2 Spatial equity and efficiency in urban health care

2.0 Introduction

This chapter lays the foundation for the remainder of the study and clarifies the relationship between the proposed study and previous work conducted on the topic. It consists of a number of building blocks, which together form the theoretical, the methodological and the contextual framework of the study. The discussion commences with a short background overview of the emergence of urban health care as a central theme in national health systems in sub-Saharan Africa and identifies the main objectives and guiding principles of public health care provision.

The first building block, describes the theoretical framework of the study. We take the guiding principles of health care as our point of departure and place them in a spatial context through a discussion of two important bodies of spatial theory: neoclassical location theory, and spatial welfare theory as formulated by (Smith, 1977). The importance of describing the neoclassical approach is that it provides us with fundamental spatial analytic constructs, which can be used to create a simplified model of the working of health care systems in geographical space. Spatial welfare theory builds upon the neoclassical approach and yet represents a more realistic framework: it enables us to move our simplified model of the spatial organization of health care closer to the complexity of reality.

The second building block describes the methodological framework. The discussion centres on the translation of previously identified theoretical constructs into appropriate methods and techniques. It consists of an exploration of the concept of accessibility, its most common operationalizations and the identification of the most appropriate type of accessibility measure for this study.

The third building block is the contextual framework which is an account of *(i)* the empirical literature on accessibility for health services and *(ii)* the applicability of analytical GIS approaches for health care planning in sub-Saharan contexts. Together, these building blocks culminate in the formulation of the conceptual framework and a step-by-step description of the remainder of the study.

2.1 Urban health care and its guiding principles

Health is a universal human right and a major focus of social and political concern worldwide. Health also is a somewhat elusive concept that can be defined in various ways. One definition comes from the World Health Organization (1978) who regards

health as 'a state of complete physical, mental and social well-being'. Such an ideal state may be desirable but is of limited practical use as it implies that most of humanity would be unhealthy at all times. More productively, we might think of health as being physically and mentally 'fit' and capable of functioning effectively for the good of the wider society. Seen from this perspective, health is viewed upon as a prerequisite for social advancement and economic progress: the healthier people are, the more likely they are able to contribute to social and economic development. Economic growth, in turn, can enable responsible authorities to provide better health care and thus improve the health status of a population. This does not mean that economic growth automatically results in improved health status or that a high Gross National Product is required before health can be improved. Much will depend on the level of available health resources and the distribution mechanism that is in place.

During the 1970s, increasing recognition was given to the fact that poverty, inadequate living conditions and malnutrition were at the root of poor health. These insights culminated in the Alma Ata declaration where WHO and UNICEF launched the worldwide 'Health-for-All' strategy. The Health-for-All strategy aims to reach and maintain the highest attainable level of health for all people. It consists of a range of services intended to promote good health, prevent illness and treat people that have become ill. It is a practical approach to making essential health care universally accessible to people in an acceptable and affordable way. At the very least, all people should have such a level of health that allows them to work productively and participate in social life (World Health Organization and Unicef, 1978). Five basic principles underlie the strategy: adoption of a multi-sectoral approach, a focus on prevention, involvement of local communities in decision-making, use of appropriate health technologies, and an equitable distribution of services. In the past decades, primary health care has become adopted throughout the developing world.

In most sub-Saharan countries, post-independence governments adopted the primary health care approach and took on a dominant role in the provision of health care with the objective to provide universal access to health services at no or little cost to the user (Berman, 1998). Initial emphasis was on extending health service coverage in rural areas. Towards the beginning of 1990s, it was increasingly recognized that rapid expansion of urban areas was irreversible and resulted in profound inequalities in health status and access to health care amongst city populations. Since then, *urban* primary health care has become and remained a major concern of public health authorities worldwide (Tabibzadeh and Rossi-Espagnet, 1989; Atkinson, 1993; Atkinson and Merkle, 1994).

Since the mid 1990's national health systems in African countries are being restructured in an effort to cope with increased demand for health care in an environment of public budget constraints. The outcome of these changes is a shift away from pure state provision of health care towards a situation with greater reliance

on market forces. Throughout sub-Saharan Africa, this takes the form of the emergence of market driven private health care provision and the introduction of cost-recovery measures in the more centrally planned public health care sector. These adjustments should be seen in the broader context of a resurge of global belief in 'free market' processes and a re-thinking of government roles in social policy (Creese *et al.*, 1998). Under these new conditions, inequalities in health status and in accessibility to health care are increasing (Lerer, 1998 56). Turshen (1999) describes this in terms of a declining role of the World Health Organization and a growing hegemony of the World Bank. This issue is taken up in further detail in Chapter 4.

The above-described developments illustrate that a new situation has emerged whereby public and private health care systems operate side by side, but each is driven by different principles. The essential distinction is that private health care provision is market-led and based upon the expectation that the health care user pays at the point of use or buys into private health insurance - generally tied to employment - that meets the costs of treatment. Public health care provision, by contrast, makes provision available on the basis of need rather than ability or willingness to pay. Curtis and Taket (1996) refer to the latter ideology as collectivist, which means that public health services are not paid for by the individual user, but are funded via some form of general taxation or compulsory insurance. It will be clear that the primary health care approach is essentially collectivist and driven by welfare objectives, such as the improvement of the quality of life and achieving maximum health benefits to the greatest number of people.

When making decisions about the provision of public health services the following guiding principles apply. The *first* is to provide services in an equitable way. Notions of fairness and social justice are invariably stated as being at the core of public health care interventions. Normally, this viewpoint entails that priority should be given to the more vulnerable population groups. The *second* objective is to provide services that are effective. Treatments should offer real benefit; resources should not be wasted on providing services that offer little health gain. The *third* objective is to provide services that are efficient. By this we mean how to provide services that maximize health benefit to society while minimizing the cost of provision (Carr-Hill, 1994). As a general rule there will be tension between the efficiency and equity objectives: providers wish to manage resources efficiently whereas the wish of users will be to have resources provided equitably. Health care provision is thus inevitably tied up with issues of resource allocation, distribution and priority setting: decisions have to be made about the nature and range of services to provide and how they are distributed amongst the members of society. Obviously, these are questions of an economic and political nature.

At the same time these are questions of a geographical nature since health care services will have to be provided at particular points in geographical space. Because of this

there will always be a degree of geographical variation between the provision and use of health services. The real question that needs to be addressed is therefore whether such geographical variation is acceptable or represents inequity. This subject is further explored in the sections that follow.

2.2 Theoretical framework: a spatial perspective on welfare maximization, equity and efficiency

In this paragraph we translate the guiding principles of urban health care into a spatial context, clarify the need for a locational decision making process (spatial policy) for distributing public health resources, and provide an ethical framework that enables us to evaluate the performance of alternative spatial distributions of health resources in terms of their contribution to health needs satisfaction.

We do so in a step-by step manner. We begin with a brief overview of neoclassical economics and its spatial extension. Taken together, this body of theoretical thought enables us to establish a simplified model for the spatial organization of health care provision. We then continue with a discussion of spatial welfare theory, which brings our model closer to reality by identifying the main shortcomings of neoclassical thought and presenting alternative solutions. Although most of the discussion is 'generic' the link to the 'production', 'consumption' and distribution of health care is apparent throughout the discussion.

Neoclassical economic theory: a basic regulatory mechanism

An appropriate point of departure for our analysis is neoclassical economics, which presents a view of how economic activity functions in capitalist society. It considers the economic processes of production, consumption and distribution of goods and services. Production concerns the creation of all the goods and services needed by society, consumption concerns their utilization by the members of society. What is produced and in what quantity depends, in theory, on what consumers are prepared to buy which, in turn, depends on the price asked by the producer and on what the consumer can afford. In neo-classical economic theory, markets automatically adjust to changes in demand and supply and tend towards a state of equilibrium at a price that brings supply and demand into balance by virtue of the intrinsic price mechanism (Dicken and Lloyd, 1990).

The concept of utility - that which individuals try to maximize - is central to the theory of consumer behaviour. Consumers maximize utility on the basis of individual tastes and preferences for goods and services and allocate their expenditure accordingly among alternative goods or services, subject to their budget constraint - represented by

income - and the prevailing set of prices. The theory of consumer behaviour is based upon the assumption of consumer sovereignty, which means that the individual consumer has perfect knowledge of the market and acts economically rational. The analysis of individual consumer behaviour can subsequently be extended to represent the collective consumption of an entire society. This is achieved by aggregating individual utility functions into a social welfare function expressing community preferences for goods and services.

The neo-classical analysis of production is analogous to that of consumer behaviour: to maximize profit, the individual producer operates at the highest level of efficiency in its use of resources and produces at the lowest cost. Producers and consumers are both guided by the desire to optimize: to get the best possible outcome for their efforts. Producers seek maximization of profits, consumers aim to minimize their expenditure in meeting their consumption needs.

When production and consumption are brought together, the price mechanism ensures that resources are distributed among alternative goods and services in such a way that no reallocation is possible without diminishing the total value of production and the overall utility or welfare derived from it (Walker, 1981). The 'invisible hand' of the market guides economic activity towards optimal resource allocation and maximization of social welfare (Goodall, 1987).

Such an ideal economic or social state is termed Pareto optimal: it is impossible to make any individual better off without at the same time making someone else worse off. Proponents of neoclassical economics consider Pareto optimality as a situation in which a society achieves optimal allocative efficiency as well as distributive equity. We will return to this issue later but first we explore how neo-classical analysis can be extended to incorporate geographical space.

Neoclassical location theory: a simplified spatial model

In this section we reconsider the processes of production, consumption and distribution but now from an explicitly geographical perspective. In doing so we adhere to the assumptions and mechanisms embedded in neoclassical economic thought. At the same time, we need to introduce some additional assumptions. These assumptions allow us to develop a simplified spatial model of the real world, which enables us to focus explicitly on the role of geographic distance in the spatial organization of health care provision.

It is assumed that consumers all live equidistant from one another on an isotropic plane in which transportation costs are uniform and exactly proportional to distance in all directions. Consumers exhibit uniform utility maximizing behaviour, which means that they have identical incomes, demands and tastes.

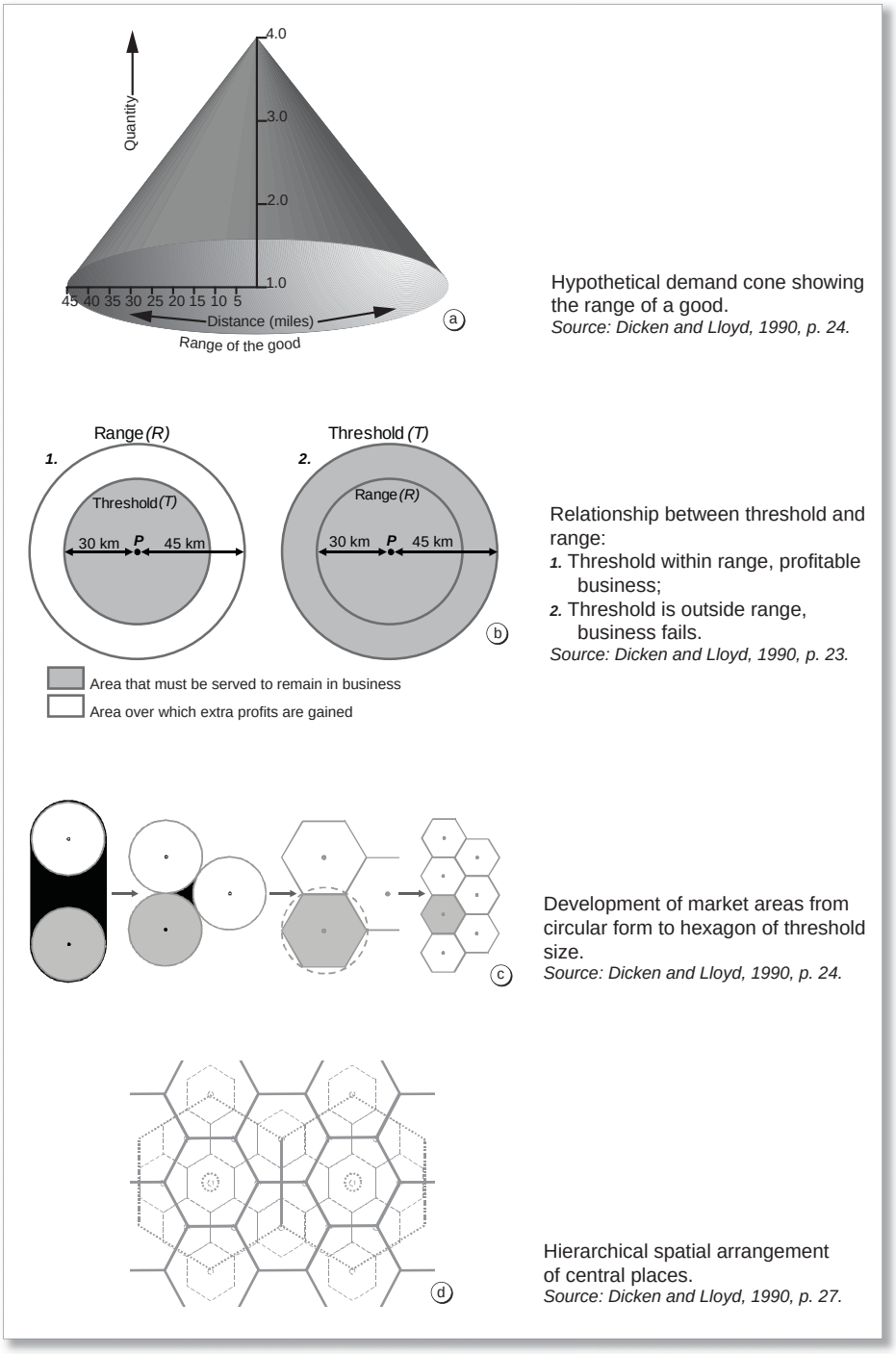

Figure 2.1: Main concepts of neoclassical location theory.

Producers and consumers both act economically as well as spatially rational. As we recall from the theory of consumer behaviour, the rate of consumption of a utility maximizing individual depends on the price or exchange value of a good or service.

In a world in which geographical space matters, this price will be made up of two elements. The first element is the market price of the good or service at the point of sale. The second element is the cost of travelling to and from that point. This means that, with a fixed budget of time, money or effort, the amount of a particular good or service or the frequency with which it will be consumed, will decrease with increasing distance from the production facility up to a point where demand for that good or service will become zero (see Figure 2.1*a*).

Adoption of this spatial perspective implies that the production of goods and services will depend on two economic mechanisms: the threshold value and the range of a good. The threshold value is defined as the minimum level of consumer demand needed for the profitable production of a particular good or service. The range of a good or service describes the effective size of this demand, which is defined by the maximum distance consumers are willing to travel to obtain a particular good or service. If the threshold exceeds the range, profitable provision of a good is not possible and the good or service will not be provided. Range and threshold value both have a spatial form as illustrated in Figure 2.1*b*.

The analysis of individual producer behaviour in geographical space can subsequently be extended to represent the collective spatial arrangement of production within a defined territory. This is explained as follows. Additional producers of the same good or service will enter the market until total consumer demand is satisfied. In doing so subsequent producers will have to take into account the existing location and market area of already established producers to avoid unnecessary competition. As more and more producers enter the market, the spatial arrangement of production will exhibit a progression from initial non-overlapping circular market areas towards hexagonal market areas. This progression is illustrated in Figure 1.1*c*. Given that there is free entry into the market and a desire to optimize, the geographical extent of individual hexagonal market areas will tend to approach the threshold size. The result is a compact, uniform lattice of production centres, each serving hexagonal market areas of identical size, equivalent to the threshold size for that good.

So far, the discussion has limited itself to the production of one good or service only. In reality, however, we will see that a diversity or 'bundles' of goods and services will be produced. Each will have a specific range and a specific threshold value. Inexpensive, regularly demanded (lower order) services will have a small spatial range and threshold value while expensive, irregular (higher order) services will have a larger spatial range and threshold value. Put together, these postulates and assumptions lead to the formation of a regular interlocking hierarchy of hexagonal lattices, each

describing the non-overlapping market area of a different service outlet. Higher order outlets are located at the centre of large hexagons; lower order outlets will locate at the centre of small hexagons (see Figure 2.1*d*).

The outcome of a perfect market in a space economy would, in theory, generate a 'spatial Pareto' situation. No alteration of what is produced and consumed *where* could make anyone anywhere better off without at the same time making someone else some*where* else worse off (Smith, 1977, p. 146). In our simplified model of the space economy, spatial Pareto optimality is seen as a situation in which a society achieves optimal spatial efficiency as well as equitable spatial distribution.

Christaller (1966) used the above-described assumptions and economic mechanisms to formulate his central place theory, which seeks to explain an apparent order in the spatial arrangement of urban settlements in terms of their sizes and their intermittent spacing. Settlements are seen as central places providing goods and services to the surrounding populations. Central place theory is concerned primarily with the locations of (retail) service activities and is described as being complementary to Weber's theory of industrial location and von Thünen's theory of agricultural location (Davies, 1967). Berry and Garrison (1958) extended the range of central place applications to the spatial organization of a hierarchy of retail services within urban areas. The analogy to the intra-urban provision of health care provision is then easily made.

What now is the relevance of neo-classical location theory for spatial health care planning? First of all, we have obtained a first insight in the way - at least in theory - in which a society may maximize human welfare in an economic as well as in a spatial sense. Furthermore, the discussion has provided us with two analytical constructs of fundamental importance: threshold value and range. On the basis of these two mechanisms, the characteristics of service supply in relation to the spatial expression of consumer demand can be described. Important also is that we are not only concerned with the location of individual service supply points, but also how they function in relation to one another in geographic space. Finally, the notion of hierarchy in service supply has been introduced. Each of the above mentioned issues are essential considerations in the structure of any health care system. Apart from the hierarchical nature of health care systems, concepts such as market area, efficiency of provision and equality of access are fundamentally important considerations.

At the same time, the normative nature of the neoclassical approach and its strict assumptions make that the model does not reflect reality sufficiently well. In the next paragraph, we move away from this rather rigid approach and re-examine some of the restrictive assumptions in an attempt to bring our model closer to reality. We do so through a discussion of spatial welfare theory as formulated by (Smith, 1977; Smith, 1994).

Spatial welfare theory: a more realistic spatial model for urban health care

The spatial welfare approach defines the state of society by the quantities of goods (and bads) produced and their distribution among consumers territorially disaggregated. Spatial welfare theory objects to the narrow perspective of neo-classical thought and demonstrates that an unregulated free market system does not culminate in welfare maximization but results in unacceptable social, economic and spatial inequality. It provides arguments why governments need to intervene in imperfect markets to assist in welfare maximization and provides an ethical framework within which rational policy judgments can be made (Smith, 1994, pp. 389-394). Analogous to health care, the welfare approach focuses on quality-of-life issues, emphasizes questions of social justice, and highlights distributional issues with a spatial dimension.

The keywords of the welfare approach are 'who gets what where and how' (Smith, 1977). The *who* refers to the population of the area under study, subdivided into groups on the basis of relevant characteristics. The *what* refers to the goods and bads enjoyed. The emphasis on *where* provides the spatial perspective and concerns the identification and understanding of place-to-place variations in welfare, which includes differential access to health services. The *how* refers to broader societal processes, such as the functioning of the economic and political system, that influence human welfare (Smith, 1977; Chisholm, 1995; Knox, 1995).

Spatial welfare theory identifies a number of shortcomings of the neo-classical perspective. The fundamental objection to neo-classical thought is its reliance on unrestricted individual consumer preferences as the origin of collective consumption patterns: the individual is seen as the primary atom and his/her preferences as the ultimate data for the welfare of society. The counter position of spatial welfare theory is that consumer preferences and the budget constraint are not autonomous but strongly influenced by the society and economy in which the individual lives. Income reflects the position in the economic, social and political structure. Poor people may not be able to express demand effectively in the marketplace, thus limiting opportunity for satisfaction (in severe cases, basic needs may not be fulfilled at maximum utility). In such situations, consumption patterns do not reflect free choice but rather display repressed preferences or constrained choice. Likewise, consumer preferences are not autonomous but subject to external influences. People are not freely deciding atomistic individuals but members of societal groups, which also influences their preferences. As a result, different groups within society are bound to have different - possibly conflicting - preferences. In most cases this will mean that the preferences of the richer will be implemented because they have the political power. The implication is that the aggregation of individual utility functions into a social welfare function is unrealistic as it erroneously assumes community consensus on how welfare maximization should be achieved. In other words, neoclassical theory correctly identifies tastes, prices and the

income constraint as the immediate determinants of consumption but fails to investigate the origins of personal preferences and of the budget constraint. Focus is on welfare maximization only, there is no concern for the way in which goods and bads are distributed among the members of society (this issue is further discussed in the next section).

Another obvious shortcoming of neoclassical thought is the assumption that individuals act as 'homo economicus'. In reality consumers and producers are not perfectly informed and rational beings capable of optimization. They will not fully maximize utility/profit as they have sub-optimal knowledge of the market and imperfect ability to process this knowledge. Instead of optimizing, they exhibit satisficing behaviour, that is, they make decisions that allow them to reach a threshold level of satisfaction. Upon achieving the threshold, they feel little incentive to strive for maximum satisfaction.

Spatial welfare theory incorporates geographical space as an element of utility or welfare but in a way that differs from neoclassical location theory. The very theoretical constructs of the isotropic plane and uniform consumer behaviour are abandoned in favour of the realistic conception that alternative locations in geographic space offer alternative combinations of goods and bads and that consumers will exhibit different spatial behaviour as a result of differences in preferences and incomes. The individual will nevertheless tend to locate himself at that place whose characteristics posses a higher level of utility than other places known to him. The concept of place utility - the spatial equivalent of economic utility - is used to describe consumption in space. Place utility refers to the net composite of (dis-) utilities, which an individual derives at some point in geographic space. In their spatial behaviour, individuals are not only constrained by imperfect knowledge but also by economic constraints - embodied by the budget constraint - that impact on them. In other words, as people arrange themselves in geographic space, they may not necessarily reveal real preferences and may have unfulfilled needs.

It is important to note that the place utility of a particular location is not only a function of the intrinsic values of the location itself, but also of its relative spatial position with respect to other places. For many sources of needs or want satisfaction, the individual depends on movement to sources of supply. Spatial accessibility with respect to, for example, health care services is, therefore, an element of real income, which impacts on human well-being.

The outcome of the process of consumption in space is a situation characterized by place-to-place variations in quality-of-life standards or human well-being: there is spatial variation of levels of need satisfaction. Depending on the magnitude of the differences in budget constraint among people, place utility can be a major force behind the emergence and persistence of socio-spatial inequality within urban areas.

This is simply because the poor have less freedom in realizing their space preferences than the more well-to-do.

The next question that needs to be addressed is how production is incorporated in spatial welfare theory. Spatial welfare theory incorporates geographical space as a specific consideration for the achievement of efficiency in production. The fact that production is driven by profit implies that the location decision is directly linked to the production costs and the level of expected effective demand that can be satisfied at a particular location. In an optimal space-economy, the spatial arrangement of production would be such that the correct services are produced (in correct quantity at correct price) at the correct locations.

In the real world, however, the spatial arrangement of production will not reach such an optimal state for at least four reasons. *First*, the process of production - like that of consumption - is not fully informed or perfectly rational. *Second*, producers are not free to locate anywhere but are restricted to settle at available locations only. *Third*, imperfect spatial mobility of production factors prevents easy adjustment to spatial changes that occur in the market (Johnston *et al.*, 1998). The *fourth* reason is the existence of spatial externalities. Put simply, an external effect is an unpriced effect. It may be a benefit received by those who do not pay for it (a positive externality) or a loss incurred by someone who is not compensated (a negative externality). A classic example of a negative externality is the air pollution resulting from a localized industrial plant. A positive externality could be proximity to valued services such as schools, parks or health clinics. The examples usefully illustrate that the externality concept has a strong geographical dimension. Activities take place at a specific location in geographic space and the side effects of these activities - whether positive or negative - emanate from these locations and affect the surrounding areas. In most cases, these effects will tend to diminish with increasing distance from the source of the externality (Pinch, 1985).

The main conclusion to be drawn from the discussion presented above is that spatial welfare theory demonstrates that an unregulated economic/spatial system is unlikely to achieve maximum social/spatial welfare unless there is some form of government intervention into the market system, so that either prices are adjusted to reflect full social costs or outputs are controlled to a socially optimal level (Walker, 1981). It is equally important to realize that until now, the discussion has only considered the production and consumption of *private* goods and services; that is goods and services produced according to market exchange principles. In the following section, we further elaborate spatial welfare theory but there the production and consumption of collectively provided *public* services in geographic space is at the centre of the discussion.

Public services, collective consumption and the necessity of spatial policy

Those goods or services that - while valued by individuals - may not be produced if their supply is left to market forces are termed public goods. Three basic characteristics define a pure public good or service (Walker, 1981). The *first* characteristic is that of joint supply: if a good or service can be supplied to one person, it can also be supplied to all other persons at identical quality at no extra cost. The *second* characteristic is non-excludability: once the good or service is supplied to one person, it is impossible to withhold it from others so that those who do not wish to pay for it cannot be prevented form enjoying its benefits. The *third* characteristic is non-rejectability, which means that once a good or service is supplied it must be equally consumed by all (Pinch, 1985). The crucial point here is that the combined characteristics of pure public goods imply that their supply leads to market failure. This can be clarified by distinguishing between the *use value* of goods and services - the extent to which they can satisfy human needs - and their *exchange value* - the prices they can command in the market. The specific characteristic of a pure public good or service is that it has a recognized use value, but because of its characteristics this does not translate into an exchange value. A rational individual wishing to maximize utility will not pay for a public good or service if it can be obtained free. The absence of an exchange value thus prevents a pure public good or service from being produced under market conditions.

In reality, public good or services are actually not consumed and benefited from jointly and equally by all. Because of this they are more appropriately termed impure public goods or services. In the context of public service provision in urban areas, geographical space contributes to impurity by means of variations in access across space. This is explained as follows. Within a defined territory, public services - theoretically equally available to all - have to be located at particular points in space. Even if these services are provided free of charge at the point of supply, individuals will have to bear the cost of travelling to the facility. This means that with a fixed budget of time, money or effort, the amount of a public good or service or the frequency with which it will be consumed will decrease with increasing distance from the facility up to a point where demand for that good or service will become zero (Dicken and Lloyd, 1990).

In this way variation in access across space, generates inequality and undermines both the joint supply and the non-excludability criteria of a public good. In case a service is delivered to the consumer - for example an ambulance service - the costs for supplying the service will increase with distance. The quality of the service - e.g. response time - to the consumer will also vary with distance. In this case, geographic space generates inequality with respect to the joint supply criterion of a pure public good.

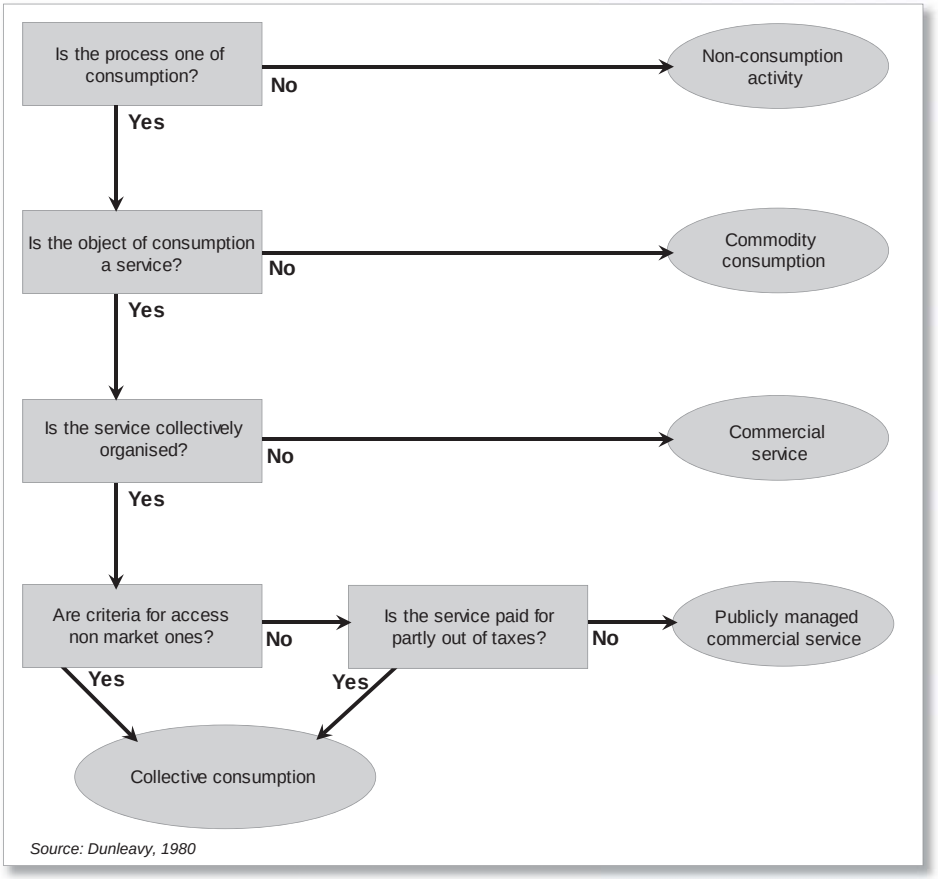

Figure 2.2: Collective consumption defined.

It is important to note that the pure theory is based upon the characteristics of the goods and services themselves and not whether they are produced within the private or the public sector of the economy. In fact, there are few *technical* reasons why services such as education and health care should be allocated by the public sector in preference to the private sector. The decision to allocate many services through the public sector is, not because of their technical characteristics, but because of social objectives translated into political action. A common reason for public provision is dissatisfaction with the inequalities in the distribution of services as allocated through private markets. Governments intervene in producing public services to assist in welfare maximization (Smith, 1995). The provision of public services is usually organized and managed by a state-operated system and financed, at least in part, through taxation. Their overall level of output is related to the general process of public resource allocation in society, which constrains what can be spent on them. Public goods, thus, differ from private goods as they are part of the process of collective consumption (see Figure 2.2).

Apart from a concern with the level of resource allocation, governments will have to devise criteria for the distribution of public resources among the members of society. As mentioned before, the objective of public resource allocation is often to achieve what is considered a fair or socially just distribution. Because of this, collectively provided services are normally consumed by users according to non-market criteria such as social need rather than ability to pay. At the same time, society wishes to obtain the best return for their investment in public resources, so that issues of efficiency and effectiveness are also important.

Since the recipients will be dispersed within a defined territory, there is also a need for a locational decision making process. This transposes the welfare objective to achieving, as much as possible, a spatially just distribution of public resources in a spatially efficient manner. Invariably, this involves making judgments between alternative spatial arrangements of provision in a defined territory. To evaluate existing situations and suggest improvements, however, we need criteria for distinguishing better from worse. This is the subject of the following section.

Evaluating the performance of alternative spatial welfare distributions

A useful starting point for the evaluation of (spatial) distributions is the principle of equality. Equality - like health - is a universal human right and refers to impartiality. A distinction can be made between arithmetic and proportional equality. Arithmetic equality refers to perfect equality of treatment where exactly the same quantity of benefits goes to all irrespective of circumstances. Proportional equality, however, is output oriented: to achieve equality in results may require inequality in resource allocation. It refers to situations where there is equality of treatment in the same circumstances but different circumstances justify differential treatment in the distribution of benefits (Smith, 1994). Proportionality, thus, requires the identification of relevant differences that justify differential allocation of resources.

Efficiency is about comparing costs (resources spent) and benefits (well being produced) of alternative resource allocations in such a way that the welfare gains to society are maximized. It refers to the way of obtaining the highest output from a finite level of resources (Carr-Hill, 1994).

Equity is a more elusive concept and not automatically interchangeable with equality. Equality is about equal shares (inputs or outcomes) while equity refers to fairness and it may be judged fair by society to be unequal. We also need to distinguish between two elements of equity: horizontal and vertical equity. Horizontal equity is concerned with equal treatment for equal need. Vertical equity, however, has a moral and ethical dimension as it refers to the effort to develop a rationale to allocate resources among people who are in unlike circumstances (Truelove, 1993). It will be clear that for vertical equity to exist, the distribution of resources must be seen by society to be fair.

What is regarded as fair and socially just, thus, depends to a large extend on social and political judgments which reflect the prevailing ethical ideology, rather than something with absolute moral authority.

The most influential ideological positions that have a bearing on the distribution of resources are libertarianism and egalitarianism (Smith, 1977). The libertarian perspective relies on the 'economic' justice of free-market competition. Differences in ability result in the more productive people and localities receiving greater benefits. The invisible hand of the market performs social justice in distribution by means of the Pareto criterion. The fact that social well-being will vary across space is accepted as socially just. The argument in favour of Pareto optimality is that it involves minimal ethical content.

Attractive as this may seem, the libertarian line of thought can be criticized on two counts. *First*, the (space) economy will not function in a perfect manner in reality for reasons elaborated before. The *second* and fundamental shortcoming of the Pareto criterion is that it is 'a-historic'. While correctly identifying tastes, prices and the income constraints as the immediate determinants of consumption (in space), the model says nothing about how these conditions originate and how the distribution of goods and bads among members of society is. If the distribution of benefits is unequal to start with, this could become more unequal. In a situation where the rich become richer and the poor no poorer, this would still be considered as a welfare improvement by the Pareto criterion. The application of the Pareto criterion in a non-growth economy, for example, would prevent redistribution in the direction of the poor, no matter how unequal the initial distribution is. This characteristic makes the Pareto criterion a deficient tool for judging alternative (spatial) distributions.

The egalitarian perspective, by contrast, emphasizes the primacy of equality in distribution while efficiency is seen as a secondary goal (Coates, 1977). The main advantage of an egalitarian point of view is that it enables to judge an existing situation in terms of the degree of inequality observed. Furthermore, alternative spatial arrangements can be judged by their tendency towards reducing inequality: the more equal the alternative arrangement the better. In short, the dictum 'the more equal the more just' (Smith, 1994) provides more convincing grounds for evaluation and action than the Pareto criterion.

A well-established concept for evaluating spatial (in-) equality from an egalitarian perspective is 'territorial justice'. Davies (1968) describes territorial justice as follows. If the objective of collectively provided public services is 'to each according to his need' then the most appropriate distribution between areas must be 'to each according to the needs of the population of that area'. Territorial justice is rooted in the principle of proportional equality - provision proportionate to need - and is about securing a geographical match between resource allocation and resource needs. Since the

approach is concerned with relative variations in resource provisions between areas, it can also be usefully applied in contexts where there is an overall shortfall in resource allocation. The concept of need thus is a relative one which should be seen in the context of existing conditions in society in relation to some socially accepted norm.

The empirical identification of variations in need in a territorial distribution requires the development of social indicators. Territorial social indicators refer to measures of social well-being in a defined territory, which may refer to a broad concept or a specific condition such as, for example, health status. To avoid the problems of the ecological fallacy, the level of territorial disaggregation should be at a sufficiently fine resolution. With territorial social indicators spatial variations in well-being can be properly assessed and spatial policies developed to counter identified disparities. Over the past decades, the social indicator approach has become widely accepted as a desirable monitoring and policy development tool (Johnston *et al.*, 1998).

With the concept of territorial justice and the social indicator approach a suitable theoretical framework has been established for describing and evaluating the performance of alternative spatial distributions of collectively provided public services in terms of welfare maximization.

Linking spatial welfare theory to urban health care

Through the discussion on spatial welfare theory a general theoretical framework is formulated. The analogy to health care provision is clear but so far has only been implicit. The subject of this section is to explicitly link spatial welfare theory to urban health care by briefly returning to the question 'who' gets 'what' 'where' and 'how' as illustrated in Figure 2.3. Some comments are also made on health seeking behaviour of patients.

The broader context of urban health care provision

In an ideal society health care systems would be capable of meeting all the health needs of a population. Experience, however, shows that this has proven to be unrealistic and contemporary health planners recognize that the need for care will almost inevitably exceed supply. In the Western world, the expressed need for and cost of health care has continued to grow despite limited population growth, rising living standards and improvements in diet, hygiene and preventative measures. In part, this is driven by changing health needs of an aging population, technological innovations and increasing personnel costs but it is also related to rising expectations in terms of the type and volume of services required. Despite the affluence of Western society, cost containment policies have been put in place since the 1970s in recognition of the fact that the expressed need for health care may actually be insatiable. Even in affluent societies, health care systems will, therefore, satisfy relative rather than absolute health needs.

Although the same essential issues play a role in the debate on health care provision in sub-Saharan Africa, it is important to stress that the context of this discussion is a totally different one. The demand for health care is also growing but the underlying causes are quite unlike those of the Western world. In sub-Saharan Africa, rapid population growth, the urbanization of poverty, malnutrition and unhealthy living conditions are the main causes of morbidity and mortality in the city. Furthermore, the severe budget constraint of the large contingent of urban poor restricts them from expressing sufficient effective demand for essential (private) health care of adequate quality. At the same time, adverse economic conditions restrict national governments to allocate adequate amounts of public resources to health care. Current levels of government expenditure on health care in sub-Saharan Africa are generally far too low to satisfy even modest health needs (Creese *et al.*, 1998).

Key issues in the ongoing world-wide socio-political and ethical debate concern the degree to which health care should be financed through collective means or should be left to market forces. Over the last decade, the welfarist orientation has lost grounds to more libertarian interpretations which promote increased privatization of health care: efficiency considerations currently outweigh equity issues. In most countries in sub-Saharan Africa the outcome of this process is the emergence of pluralist health care, that is, side by side functioning of market driven private health care provision together with collectively provided public health care. The strength of spatial welfare theory is that it provides a unifying framework that can be used to describe the spatial arrangement of private as well as public health provision. Since the interrelations between economic development, urbanization and health care provision in sub-Saharan Africa are all essential issues, they are separately discussed in more detail in Chapter 4.

Demand, supply and spatial distribution of health care in urban areas

Urban areas exhibit variations in the spatial distribution of health needs. Such variations occur as a result of permutations in population densities, socio-economic, demographic and living conditions which all impact on the health status of people. The existence of spatial variations in levels of well being - of which health needs are part - within urban areas closely matches spatial welfare theory as was previously illustrated through the examination of the concept of place utility. Also, cities are not static but evolve over time. If we consider the process of rapid urbanization that is ongoing in sub-Saharan Africa, it will be clear that the spatial distribution of health needs is dynamic. Changes occur not only because of urban expansion and densification processes, but also because of evolving health needs as part of the epidemiological transition.

The fact that health needs will vary across space and time implies that the spatial structure of health care provision has to adapt to - if possible even anticipate - these changes. This makes that the spatial arrangement of health care provision in rapidly

changing urban settings is an issue that requires ongoing attention of health planners. As illustrated in Figure 2.3, the spatial organization of private health care provision is governed by the profit principle - satisfying effective demand in an efficient manner. Public health care provision, by contrast, aims to maximize health needs satisfaction in an equitable manner. Spatial welfare theory demonstrates that this does not only require a process of mobilizing collective resources for health care but also a form of spatial planning for distributing these resources to a spatially dispersed population. The discussion on spatial welfare theory has indicated that territorial justice is a suitable concept for distributing public health resources in geographic space.

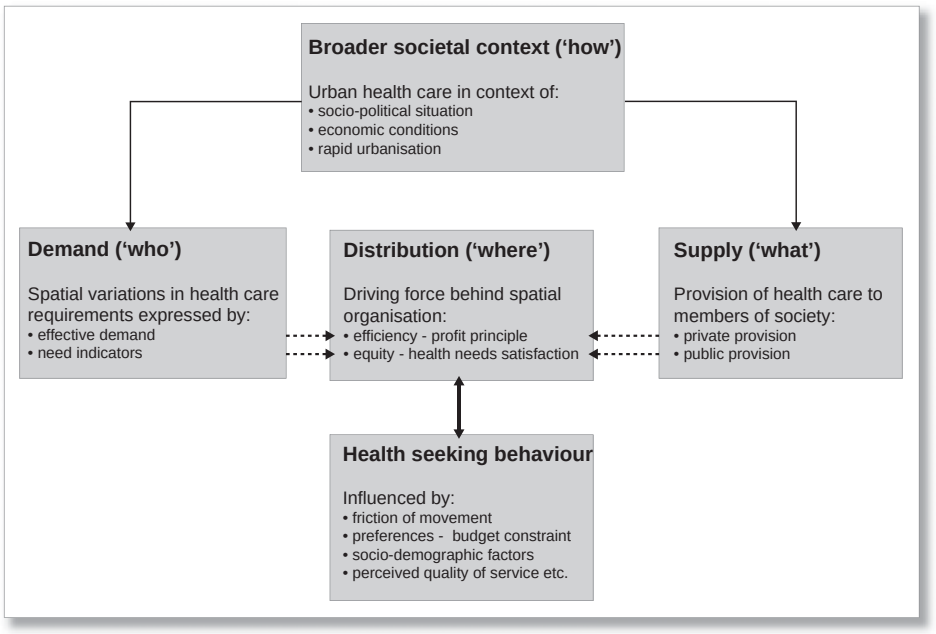

Figure 2.3: A working model of spatial welfare theory in urban health care.

The adoption of the 'health needs' objective as the basic regulatory mechanism for the spatial allocation of public health services is, however, not without problems. This is because health needs are hard to define and measure. In addition to medical criteria, health needs must be understood in the context of the individual and the social, technical, political and economic environment that exists at a particular time. Value judgments will play a role regardless of the definition adopted, and planning outcomes may be affected by the way in which health need is operationalized. As described previously, a way forward is to establish practical indicators of health needs based on empirical evidence. The establishment of spatially disaggregated health indicators in sub-Saharan cities is, however, generally problematic given the paucity of reliable, up to date and disaggregated data (also see Chapter 3).

Health seeking behaviour: access and utilization

An important contribution of spatial welfare is its dismissal of the neoclassical assumptions of equal friction of distance in all directions and of absolute rationality in human (spatial) behaviour. Obviously, the friction of distance is not equal in all directions but largely determined by existing road patterns and restricted by physical barriers to travel. Furthermore, it is not a function of physical distance alone but influenced by the available modes of transportation, travel time, costs, congestion level, and so forth.

Health care utilization - the outcome of the process through which individuals decide which particular medical facility to use - will not occur under conditions of perfect information and will not necessarily be economically and spatially rational (Golledge and Stimson, 1997). Commonly recognized co-determinants of health seeking behaviour include socio-economic status, demographic characteristics, health care costs, type and severity of illness, perceived quality of care and, friction of distance (Joseph and Phillips, 1984; Phillips, 1990).

As illustrated in Figure 2.3 health seeking behaviour can be viewed upon in two ways. Firstly, as the *outcome* of a given spatial arrangement of health care provision. Second, it can be used as the *input* for the planning of health care interventions aimed at increasing health needs satisfaction. Regularities in revealed health seeking behaviour can be identified and subsequently used to identify under-serviced areas and set priorities for health care interventions. Proper health care planning thus requires understanding of the factors that govern actual health seeking behaviour. The strength of such an approach is that it roots health-seeking behaviour in the social and spatial context in which decisions are made. It also implies a move from a normative towards a descriptive planning framework that incorporates actual patterns of spatial behaviour. The analysis of the factors governing health seeking behaviour is the subject of Chapter 5.

The argumentation presented above has shown the strength of spatial welfare theory as a theoretical framework for describing the functioning of urban health care (whether private or public) in geographic space. In the next section we move away from theoretic abstraction and orient the discussion to the methodological framework of this study.

2.3 Methodological framework: accessibility analysis

This section identifies appropriate methods and techniques of spatial analysis that can be used in a GIS environment for the representation and modelling of a spatially equitable and efficient health care system. To appreciate what contemporary GIS-

based spatial analysis has to offer, this section commences with a brief overview of how this subject area has developed over time. Attention then turns towards an exploration of the concept of accessibility, its most common operationalizations and the identification of the most promising type(s) of measure for this study.

Spatial analysis and human geography

The origins of spatial analysis lie in quantitative geography, which developed into an important school of thought in the 1950s and 1960s (Johnston, 1979). The spatial analytical approach was a reaction to the regional paradigm in human geography with its emphasis on the unique and the particular (regional differentiation and regional synthesis). The spatial analytical approach, by contrast, emphasized the importance of developing generalized laws, models and theories of spatial organization within a logical positivist framework. (Berry and Marble, 1968) have described the primary goal of the approach as building accurate generalizations with predictive power by precise quantitative description of spatial distributions, spatial structure and organization, and spatial relationships.

The introduction of quantitative techniques in human geography was an important outcome of the spatial analytical approach. Relevant theories, statistical techniques and mathematical methods were incorporated from other systematic sciences and extended, modified and applied in geographical research (Johnston, 1979, p. 54). Two main fields of study characterize spatial analysis: statistical spatial data analysis and spatial modelling. Much of the spatial modelling work of that time were mathematical interpretations of neo-classical economics, central place theory or classical mechanics. Research efforts emphasized the development of spatial theory and quantitative techniques; the significance of the research problem itself was often secondary (Goodall, 1987).

Up to the early 1970s there was great interest in and widespread use of spatial analytical methods by human geographers. After that, the popularity of model-based approaches declined and during the 1980s spatial analysis became largely forgotten (Rushton, 1993; Maguire, 1995). The disillusion with the spatial analytical approach was caused by two important factors. The first was the recognition that many of the models did not properly represent human behaviour (a lack of credibility) and that their outcomes were not easy to interpret or practical to use (a lack of acceptability). The result was a real paucity of applied models in use by planners in either the public sector or the private sector (Clarke, 1997). The second factor was the changing perspective of planners and decision-makers. There was concern that the mechanical approach of quantitative geography failed to respond to broader issues of equity, pollution as well as the increase in uncertainty posed by global crises of energy and economy (Harris and Batty, 2001). Increased academic interest in understanding how individuals behave and structure their spatial organization, as well as the resurgence of

interest in public policy issues made that behavioural and radical geography came to the forefront in mainstream geography. Starting from the mid-1990s, however, a reappraisal of spatial analytical techniques can be observed, largely as a result of the widespread proliferation of GIS.

GIS and spatial analysis

Whereas spatial analysis has traditionally been the domain of quantitative geography, GIS has emerged from rather different origins. Rooted in computer science and computer graphics, the early GIS systems were primarily developed as tools for the storage, retrieval and display of geographical data. Capabilities for spatial analysis were very limited. Since then, GIS systems have matured considerably and modular multi-purpose packages with broad functionality have come to dominate the market. What remains, is that the vast majority of GIS applications is oriented towards database automation and relatively simple mapping and query operations. Most GISs have been established with the objective to increase the efficiency and effectiveness of the organization.

Starting from the mid-1990s a substantial number of publications illustrate a growing (academic) interest in the possibilities of using spatial models in GIS for planning purposes. A number of developments are responsible for the reappraisal of spatial analysis: technological progress, a maturing community of GIS users and academic advancement in the field of spatial analysis itself (Fotheringham and Rogerson, 1994; Maguire, 1995; Birkin *et al.*, 1996; Longley and Batty, 1996).

Technological progress stimulates spatial analysis in at least three ways. To begin with, the availability and affordability of powerful desktop computers has substantially removed the processing bottleneck that formerly restricted spatial analysis. Furthermore, GIS software provides spatial analysts with a support environment with excellent tools for spatial data management and visualization. Last but not least, - in the Western World - there is generally a much greater availability of digital data that can be utilized as inputs into spatial models. New techniques of data capture -such as remote sensing and Global Positioning Systems- have come within reach of most GIS users and digital (spatial) databases are increasingly seen as a tradable commodity and strategic resource.

A second development concerns the changing requirements of a maturing GIS community. The initial focus on database creation and routine information management is slowly evolving and more and more users also want to utilize GIS for the analysis of their data. Although the demand for such functionality is still restricted to more experienced users and to application areas such as business and service planning, overall demand will further rise in future. This development illustrates the evolution of GIS from a transaction-processing tool towards a decision-support tool,

capable of sophisticated analysis and modelling operations. It also reflects the desire of organizations to gain competitive benefits from the use of GIS (Openshaw and Clarke, 1996).

A third development responsible for the reappraisal of spatial analysis comes from academic advancement in the field of spatial analysis itself. Ongoing re-specification and reformulation of models has resulted in a new generation of analytical models with a much higher degree of credibility and acceptability. There has been a trend away from crudely specified, large-scale models towards data-rich models specifically designed for tightly defined application areas. The deterministic neo-classical foundation - a common characteristic of 'older' models - has been supplanted by probabilistic techniques capable of a more appropriate representation of human spatial behaviour. Not only have existing methods improved, but also a whole range of new methods (e.g. micro simulation, neural networks, and cellular automata) has developed alongside.

The combined impact of these developments is that many basic and some advanced spatial analysis capabilities have become incorporated in commercial GIS software in recent years. Despite this extension, many professional spatial analysts consider the current level of analytical functionality in commercial GIS as deficient and not reflecting the advances made in the last ten years (Openshaw and Clarke, 1996; Hendriks and Ottens, 1997; Fotherinham, 2000). Most of the academic modelling tools continue to be developed as separate modules (mostly research prototypes), which can be coupled to the GIS through a data transfer mechanism of varying sophistication. It is worthwhile noting that the coupling approach has strong similarities with the system architecture of Spatial Decision Support Systems.

What about the future of spatial analysis in GIS? It is very probable that further enrichment of analytical functionality in GIS can be expected as a result of the progression of above described developments. It should be equally clear, however, that spatial analysis is outside the mainstream of GIS applications and will remain so for the foreseeable future (Longley and Batty, 1996). The success of analytical GIS will, therefore, very much depend on its ability to provide real support to decision-making by offering acceptable and robust methods of problem solving to the planning community (Clarke, 1997).

The challenge for spatial analysts is, thus, to further enhance their analytical methods, to support their practical application in planning contexts, and to educate end-users about the benefits and pitfalls of using them. In the next section, focus is on the accessibility concept which is an important field of study within spatial analysis.

Accessibility: a concept and tool for spatial equity analysis

Spatial planning of health care facilities at the intra-urban level concerns the identification of suitable locations for a given number of health facilities in a defined territory, in such a way that the health needs of a spatially dispersed population are served in an optimal way. Basic to this type of location planning is the concept of accessibility. Stated in general terms, accessibility relates to the ability of people to overcome the friction of distance to avail themselves of services at fixed points in space. As argued before (see section 2.2), access to sources of need satisfaction has an important bearing on human well-being: accessibility can be viewed upon as a scarce resource, which is dependent on distribution and redistribution through the planning process (Smith, 1974). Accessibility is used as a mechanism through which we can approximate the degree of health needs satisfaction - and/or predict the expected number of patients - for a potential location.

Although accessibility is a familiar planning objective and a frequently used term in the geographic literature, it is a slippery notion, which has been defined, and operationalized in a variety of ways. We therefore first need a clear insight in the accessibility concept, only then we can investigate how accessibility can best be given operational form in the context of this study.

A very useful framework that introduces the notion of accessibility in the context of service provision is provided by Moseley (1979). Moseley systematizes accessibility as consisting of three components: *(i)* people, *(ii)* the activities or opportunities that they require, and *(iii)* the transport or communication link between the two. The strength of this framework is that it makes explicit that accessibility varies according to the characteristics of people, of the activity or opportunity, and of the transportation infrastructure. Accessibility is the outcome of the combined characteristics of the three components mentioned above. Moseley's framework explicitly positions accessibility as a concept consisting of a socio-economic as well as a spatial dimension between which important relationships exist.

A comparable but more elaborate framework for understanding accessibility is described by Geurs and Ritsema van Eck (2001). They present accessibility as the outcome of four interrelated components: a spatial component, a transport component, a temporal component and an individual component (see Figure 2.4).

The spatial component of accessibility describes two elements: *(i)* the spatial distribution of opportunities (and their characteristics) and *(ii)* the spatial distribution of demand for opportunities (and their characteristics). In general terms, it refers to consumption in space (place utility) and production in space (space utility) as previously described in section 2.2. In terms of health care, it reflects the spatial arrangement of health care provision in relation to the spatial variation of health needs

of people. In the case of private health care the attuning mechanism is effective demand; in the case of public health care it is some indicator of health needs.

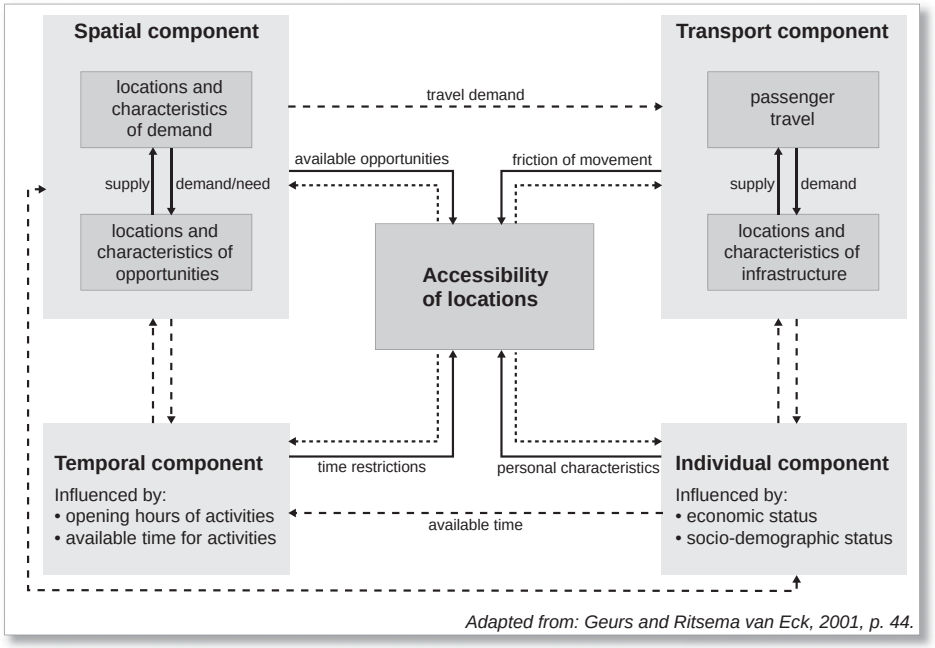

Figure 2.4 : Components of accessibility and their relationships.

The spatial arrangement of opportunities is normally represented by point locations; attractiveness is mostly measured in terms of variations in size and functional make up. Another important consideration concerns the representation of the spatial variation in demand or need. This requires some form of spatial disaggregation: the breaking down of a bounded region into smaller zones, thus grouping people by proximity. For analysis at the intra-urban scale, the definition of spatial units for which demand or need is estimated, can be based upon the boundaries of administrative units (e.g. census tracts, wards). An important motive for using such units is the link to descriptive data (e.g. census data) that can be used to approximate demand or need. A disadvantage is that these units may be arbitrary in size and shape and internally heterogeneous (problem of the ecological fallacy).

In the more recent literature, a trend can be observed towards using smaller geographic units, the use of postcode areas in particular. Another approach is to subdivide a study area into a regular tessellation of some kind (squares, hexagons). The advantage of such an approach is the near elimination of the ecological fallacy (Spiekerman and Wegener, 2000). A disadvantage is the weaker link to secondary data sources. In most cases, the spatial component is measured for a resident who begins a trip to an

opportunity from the place of residence but alternative approaches are described that incorporate multi-stop trips and trip chaining (Arentze *et al.*, 1994).

The transport component consists of three elements *(i)* the location and characteristics of the transport infrastructure, *(ii)* the demand for travel and *(iii)* observed infrastructure use (mobility). Transport itself is not the desired product but a means to an end. The purpose of the transport component is to describe how people overcome the friction of distance that separates the point of demand from the point of supply (Pacione, 1989). Friction of distance is normally described in terms of travel mode, time, costs or effort. It can be estimated by Euclidian distance, some modification of it such as, for example, the Manhattan metric or actual distance/time calculation along a street network (Hanson and Schwab, 1986). The latter is generally considered to give the best approximation (Geertman and Ritsema van Eck, 1995). Generally, travels patterns show some form of distance decay, that is, the interaction between two locations declines with increasing disutility (travel distance, time, cost and effort). The perception of friction of distance is not given but varies with transport mode, purpose of trip, characteristics of the household and characteristics of the attractiveness of the destination.

The temporal component of accessibility describes the *(i)* availability of activities at different moments in time and *(ii)* the times at which individuals are able to participate in activities. An example of the first is that accessibility to health care depends on the opening hours and days of medical facilities. An example of the second occurs if the opening hours of a medical facility coincide with the working hours of an individual. This component shows that accessibility is not merely about making connections in physical space but also needs to take available time into account. The incorporation of the temporal component of accessibility finds its origins in the space-time studies from (Hägerstrand, 1970).

The individual component of accessibility refers to the characteristics of individuals that impact on accessibility. Approaches that incorporate individual characteristics and preferences do exist - and will be touched upon in the following - but their practical application is relatively scarce. By far the most common approach is that the individual component of accessibility is incorporated into accessibility measures by disaggregation of the population under study using economic and socio-demographic characteristics.

Obviously, the four components of accessibility are interrelated: the spatial component is an important factor determining travel patterns; it may also introduce time restrictions and influence people's opportunities. In turn, the individual component influences travel demand, the available time for activities and the demand/need for opportunities. Furthermore, accessibility also influences the contributing components through feedback relationships: accessibility is a location factor for people and

activities and influences travel demand, people's opportunities and the time needed to participate in activities (Geurs and Ritsema van Eck, 2001).

Accessibility measures: an overview and evaluation of approaches

The subject of this section is to give operational form to the concept of accessibility. It shortly describes and evaluates the following measures *(i)* container measures, *(ii)* simple distance measures, *(iii)* cumulative opportunity measures, *(iv)* composite measures, *(v)* time-space approaches and *(vi)* utility-based approaches.

Container measures

The simplest and crudest measure of accessibility - mainly used in the political science literature - is the container approach: access is determined on the basis of a simple count of facilities or services by some administrative area. There is no consideration for the structure of the transportation network, the frictional effect of distance or properties of the supply side (Talen and Anselin, 1998). Although this approach has some merit at higher levels of spatial aggregation - when resources are allocated by the political process to individual administrative units - it is not appropriate for analysis at more detailed geographic scales such as that of intra-urban health care provision. It also does not reflect any of the previously identified components of accessibility.

Simple distance measures

The simplest distance measures operationalize accessibility as the straight line distance between two locations in geographic space. This type of measure is often used in situations where standards exist in terms of maximum travel time or distance to an opportunity (e.g. every person must be able to reach a health facility within 30 minutes travel time). Accessibility is measured from a demand location to a single (the closest) opportunity. This type of approach is based upon the notions of neo-classical location theory. It is very commonly used and mostly operationalized on the basis of standard GIS functionality. In their simplest form, this is done through the generation of service areas around opportunities using functions such as buffers or Thiessen polygons, often in combination with overlay and query procedures (Doherty *et al.*, 1996). An advantage of simple distance measures is that the data requirements are modest and that results can be interpreted easily. This type of approaches also has clear disadvantages. *First*, demand is assigned to the nearest - mostly Euclidian distance - opportunity in a deterministic manner. In other words, the transport, the individual and the temporal components of accessibility are neglected. A *second* problem with this type of approach is that - although it has some merit for evaluating existing situations - it is not very suited for forecasting and scenario-building.

Cumulative opportunity measures

Cumulative opportunity measures indicate the number of opportunities - or number of potential customers - that can be reached within a given travel time or distance

(Breheny, 1978). All potential opportunities within a given cut-off distance/time are weighted equally. Accessibility increases if more opportunities can be reached within a given travel time or distance. This increase can be the result of either a change in the ease of reaching destinations (a change in the transport component) or a change in the number of available opportunities (a change in the spatial component). They incorporate both the spatial and the transport component of accessibility but do not evaluate their combined effect (as is the case in composite measures). The individual component can be approached through socio-economic disaggregation of the population under study.

A well-known research tool from which cumulative opportunity measures are derived is the 'location profile'. A location profile is a graph that for a single location depicts the cumulative number of opportunities for a growing distance range. De Jong and Ritsema van Eck (1997) describe two basic accessibility indicators that are derived from the location profile and used in a GIS environment. The proximity count is the number of opportunities within a given (a fixed maximum acceptable) distance range from a particular starting location. The threshold distance is the shortest distance to a given (a fixed minimum) number of opportunities. On the basis of these measures, a tessellated accessibility surface can be constructed that shows the general pattern of accessibility, or expected number of customers, for a new activity for each location in a bounded region.

The main advantage of cumulative opportunity measures is that their results are straightforward and easy to interpret by non-specialists. Other important advantages are that the data requirements are relatively undemanding and that alternative scenarios can easily be generated. An obvious disadvantage is the lack of differentiation between opportunities that are near to the demand origin and those just within the specified cut-off distance. *Second*, all opportunities are equally desirable despite potential differences in their characteristics. *Third*, the choice of a cut-off travel distance or time (the range) is a somewhat subjective matter. Frequency distributions of travel times or distances from a travel survey can, however, provide relevant information for the determination of an appropriate cut-off value. Another weak point of the proximity count as a measure for potential market size, is that it is based upon deterministic behavioural assumptions: it is assumed that all customers will always visit the nearest opportunity (Ritsema van Eck and de Jong, 1999).

Composite measures

Composite measures of accessibility have a long and rich history and were originally introduced to human geography by the 'social physics' school of thought. At the basis of this type of accessibility measures lies the gravity model, which utilizes the principles of Newtonian physics to explain social phenomena. Briefly, the gravity model stipulates that two places will interact with each other in proportion to the product of their size and inversely according to some function of the distance between them

(Abler *et al.*, 1972, p. 221). The accessibility measure is termed composite as it captures two parameters (the spatial and the transport component) into a single index: *(i)* scale impacts which are operationalized in terms of the utility of the destination location; and *(ii)* distance impacts operationalized in terms of separation units of some kind. The accessibility of a particular location is seen as a function of its relative proximity to alternative destinations of varying utility in a given spatial system.

Because of the probabilistic nature of composite measures they assume non-deterministic fuzzy market areas. Composite measures are mostly used to explain why some localized activities attract more users than do others and to explain the way in which an activity impacts on the surrounding area in terms of customer flows (Haynes and Fotheringham, 1984, p. 11). Generally speaking, the more accessible a location is, the higher the intensity of possible spatial interaction with surrounding locations becomes. Composite accessibility measures describe aggregate human behaviour, not individual behaviour.

Well-known analytical tools for estimating interaction potential are the potential model and the spatial interaction model. Whereas the potential model considers places one at a time with respect to interaction potential with all other locations in the system, spatial interaction models describe the pair wise interaction flows themselves among all locations in the system. When interaction potential is computed for a large number of locations in a bounded region, a map of a potential surface can be constructed.

The main advantage of composite approaches is that they offer a probabilistic view on human spatial behaviour and that their data requirements are modest. Furthermore, some argue that the measure corresponds with a concept that most non-specialists would accept: it denotes a range of choice offered by the land use and transport components in the form of a sum of potential destinations (Geurs and Ritsema van Eck, 2001). Compared to measures of cumulative opportunity, however, the calculated accessibility indicator is harder to interpret. Calculated values have no meaning by themselves and can only be compared to other sites, in the sense that a higher value indicates a more accessible location (Geertman and Ritsema van Eck, 1995; de Jong and Ritsema van Eck, 1997).

The most important disadvantage of gravity-based measures is that they have only limited theoretical justification other than the analogy to a law of physics. Despite numerous efforts (e.g. Wilson, 1970), a link to a theory of human spatial behaviour was never sufficiently established. A second disadvantage is how to reconcile the individual component of accessibility with a model that describes aggregate outcomes. In practice this problem is reduced by breaking down the population under study into socio-economic and demographic strata, allowing each subgroup to have its own characteristics of consumer behaviour. A third disadvantage is that of self-potential: in their basic form composite measures are extremely sensitive to short distances.

Finally, composite measures require the estimation of a distance decay function to describe the friction of distance. For plausible results, the calibration should be estimated using recent empirical data of spatial travel behaviour in the study area. The distance decay function should reflect the characteristics of the opportunity and of demand. Calibration is compounded by two theoretical issues *(i)* the fact that stated/revealed behaviour is not necessarily the same as preferred behaviour, and *(ii)* the extent to which human behaviour itself may change if the spatial arrangement of opportunities changes (Horton and Reynolds, 1971; Breheny, 1978; Rushton, 1993; Golledge and Stimson, 1997). The implication is that travel surveys should pay due attention to the identification of reasons behind stated/observed behaviour (also see Chapter 3).

Time-space approaches

Time-space approaches operationalize accessibility in a rather different manner than those described previously. Time-space approaches were developed, in part, as a reaction to the shortcomings of gravity-based measures and are related to the work of Hägerstrand (1970). At the heart of time geography is the notion that all of the actions and events that sequentially make up an individual's existence have both temporal and spatial attributes. The time-space approach attempts to define a bounded region of time and space at the individual level and to identify the main factors that constrain the individuals' freedom to occupy certain space and time locations. Movement of the individual in the spatial-temporal environment is restricted by three categories of constraints. *Authority constraints* refer to the limitations of access in place and time as enforced on individuals by, for example, authorities. *Capability constraints* refer to limitations of movement because of *(i)* physiological necessities and *(ii)* available transportation means. *Coupling constraints* refer to the 'where, when and for how long' the individual must join other individuals in order to finish a task (Johnston *et al.*, 1998).

By identifying these constraints it would become possible to deduce reasons why a particular individual follows one space-time path rather than another. Concern is with the space-time constraints of individuals and not primarily with the decision process that guides space-time behaviour. The significance of the motivations that underlie spatial behaviour is recognised but these are viewed upon as being elusive and therefore difficult to handle[1] (Golledge and Stimson, 1997).

The opportunities that can be reached given time and space constraints can be seen as a measure of accessibility. Accessibility can be improved either by expanding the action

[1] Geurs and Ritsema van Eck (2001) report on more recent research efforts that relate time-space approaches to the decision-making process by means of the random utility concept.

space of the individual or by placing more opportunities within its bounds. The main difference with cumulative opportunity and composite measures is their explicit link to the individual and the temporal components of accessibility. Time-space accessibility measures are able to reveal differences in individual accessibility, where composite and cumulative opportunity measures are aggregate measures that assign the same level of accessibility to all persons or members of a socio-economic group, that share the same origin location (Geurs and Ritsema van Eck, 2001).

The advantage of a time-space approach is that it accounts for each of the components of the accessibility concept (see figure 2.4). The fact that it is a disaggregate approach also is its main disadvantage as it has rather demanding data requirements to be implemented. Because of the large data requirements, applications are often restricted to smaller study areas and/or subsets of the population.

Utility-based accessibility approaches

Utility-based accessibility approaches assign particular importance to conceptualizing the decision process to explain and model individual spatial choice behaviour. This type of approaches is mostly operationalized using discrete choice models. These models are generally rooted in random utility theory (Thurstone, 1927), in which the probability of an individual making a particular choice depends on the utility of that choice relative to the utility of all choices. The expected choice behaviour is captured using a utility function that reflects the attractiveness of the destination, the travel impedance, the socio-economic characteristics of the individual, and their tastes and preferences (Borgers and Timmermans, 1993). Discrete choice models sometimes take a form similar to that of gravity-based measures, but given their solid theoretical foundation, they have an advantage over gravity-based measures (Borgers and Witlox, 1997; Handy and Niemeier, 1997).

The main advantage of utility-based approaches is their solid theoretical foundation in economic theory and that they encompass all components of accessibility. The disaggregate approach, again is also an important disadvantage as data collection is a very demanding task. For every individual one needs to record what his/her perception is of all the available alternatives and how they are weighted. The conceptualization of the decision making process requires considerable interaction with respondents. Because of the large data requirements, utility-based applications are often restricted to subsets of the population and smaller study areas. Another disadvantage is that the outcomes of a utility-based measure are not easily interpreted by non-specialists (Geurs and Ritsema van Eck, 2001).

Identification of suitable measures for urban health care planning

The question that remains is which of these measures is the most appropriate for urban health care planning in sub-Saharan Africa. This question is answered by looking at factors such as *(i)* theoretical and methodological soundness, *(ii)* ease of

interpretation, *(iii)* data needs, and *(iv)* usability for urban health care planning in the sub-Saharan context.

We begin with excluding container and simple distance approaches for two reasons. *First*, their theoretical soundness is too limited (they neglect the transport, the individual and the temporal components of accessibility). *Second*, these measures are not suited for forecasting and scenario-building which is an essential consideration for location planning in the urban health sector.

Despite their theoretical elegance, time-space and utility-based measures do not appear to be the most appropriate for this type of study for two reasons. The first reason is practical and refers to the difficulty in obtaining data. The conceptualization of the decision making - in the case of utility measures - process requires considerable interaction with respondents. In practice it will be close to impossible to collect such data in the field. The second reason is conceptual. Since this study sets out to establish a socio-economic typology of patients and model their expected spatial choice behaviour - a more aggregate approach -, it should not rely on a theoretical framework that focuses on individual level explanation.

Gravity-based operationalizations of accessibility feature prominently in the empirical literature. Despite their theoretical limitations, composite approaches are successfully applied to describe more aggregate spatial behaviour and have undemanding data requirements. An important advantage is their probabilistic view on human spatial behaviour. Composite measures should therefore not be excluded and are considered to be a useful option for this study.

Cumulative opportunity measures are also suitable for a number of reasons. *First*, their more aggregate nature matches the proposed level of analysis for this study. *Second*, the interpretation of the performance measure of accessibility is straight forward which is an important advantage over gravity-based methods. *Third*, the data requirements of comparative approaches are comparable to gravity approaches. *Fourth*, cumulative opportunity measures can be usefully applied in situations in which norms exist with respect to maximum travel distance or minimum threshold (this is often the case in health care). *Fifth*, advanced comparative measures have recently been developed that take spatial competition effects into account (Ritsema van Eck and de Jong, 1999). Accessibility surfaces can be constructed for situations where a new facility is planned as part of a chain of facilities and the objective is not to maximize the market area of the individual facility alone but rather that of the chain as a whole. This clearly is a crucially important consideration in planning for new public health care facilities. To summarize the above, it appears that the most sound way to represent and model a spatially equitable and efficient urban health care system in the sub-Saharan context, is to use a 'what if' type of GIS-based approach on the basis of a composite or a cumulative opportunity measure of accessibility.

2.4 Contextual framework: GIS based accessibility analysis in sub-Saharan Africa

The former paragraphs have established that there is a need for systematic location planning in the public health sector, and that accessibility analysis provides a useful conceptual and methodological toolkit to support such planning efforts. In this section, focus is on *(i)* the empirical literature on accessibility analysis for (health) services, and *(ii)* the applicability of 'what if' accessibility analysis in a GIS environment in Sub-Saharan Africa.

Accessibility analysis in (health) service planning: the empirical literature

The empirical literature on location planning for services has a long history and is overwhelmingly vast and varied. It covers a host of application areas which range from health care (Wellings, 1983; Oppong and Hodgson, 1994; Bullen *et al.*, 1996; Martin *et al.*, 2002), education (Pacione, 1989; Densham and Rushton, 1996), day-care centres (Truelove, 1993) and emergency services (Ritsema van Eck, 1993; Peters and Brent Hall, 1999) to retailing outlets (de Jong and Krygsman, 1999), parks (Erkip, 1997; Nicholls, 1999), libraries and public playgrounds (Talen and Anselin, 1998).

Common to all is that the concept of accessibility - and some operational form of it - is a central element in the analysis. Despite this similarity, the range of methodological approaches and application contexts is very diverse. A structured overview of a selection of academic publications in this field is presented in Table 2.1. The overview is certainly non-exhaustive and skewed towards health service applications but a number of qualitative conclusions can, nevertheless, be inferred from it.

Accessibility analysis becomes GIS based

Table 2.1, clearly shows that the period up to 1990 can be considered as the pre-GIS era. This is followed by a period of transition from 1990 - 1995 in which GIS makes its entry in the empirical field of accessibility analysis. After 1995, accessibility analysis is predominantly conducted in a GIS environment, regardless of the field of application. Obviously, the emergence of GIS as a tool in empirical service planning closely correlates with the general uptake of GIS technology, and with the earlier discussed rediscovery of spatial analysis.

Table 2.1 also shows how GIS impacts on the way in which accessibility analysis is operationalized. In the pre-GIS era, spatial disaggregation is predominantly based upon some form of administrative zone; the friction of distance is generally calculated in terms of Euclidian distance. With the uptake of GIS, spatial disaggregation is increasingly tessellation based (although zones remain used); and network calculations become the dominant way of representing the friction of distance.

Author	Year	Approach	Measure	Spatial disaggregtion	Socio-economic disaggregtion	Friction of movement	Utility of opportunity	Focus	Country	GIS	Service
Truelove	1993	'What is'	Cumulative	Census tract, Enumeration area	Income, age	Euclidian	Yes	Urban	Canada	No	Day care center
Densham and Rushton	1996	Location allocation	Distance	Education district	Population, children	Euclidean + correction	No	Regional	United States	Yes	Education
Pacione	1989	'What if'	Composite	Postcode zone	Population, age	Euclidian	No	Urban	Great Britain	No	Education
Breheny	1978	'What is'	Cumulative	Census zone	Population, children	Travel cost	Yes	Urban	Great Britain	No	Education
Fortney	1996	Location allocation	Distance	Census tract	Population, children	Euclidian	Yes	Urban	United States	No	Education
Brent Hall et al.	1995	'What is'	Composite	District	Population, fertile females	Network	No	Regional	Costa Rica	Yes	Health service
Secondini et al.	1996	'What if'	Composite	Census tract	Population, size	Euclidean	Yes	Regional	Italy	Yes	Health service
de Jong et al	1991	'What if'	Cumulative	Tessellation	Population, size	Network	Yes	Urban	Canada	Yes	Health service
de Jong and Amer	2002	'What if'	Cumulative	Tessellation	No	Network	No	Urban	Tanzania	Yes	Health service
Martin et al.	2002	'What is'	Cumulative	Tessellation	Population, size	Network	No	Regional	United Kingdom	Yes	Health service
Marks et al.	1992	'What is'	Distance	Zipcode zone	Population, elderly	Euclidian	Yes	Regional	United States	Yes	Health service
Walsh et al.	1995	'What is'	Distance	Zipcode zone	Population, size	Network	Yes	Regional	United States	Yes	Health service
Bullen	1996	'What is'	Distance	Parish boundary Urban ward	Population, size	Euclidian	No	Regional	Great Britain	Yes	Health service
Doherty et al	1996	'What is'	Distance	No	Population, size	Euclidean + correction	Yes	Urban	South Africa	Yes	Health service
Hyndman et al	1997	'What is'	Distance	Census district	Population, age cohort	Network	Yes	Regional	Australia	Yes	Health service
Peters and Brent Hall	1999	'What is'	Distance	Tessellation	Emergency caller Population, size	Network	No	Regional	Canada	Yes	Health service
Perry and Gesler	2000	'What is'	Distance	Village (Point)	Population, size	Euclidian + Travel time	No	Regional	Bolivia	Yes	Health service
Hyndman et al.	2001	'What is'	Distance	Census district	Population, disadvantaged	Euclidian	Yes	Urban	Australia	Yes	Health service
Mallick and Routray	2001	'What is'	Distance	Village (Point)	No	Euclidean	Yes	Regional	India	Yes	Health service
Lovett et al.	2002	'What is'	Distance	Parish/Ward	Population, size	Network	No	Regional	United Kingdom	Yes	Health service
Parker and Campbell	1998	'What is'	Distances	Postcode zone	Patient number	Euclidian Network	No	Regional	Great Britain	Yes	Health service
Khan	1992	'What is'	Composite	SMSA	Population, size	Euclidean	No	Regional	United States	No	Health service
Wellings	1983	'What if'	Cumulative	Tessellation	Population, size	Euclidian	Yes	Regional	Lesotho	No	Health service
Massam et al.	1986	Location allocation	Distance	Village (Point)	Population, size	Euclidian	Yes	Regional	Zambia	No	Health service
Annis	1981	'What is'	Distance	Village (Point)	Population, size	Euclidean + correction	Yes	Regional	Guatemala	No	Health service
Knox	1982	'What is'	Distance Cumulative Composite	Census tract	Population, size	Network	No	Urban	United States	No	Health service
Ayeni et al.	1987	Location allocation	Distance	Village (Point)	Population, size	Euclidian	No	Regional	Nigeria	No	Health service
Hodgson	1988	Location allocation	Distance	County	Population, size	Euclidian	Yes	Regional	India	No	Health service
Mulvihill	1979	Location allocation	Distance	Home location of individual (Point)	Population, poor	Euclidean	Yes	Urban	Guatemala	No	Health service
Oppong and Hodgson	1994	Location allocation	Distance	Village (Point)	Population, size	Euclidian	Yes	Regional	Ghana	No	Health service
Geertman	1995	'What if'	Composite	Tessellation	Population, size	Network	-	Regional	Netherlands	Yes	Housing
de Jong and Ritsema van Eck	1997	'What if'	Cumulative	Tessellation	Population, size	Network	Yes	Regional	Netherlands	Yes	Jobs
Talen and Anselin	1998	'What is'	Container Composite	Census tract	Population, size	Network	No	Urban	United States	No	Playground
White et al	1997	'What is'	Distance	Enumeration district	Population, size	Euclidian	No	Regional	United Kingdom	Yes	Post office
Nicholls	1999	'What is'	Distance	Census unit	-	Euclidian Network	Yes	Urban	United States	Yes	Public park
Erkip	1997	Container	Container	Districts	Population, low income	Travel time	Yes	Urban	Turkey	No	Public parks

Author	Year	Approach	Measure	Spatial disaggregation	Socio-economic disaggregation	Friction of movement	Utility of opportunity	Focus	Country	GIS	Service
de Jong and Ritsema van Eck	1997	'What if'	Cumulative	Tessellation	Population, size	Network	Yes	Urban	Netherlands	Yes	Retail service
de Jong and Krygsman	1999	'What if'	Cumulative	Tessellation	Population, size	Network	Yes	Urban	South Africa	Yes	Retail service
Ritsema van Eck and de Jong	1999	'What if'	Cumulative, Composite	Tessellation	Population, density	Network	No	Urban	Netherlands	Yes	Retail service
Makin et al.	1997	'What if'	Time space	Home location of individual (Point)	Individual	Network	Yes	Simulated	Simulated	Yes	Retail service
Huff	1963	'What is'	Composite	Statistical unit	Population, size	Travel time	Yes	Urban	United Kingdom	No	Retail service
Hanson and Schwab	1986	'What is'	Cumulative	Home location of individual (Point)	No	Euclidian	No	Urban	Sweden	No	Retail service
Handy and Niemeier	1997	'What is'	Time space Composite	Neighborhood	No	Network	Yes	Urban	United States	No	Retail service

Table 2.1: Overview and characteristics of the empirical literature on accessibility analysis for (health) services.

In the most recent literature, separate network calculations are made to accommodate for both public and private modes of transport (e.g. Martin *et al.* 2002; de Jong and Amer, 2002). The individual component of accessibility was and remains mostly incorporated by disaggregation of the population under study using economic and socio-demographic characteristics. The utility of opportunities is incorporated in nearly all studies.

GIS based accessibility analysis for health services

When looking at the type of accessibility measure used in health service analysis, a number of observations can be made. *First*, most GIS operationalizations in the health sector are based upon Euclidian distance measures. The analysis is focused on the existing situation; there is little attempt to develop scenario's that can be used by decision makers to evaluate the effects of alternative planning interventions. Cumulative opportunity and composite measures have such 'what if' capability but they appear much less frequently; time space approaches have not been encountered in the empirical health sector literature (probably this is related to the difficulty in obtaining proper data).

The emphasis on 'simple' distance measures can be interpreted as an indication of the relative immaturity of the average GIS user in the health field (likely this is combined with a limited conception of the accessibility concept itself). The more 'advanced' measures of accessibility are used by specialized academics - geographers and GIS practitioners - with a traceable history in the field of spatial analysis and GIS. The latter category of applications is not focused on the analysis of existing situations but more oriented towards 'what if' scenario development (Higgs and Gould, 2001). These types of applications normally make use of a specialized modelling tool, which is coupled with the GIS.

Second, most publications are studies of potential accessibility; the use of GIS in studies that incorporate actual health care utilization are relatively rare (Gatrell and Senior, 1999).

Third, if we take a geographic perspective, it is clear that most GIS applications in the health field come from the developed world. Their focus is predominantly at the regional/rural scale (although intra-urban applications do exist). If we look at the developing world, it is very apparent that applications with an intra-urban focus - apart from publications by Doherty *et al.* (1996) and de Jong and Amer (2002) - are very scarce. The conclusion that can be drawn from this is that there remains considerable scope for further research in GIS based accessibility analysis in the *urban* health sector in the developing world. The applicability of such a type of analysis in Sub-Saharan Africa is shortly commented upon in the following pages.

Optimal versus acceptable spatial solutions

Before doing so, however, a few words need to be said about approaches that use optimization techniques and those approaches that focus on providing 'acceptable' solutions to planners and decision makers. A very large literature exists on the use of optimization methods in location planning that comes under the location-allocation umbrella (Bach, 1980; Beaumont, 1980; Leonardi, 1981; Ghosh and Rushton, 1987). Location-allocation models jointly optimize the location of centres and the allocation of consumers to those centres. This type of optimization problems requires the specification of an objective function (e.g. the maximization of consumer benefits) and a set of constraints such as minimum levels of supply. The approach demands finding a spatial solution that maximizes the objective function while, at the same time, not infringing the constraints. Hodgart (1978) and Hanssen *et al.* (1987) provide good general reviews of this family of models.

For an overview of location-allocation approaches for health service planning in the developing world - mostly dating back to the 1970s and 1980s - see Rahman and Smith (2000) and Rushton (1984). From Table 2.1, it can be seen that location-allocation approaches were quite popular in health planning in the developing world during the pre-GIS era. Nearly all of these location-allocation applications are concerned with the provision of regional/rural services. This is a clear reflection of the rural orientation of development thinking of that period. GIS-based location-allocation approaches are surprisingly scarce, despite the fact that it has become part of standard GIS functionality. Given that the literature on location-allocation approaches is still 'alive', it appears that proponents of optimization approaches hardly cross-fertilize with academic GIS practitioners.

The principal arguments advanced by proponents of optimization methods is, that the possible range of solutions in geographic space is such that only computer-based search procedures can evaluate all possibilities and suggest optimal policy interventions

(Hodgart, 1978; Oppong and Hodgson, 1994). A number of arguments can, nevertheless, be formulated that provide counter evidence to this viewpoint. Firstly, one can argue that in a decision support environment one should not look for one unique optimum solution: for a decision maker it is more useful to work with a set of acceptable locations as a basis for evaluating several 'what if' scenarios. Secondly, accessibility is rarely the only factor that determines the location decision, restrictions such as, for example, the availability of sites will inevitably exist in reality. Thirdly, by providing planners with 'what if' tools, the congruity between the problem as seen by the decision maker and the modelling tool improves. It allows decision makers to generate their own solutions to problems and examine in more detail the consequences of their decisions (Densham and Rushton, 1996; Jankowski and Ewart, 1996; Miller, 1996). From the above, it is concluded that for this research the need is for a 'what if' scenario type of approach rather than using a restrictive global optimization solution.

Applicability of GIS based accessibility analysis in sub-Saharan Africa

The former paragraph established that GIS-based accessibility analysis offers a useful toolkit that can be used to support location planning in the urban health sector. Furthermore, there is abundant literature pointing to the value of GIS for health care planning within research and practical applications (Gatrell and Loytonen, 1998; Gatrell and Senior, 1999; Albert *et al.*, 2000; Maheswaran and Craglia, 2004). At the same time, there is little evidence, certainly in the published literature, that GIS is being used in strategic decision-making. Practical applications - especially in urban health care - in sub-Saharan Africa are currently virtually non-existent.

This raises the question in how far contemporary accessibility analysis is applicable in African (urban) health planning practice of today. Obviously, this question needs further evaluation upon completion of the study but a few preliminary comments can be made at this stage.

To begin with, the discussion on the applicability of spatial analysis in Developing Countries is not new. Rushton's (1984) review indicates that after a promising start in the 1970s, the application of spatial analytical techniques (mostly rural location-allocation applications) in developing countries generally became rejected in the 1980s by planners and the major development agencies such as the World Bank, Unesco, the United Nations and USAID. In stead, many institutions embraced the 'urban functions in rural development methodology' to bring urban service functions to rural populations (Rondinelli, 1984). This approach emphasizes provision of services as an integrated package based on an analysis of settlement hierarchy. It is qualitative and stresses the use of easily available data and of analytical methods that do not require sophisticated technical skills or computing facilities (Tewari, 1992).

Spatial analysis was dismissed on the grounds of being too sophisticated for local circumstances given the problems of *(i)* availability of computing resources, *(ii)* the lack of trained personnel, *(iii)* limitations of the methods to capture the complexity of the planning problem, *(iv)* comprehensibility to decision makers, and *(v)* data availability (Rushton, 1984, p. 161). If we contrast these arguments with the previously described forces behind the recent reappraisal of spatial analysis in the Western World (technological progress, a maturing GIS community, increasing availability of digital data, and academic advancement in spatial analysis) a few observations can be made.

Computing resources no longer form a significant barrier for spatial analysis

Hardware and software barriers to adoption of spatial analysis have much reduced since the 1970s. With the advent of affordable desktop computers, hardware is no longer a serious bottleneck. Adequate computing facilities have rapidly become available in academic institutes and in public institutions in many African countries. Software also is much less a problem. There is considerable evidence of the adoption of GIS in these institutions in recent years, which means that standard GIS functions for spatial analysis are now generally available. More advanced spatial modelling tools, however, are much less widespread as that they are still mostly research prototypes which have to be coupled to GIS. Birkin *et al.* (1996) come to the conclusion that it is only by linking such spatial modelling routines that GIS can play a more proactive role in the health care planning process. In short, the current hard- and software environment is much more conducive for spatial analysis than some decades ago although advanced modelling tools are not generally available in sub-Saharan Africa.

Scarcity of trained personnel remains

An important difference with the Western World is that the adoption of IT and GIS in sub-Saharan Africa is more recent and many (G) IT projects are still in their initial stages. Given this situation, experienced local GIS practitioners with expertise in the use of spatial analytical techniques will still be scarce. Foley (2002) reports that health authorities have made limited use of GIS because of a lack of IT expertise amongst staff, a limited appreciation of spatial aspects of planning, limited resources and the complexity of the planning process itself. Although Foley describes the situation in the United Kingdom there is every reason to believe that, in this respect, the situation in sub-Saharan Africa is even less favourable. The implication is that close collaboration is needed between GIS academics with expertise in spatial analysis and health professionals to advance strategic applications of health based GIS work in this part of the world. Apart from collaboration, training can remedy this situation. This requires that spatial analysis become incorporated in GIS education aimed at professionals involved in location planning of urban services.

Limitations of analytical methods have substantially reduced

As discussed previously, a common problem with spatial techniques developed during the era of quantitative geography was that the models did not represent human

behaviour sufficiently well. Since then, however, a new generation of analytical models has been developed with a much higher degree of credibility. There has been a trend towards data-rich models designed for specific application areas. The deterministic neo-classical foundation has been supplanted by probabilistic techniques capable of a more appropriate representation of human spatial behaviour.

Comprehensibility to decision-makers has improved
Another important shortcoming of the early quantitative approaches was that decision makers did not generally accept them as a useful tool to work with. In many cases this was caused by a mismatch between the formulation of the problem by the analyst and the problem as seen by the decision maker. Decision makers desired an analysis system to help them generate alternative 'what if' scenarios, analysts tended to formulate the location problem as an optimization problem. As illustrated during the discussion on accessibility measures, current approaches show much more recognition for the desires of decision-makers. The emphasis on optimization has been replaced with 'what if' approaches that aim at identification of good though possibly sub optimal solutions. Finally, with the excellent visualization capabilities of GIS, the interpretability of the access measures has also improved.

Data availability remains an important bottleneck
The availability of secondary data that can be utilized as inputs into spatial models remains extremely scarce in sub-Saharan Africa (although some improvement can be expected with the advent of high resolution imagery such as IKONOS). Often quoted data problems for analytical planning applications in the developing world pertain to the lack of disaggregated data, differences in classification of data, variations in frequency of updating, data inconsistency in time and aerial unit and the general lack of socio-economic data at a useful spatial resolution (Yeh, 1991). Given the dynamic nature of urban areas, the shortage of resources for routine data collection and the initial stages of G(IT) adoption, the availability of up to date and reliable data is a major problem - also see Chapter 3 -, which is not likely to change in the short term.

In conclusion, the applicability of spatial-analytic GI techniques in sub-Saharan Africa is in a better position than it was a few decades ago. Data availability remains a considerable bottleneck. The scarcity of trained personnel remains but can be remedied through collaborative approaches between health professionals and experienced GIS practitioners. On the positive side, the computing environment has much improved and current analytical tools are more robust and more geared to the needs of the decision maker.

Ultimately, however, the success of GIS-based accessibility analysis in sub-Saharan Africa will depend on its ability to provide real support to decision-making by offering acceptable and robust methods of problem solving to the health planning community.

2.5 Conceptual model and research design

In this final paragraph the different threads of the previous discussion are brought together in the general conceptual model and research design of this study.

General conceptual model

Figure 2.5 illustrates how the theoretical, methodological and contextual framework link into the general conceptual model. The figure shows the imperatives of public health care which were the starting point of the analysis of this chapter.

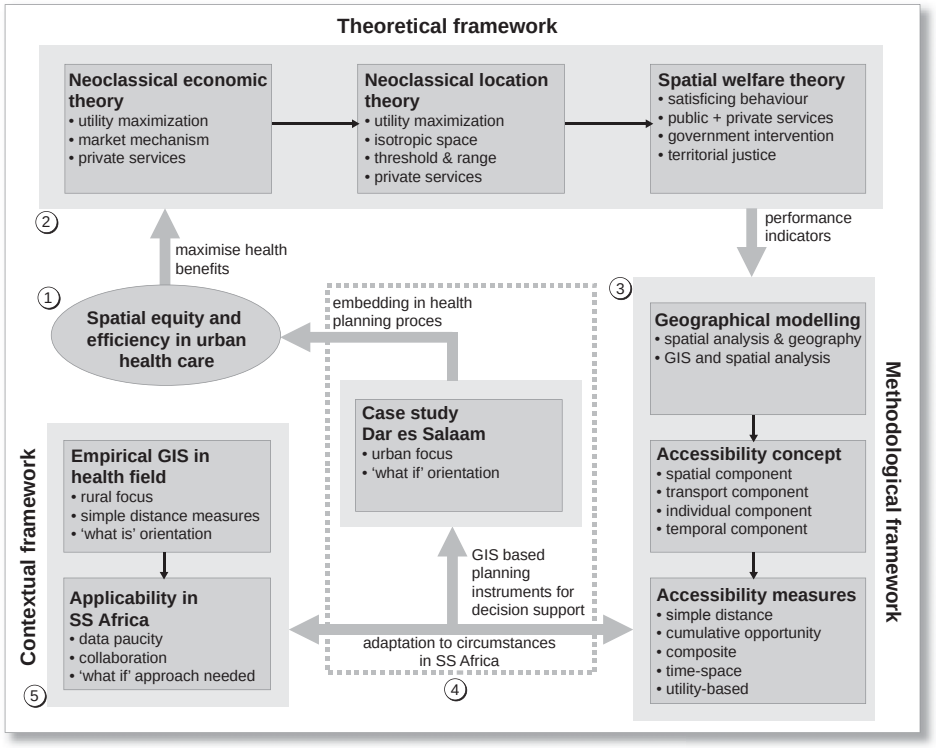

Figure 2.5: General conceptual model.

The link to economics is logical as it is the main discipline concerned with welfare maximization. The economic construct of utility maximization is transposed into geographical space through the discussion on neoclassical location theory. The added value of spatial welfare theory is not only that it provides a broader and less rigid framework than neoclassical location theory, but also that it incorporates the functioning of public services. Territorial justice is the guiding principle of distributing public services in geographic space; its evaluation requires the establishment of spatial performance indicators. Spatial performance indicators, in turn, can be described using the accessibility concept and operationalized with accessibility measures of varying

complexity. The function of the contextual framework is to contrast methodological advancement with real world applications. The first outcome is that most health care applications are in early stages of development. They normally have a rural orientation, are based upon straightforward distance measures. Focus is on the analysis of existing situations rather than on scenario development. The second outcome is that there is a (need for but) near complete absence of 'what if' type of urban health care applications that can support decision-makers in sub-Saharan Africa. This study aims to partly fill that gap.

This is done by structuring the remainder of the research around an empirical case study of the city of Dar es Salaam in Tanzania. The assumption is that Dar es Salaam is a more or less typical example of a large city in the sub-Saharan region. This means that the overarching aim of the study - to develop a GIS based planning approach that contributes to equitable and efficient provision of urban health services in sub-Saharan cities - remains intact.

Research design

Following from the general conceptual model, a number of research stages can be defined for the case study (see Figure 2.6). To begin with, a description of the broader context of urban health care provision in Dar es Salaam is needed. This is done by a closer analysis of the elements 'urban' and 'health care'. The 'urban' element consists of a description of the development process over time and current spatial structure of Dar es Salaam. The 'health care' element is addressed by providing an overview of the dynamics of the health system in Tanzania and Dar es Salaam over time. Taken together, this covers the 'how' element of spatial welfare theory. The second stage relates to the various data sources needed to implement the Dar es Salaam case study and to the setup of the conceptual data model.

As illustrated in Figure 2.6, secondary data needed to be compiled and complemented with primary data, which were collected during field surveys in 1995 and in 2000. Both secondary and primary data feed into the conceptual data model, which is set up to match the previously identified components of accessibility. This setup covers the 'who' gets 'what' and 'where' of spatial welfare theory and enables proper operationalization of accessibility measures during the analysis stage.

The third stage is the conceptual model for the analysis, which consists of three phases. The first phase is the analysis of socio-spatial health seeking behaviour (revealed as well as stated). Its aim is to obtain insight in the main user groups of (public) health care and isolate the main factors that determine health seeking behaviour across different socio-economic strata. Identified behavioural regularities will be used as input in subsequent stages of the analysis. The second phase is to evaluate existing accessibility levels to governmental health care in Dar es Salaam.

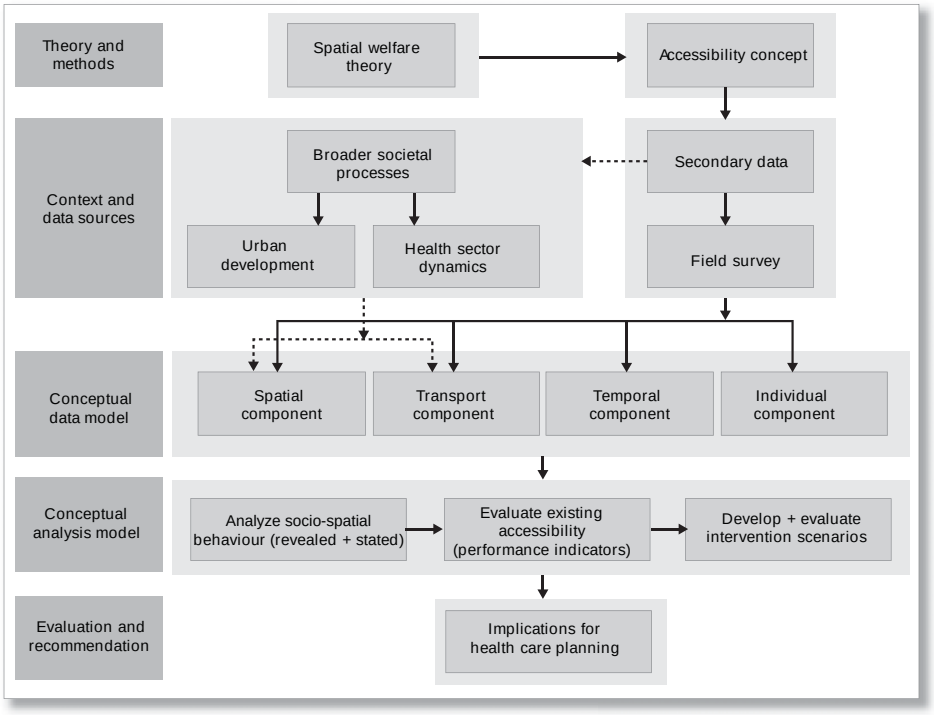

Figure 2.6: Research design of case study.

The third phase is to develop alternative scenarios of the 'what if' type that improve overall accessibility levels.

The fourth and final stage of the research is to reflect on the outcomes of the case study work, and comment on the applicability of GIS-based analytical techniques for strategic urban health care planning in sub-Saharan Africa.