score is deliberately non-specific to cancer type. This score is

meant to outline how easy it would be for a population location to reach a first entry point for

oncology care. Here, we are only focusing on surgery, chemotherapy, and radiotherapy

treatments. The same technique could be used on a specific cancer type, the method will re-

main the same, only the supply variable used in the accessibility score will change. We

should mention that spatial accessibility is better suited for pathologies that are relatively well

handled across the whole country. Accessibility for rare diseases like pediatric cancer or

complex cancers that require a specific expertise is less informative because only a handful of

care centers are indicated.

Similarly, we could compute an accessibility score that is focused on specific kinds of

stays: our web application lets the user pick between surgery, chemotherapy, or radiotherapy

as supply variable.

The quality of oncology care is linked with the care centers' volumetry. A care center with

a very low activity is less likely to provide decent care. As a result, the *French National Insti-*

tute of Cancer (INCa) defined several thresholds (36) that forbid care centers with very low

activity to keep operating. Similarly, the care quality in a saturated care center won't be good

either, since patients are more likely to wait longer before diagnosis or between interventions.

While it is easy to spot care centers with low activity, it is harder to judge if a care center is

over-crowded, and we should be careful when attributing new activity to the hospitals. We

based the 20% max growth out of the previous centers' activity increase. This percentage

could be tailored to the center cluster or current activity. Volumetry is not the only factor

determining care quality. More sophisticated indicators like average delay between diagnosis

and first treatment can tell whether a care center is in line with the care pathways recommen-

dations. Care centers with activities lower than the thresholds, or with a large proportion of

degraded pathways should be handled with care by our algorithm.

Accessibility optimization depends on many factors and healthcare professionals will not

have the same uses for our algorithm. Some may consider that for a care center to grow an-

other should decline, where others would rather not decrease any centers' activities. Moreo-

ver, the healthcare planning is very different from a region to another, and even within the

regions departments are showing disparities. Hence, we cannot expect the algorithm to be

used with the same parameters on every region. For all these reasons, we believe that provid-

ing a web application to run the algorithm and choose the parameters is the most useful way

to the help healthcare professionals improve the current situation.

Our work is in line with the *French Cancer Plan* (37) that emphasizes the importance of

increasing accessibility to oncology care as well as minimizing disparities across the country.

The government mandated *INCa* to work on the accessibility development. This study and

the web application we developed could help when attributing the care centers authorizations.

Working closely with researchers from *INCa* and public health professionals could have a

major impact on the oncology care spatial organization in metropolitan France, benefiting

millions of patients.

We ran this method in metropolitan France, but it could work on any country if data on

hospitals and municipalities are available.

Acknowledgements

The authors thank Thomas Ansart and the *Sciences Po* cartography workshop for helping us to improve our figures. We also thank Julien Guerin and Johan Archinard, from the *Data Factory at Institut Curie*, who helped us to deploy our web application. Finally, we thank Hakim Idjis, Marc-Felix Degni and Olivier Auliard and their team at *Capgemini Invent*, who helped us to extract the driving routes with OpenRouteService.

References

1. Johnson AM, Hines RB, Johnson JA, Bayakly AR. Treatment and survival disparities in lung cancer: the effect of social environment and place of residence. Lung Cancer Amst Neth. 2014 Mar;83(3):401–7.

2. Dejardin O, Jones AP, Rachet B, Morris E, Bouvier V, Jooste V, et al. The influence of geographical access to health care and material deprivation on colorectal cancer survival: Evidence from France and England. Health Place. 2014 Nov 1;30:36–44.

3. Haynes R, Pearce J, Barnett R. Cancer survival in New Zealand: ethnic, social and geographical inequalities. Soc Sci Med 1982. 2008 Sep;67(6):928–37.

4. Wang F. Measurement, Optimization, and Impact of Health Care Accessibility: A Methodological Review. Ann Assoc Am Geogr Assoc Am Geogr. 2012;102(5):1104–12.

5. Khan AA. An integrated approach to measuring potential spatial access to health care services. Socioecon Plann Sci. 1992 Oct;26(4):275–87.

6. Guagliardo MF. Spatial accessibility of primary care: concepts, methods and challenges. Int J Health Geogr. 2004 Feb 26;3(1):3.

7. Zahnd WE, Josey MJ, Schootman M, Eberth JM. Spatial accessibility to colonoscopy and its role in predicting late-stage colorectal cancer. Health Serv Res. 2021;56(1):73–83.

8. Alahmadi K, Al-Zahrani A, Al-Ahmadi S. Spatial Accessibility to Cancer Care Facilities in Saudi Arabia. In 2013.

9. Launay L, Guillot F, Gaillard D, Medjkane M, Saint-Gérand T, Launoy G, et al. Methodology for building a geographical accessibility health index throughout metropolitan France. PLOS ONE. 2019 août;14(8):e0221417.

10. Gusmano MK, Weisz D, Rodwin VG, Lang J, Qian M, Bocquier A, et al. Disparities in access to health care in three French regions. Health Policy Amst Neth. 2014 Jan;114(1):31–40.

11. Gao F, Kihal W, Le Meur N, Souris M, Deguen S. Assessment of the spatial accessibility
to health professionals at French census block level. Int J Equity Health. 2016 Aug
2;15(1):125.

12. Wang F. Why Public Health Needs GIS: A Methodological Overview. Ann GIS.
2020;26(1):1–12.

13. Weiss DJ, Nelson A, Vargas-Ruiz CA, Gligorić K, Bavadekar S, Gabrilovich E, et al.
Global maps of travel time to healthcare facilities. Nat Med. 2020 Dec;26(12):1835–8.

14. Bauer J, Klingelhöfer D, Maier W, Schwettmann L, Groneberg DA. Spatial accessibility
of general inpatient care in Germany: an analysis of surgery, internal medicine and neu-
rology. Sci Rep. 2020 Nov 5;10(1):19157.

15. Church RL. Location modelling and GIS. Geogr Inf Syst. 1999;1:293–303.

16. Luo J, Tian L, Luo L, Yi H, Wang F. Two-Step Optimization for Spatial Accessibility
Improvement: A Case Study of Health Care Planning in Rural China. BioMed Res Int.
2017 Apr 18;2017:e2094654.

17. Tao Z, Cheng Y, Dai T, Rosenberg MW. Spatial optimization of residential care facility
locations in Beijing, China: maximum equity in accessibility. Int J Health Geogr. 2014
Sep 1;13(1):33.

18. Krugman P. Opinion | Why Inequality Matters. The New York Times [Internet]. 2013
Dec 16 [cited 2022 May 2]; Available from:
https://www.nytimes.com/2013/12/16/opinion/krugman-why-inequality-matters.html

19. Meyer D. Equity and efficiency in regional policy. Period Math Hung. 2008 Mar
1;56(1):105–19.

20. Culyer AJ, Wagstaff A. Equity and equality in health and health care. J Health Econ.
1993 Dec;12(4):431–57.

21. Hemenway D. The Optimal Location of Doctors. N Engl J Med. 1982 Feb
18;306(7):397–401.

22. Fried C. Rights and health care--beyond equity and efficiency. N Engl J Med. 1975;

23. Oliver A, Mossialos E. Equity of access to health care: Outlining the foundations for ac-
tion. J Epidemiol Community Health. 2004 Sep 1;58:655–8.

24. Lloyd S. Least squares quantization in PCM. IEEE Trans Inf Theory. 1982
Mar;28(2):129–37.

25. Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters
in large spatial databases with noise. In: Proceedings of the Second International Con-
ference on Knowledge Discovery and Data Mining. Portland, Oregon: AAAI Press;
1996. p. 226–31. (KDD'96).

26. Luxburg UV. A Tutorial on Spectral Clustering. 2007.

27. Luo W. Using a GIS-based floating catchment method to assess areas with shortage of
physicians. Health Place. 2004 Mar 1;10(1):1–11.

28. Luo W, Qi Y. An enhanced two-step floating catchment area (E2SFCA) method for
measuring spatial accessibility to primary care physicians. Health Place. 2009 Dec
1;15(4):1100–7.

29. Murad A, Faruque F, Naji A, Tiwari A. Using the location-allocation P-median model for
optimising locations for health care centres in the city of Jeddah City, Saudi Arabia.
Geospatial Health. 2021 Oct 19;16(2).

30. Shavandi H, Mahlooji H. A fuzzy queuing location model with a genetic algorithm for
congested systems. Appl Math Comput - AMC. 2006 Oct 1;181:440–56.

31. Casado S, Laguna M, Pacheco J. Heuristical labour scheduling to optimize airport pas-
senger flows. J Oper Res Soc. 2005 Jun;56(6):649–58.

32. Wang F, Tang Q. Planning toward Equal Accessibility to Services: A Quadratic Pro-
gramming Approach. Environ Plan B Plan Des. 2013 Apr 1;40(2):195–212.

33. Li X, Wang F, Yi H. A two-step approach to planning new facilities towards equal acces-
sibility. Environ Plan B Urban Anal City Sci. 2017 Nov 1;44(6):994–1011.

34. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al.
SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods.
2020 Mar;17(3):261–72.

35. Bertsimas D, Tsitsiklis J. Introduction to Linear Optimization. Athena Scientific. 1998.

36. Institut National du Cancer. Mesure des activités soumises à seuil. 2017.

37. Buzyn A. Le Plan cancer 2014-2019 : un plan de lutte contre les inégalités et les pertes
de chance face à la maladie. Trib Sante. 2014 Jul 15;n° 43(2):53–60.