UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI "M.FANNO"

CORSO DI LAUREA MAGISTRALE IN ECONOMICS AND FINANCE

TESI DI LAUREA

"EVALUATION OF A DIAGNOSTIC TEST: THE NEPHROCHECK® TEST FOR THE EARLY DETECTION OF ACUTE KIDNEY INJURY"

Relatore:

CH.MO PROF. ENRICO RETTORE

Correlatore esterno:

CH.MO PROF. CLAUDIO RONCO

Laureanda: CHIARA LEVANTE

Matricola N. 1081926

ANNO ACCADEMICO 2015 – 2016

Il candidato dichiara che il presente lavoro è originale e non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere.
Il candidato dichiara altresì che tutti i materiali utilizzati durante la preparazione dell'elaborato sono stati indicati nel testo e nella sezione "Riferimenti bibliografici" e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo alla pubblicazione originale.

Firma dello studente

_____

ABSTRACT

Economics is a science that finds broad application in health care. By means of assessing benefits and costs associated with a health care program or the introduction of a new health technology, it represents a powerful tool in the decision-making process of allocating scarce resources in order to improve health outcomes. Acute Kidney Injury (AKI) is a frequent and severe complication affecting many hospitalized patients after cardiac surgery. As described in this work, AKI has a negative impact on short- and long-term clinical outcomes, as well as on health care costs. Prevention and early detection of the disease may help avoiding negative outcomes and AKI-associated costs. The present dissertation focuses on the evaluation of the NephroCheck® Test as a diagnostic tool for the early detection of AKI in a population of patients undergoing cardiac surgery. Specifically, this work assesses the ability of the NephroCheck® Test to predict the probability of developing AKI after cardiac surgery and evaluates its accuracy as a diagnostic test.

SOMMARIO

L'economia è una scienza che trova ampia applicazione nel settore della sanità. Valutando i benefici ed i costi associati ad un programma sanitario ovvero all'introduzione di una nuova tecnologia sanitaria, essa rappresenta un importante strumento all'interno del processo decisionale per l'allocazione di risorse scarse finalizzato al miglioramento degli outcome clinici. Il danno renale acuto (AKI, Acute Kidney Injury) è una complicanza frequente e grave, che affligge molti pazienti ospedalieri successivamente ad intervento cardio-chirurgico. Come descritto in questo lavoro, AKI esercita un impatto negativo sugli outcome clinici di breve e di lungo termine, così come sui costi del sistema sanitario. La prevenzione e l'individuazione precoce del danno possono aiutare ad evitare tali outcome negativi ed i costi associati ad AKI. Questa dissertazione è incentrata sulla valutazione del NephroCheck® Test quale strumento per la diagnosi precoce del danno renale acuto, in una popolazione di pazienti sottoposti ad intervento cardio-chirurgico. In particolare, il presente studio VALUTA l'abilità del NephroCheck® Test di predire la probabilità di sviluppare AKI successivamente ad un'operazione cardio-chirurgica e ne stima l'accuratezza quale test diagnostico.

INDEX

INTRODUCTION

The increasing cost of health and medical care services requires the evaluation of new therapies, interventions, devices, medical technologies and diagnostic instruments, in order to ensure that the limited resources are directed towards those programs that are expected to deliver the highest outcome. This assessment process covers multiple interrelated aspects. It is not limited to the benefits and costs associated to the introduction of a new health care program or technology, but it must include an evaluation of the clinical efficacy of the intervention and its feasibility in clinical practice, as well as considerations about longer-term impacts. It turns out that the evaluation of health care programs is a multidisciplinary process, in which different disciplines contribute and must be integrated together. Clinical issues are likely to have relevant economic consequences as well as organizational and social impacts that must be accounted for when it comes to evaluate an intervention in health care.

Acute Kidney Injury (AKI) is an increasingly common and severe complication affecting hospitalized patients that implies damage to the kidney. It is particularly common following cardiac surgery and in intensive care unit (ICU), where incidence can reach 40% and 25%, respectively (De Smedt et al. 2012). It is associated with adverse clinical outcomes, including increased morbidity and in-hospital mortality, longer hospital stay, additional treatments needed. AKI has also a long-term impact on patient's clinical outcomes, with increased mortality, possible progression to chronic kidney disease (CKD) and resulting negative economic effects (Mariscalco et al. 2011, and De Smedt et al. 2012). The annual cost of AKI-related inpatient health care in England has been estimated around £1,2 billion by Kerr et al. (2014).

AKI happens fast and, unfortunately, there is still no specific treatment designed for it. Nevertheless, early detection of the injury may facilitate management and quicken identification and initiation of an effective therapy. However, current diagnostic methods – based on serum creatinine (SCr) as a marker of kidney function and on urine output – are confounded by many other factors and change slowly after the injury has already occurred. Consequently, this prevents early recognition of AKI cases and hampers timely interventions, a fact which might explain the extremely negative outcomes related to this disease.

Based on the principle that timely AKI recognition may help dealing with the injury effectively, through careful monitoring of patient's conditions and initiation of appropriate therapies, latest studies have focused on the search for more efficient diagnostic methods,

including AKI prediction scores and discovery and validation of other markers as possible alternatives to SCr. As a result, many urinary and serum proteins have been identified and investigated as possible early *biomarkers* (acronym for biological markers), which can help recognizing patients at risk of developing AKI. Two of these biomarkers, combined together, seem to have the best performance in terms of early detection of AKI risk (Kashani et al. 2013): urine insulin-like growth factor-binding protein 7 (IGFBP7) and tissue inhibitor of metalloproteinases-2 (TIMP-2). The NephroCheck® Test (Astute Medical, San Diego, CA, USA) is an immunoassay that has been developed to measure the concentration of TIMP-2 and IGFBP-7 in urine and to combine them together for the provision of a single numerical test result.

The present dissertation has the purpose of evaluating the NephroCheck® Test as a diagnostic instrument to detect AKI early, in a population of patients undergoing cardiac surgery. A retrospective study has been conducted at the San Bortolo Hospital and the International Renal Research Institute of Vicenza (IRRIV), under the supervision of the Director of the Department of Nephrology, Dialysis, Transplantation and of the IRRIV, Prof. Claudio Ronco. A team of specialists and researchers, including physicians and biologists, has also been involved in order to conduct the study with a multidisciplinary approach. Deep knowledge of the clinical issues and aspects related to AKI as well as an insight into diagnostic procedures helped in the outline and development of the research project. Indeed, in order to evaluate the predictive ability of the NephroCheck® Test as a diagnostic device, a logistic regression model for the AKI prediction has been constructed. The scope was to build a model predicting the probability of developing AKI after cardiac surgery, based on a set of patient's characteristics at admission, surgery procedures and therapies. Subsequently, the NephroCheck® Test results had to be added to the model, in order to evaluate any improvement in the predictive ability of the model attributable to the test.

This dissertation is structured as follows.

Chapter 1 provides a brief introduction on the economic evaluation of health care programs and explains why it has become an important component of a more comprehensive decision-making process concerning resource allocation among alternative projects. Tools and main issues to be addressed, as well as the main types of economic evaluation in health care are shortly described.

The relevant problem is introduced in *Chapter 2,* from a slightly clinical point of view. Researchers and clinicians have struggled, over the past years, due to the lack of a widely

accepted definition and classification of AKI according to its severity; this problem has also hampered comparability of studies' results. This chapter describes in details the latest definition and classification of AKI, adopted also in the course of this study. The issue of AKI diagnosis is also dealt with and the role of novel biomarkers is outlined. After introducing the relevant framework and the features that are deemed necessary for a successful biomarker, the latest studies investigating the performance of TIMP-2 and IGFBP-7 are briefly reviewed. This provides a scenario to understand the potential of the NephroCheck® Test as a new diagnostic instrument.

Chapter 3 provides a theoretical framework for logistic regression model and receiver operating characteristic (ROC) curve analysis. ROC curves are, indeed, largely employed to evaluate the performance of classifiers and diagnostic devices that are meant to discriminate between diseased and non-diseased individuals.

A brief review of the existent literature dealing with clinical outcomes following AKI and consequent economic burden is provided in *Chapter 4*. The impact of AKI is described into more details, based on the findings from most recent single-centre studies. It will be clear that the lack of a uniform definition and classification of AKI has hampered, so far, comparison among studies and has made it difficult to get clear-cut results in terms of impact on clinical outcomes and healthcare costs. Nevertheless, all the presented results point in the same direction, signalling that AKI has, indeed, a relevant impact on healthcare expenditure as well as it negatively affects clinical outcomes. From the review of the literature in this field, some important facts can be drawn about the relevance of AKI: AKI is an issue also in its milder cases, being always associated with increased health care costs and adverse clinical outcomes, and these negative effects are likely to prolong over a longer-term, beyond the hospitalization period.

Finally, *Chapter 5* illustrates the empirical results of a retrospective cohort study for evaluating the NephroCheck® Test as a diagnostic device. The objective of the study was to assess whether the NephroCheck® Test improves AKI prediction through a multiple logistic regression model, which accounts for other patients' characteristics as well as surgical procedures and post-operative parameters. To do so, data from two cohorts of patients undergoing cardiac surgery during 2014 and 2015 in the San Bortolo Hospital of Vicenza were considered. For the 2015 cohort only, the NephroCheck® Test was available and employed. This chapter illustrates the steps for selecting the relevant variables to be included in the base model and analyses the performance of the model before and after including the

results of the test. Finally, by means of comparing the performance of the models in discriminating between AKI and non-AKI individuals, it can be evaluated if the NephroCheck® Test is a powerful and effective diagnostic tool.

1 ECONOMIC ANALYSIS

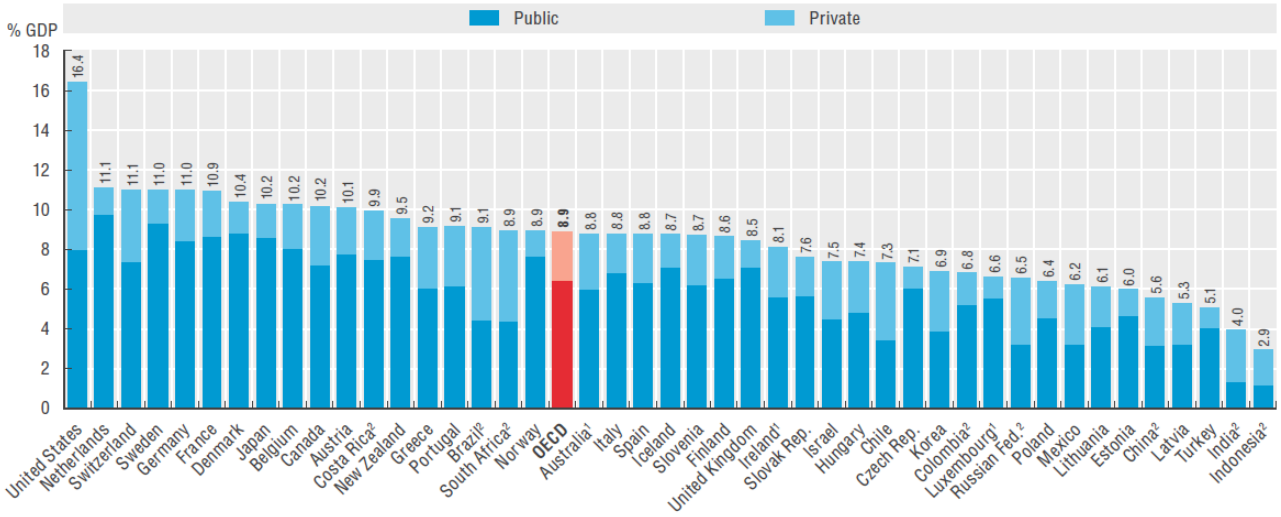
1.1 HEALTH ECONOMICS AND HEALTH TECHNOLOGY ASSESSMENT

Economics is the study of resource allocation among alternative uses when resources are scarce. It is the science that analyses how individuals and societies choose to allocate scarce resources, among competing alternative uses, to satisfy their unlimited wants and needs, and to redistribute the products of this allocation. In particular, economic evaluation can be defined as a tool for assessing the benefits and the costs of competing uses of scarce resources. To pursue this aim, economic evaluation embeds two aspects. Firstly, both resources to be invested (inputs or costs) and the products (outputs or benefits) of a particular activity are analysed. Secondly, the problem of scarcity of resources must be addressed by considering how inputs might be allocated among alternative uses.

Economics has found broad application in many fields, including health care. Health economics (HE) is the discipline that, employing economic concepts and tools, studies how to allocate scarce resources to improve health (see Kobelt 2013). In the context of health care, economic evaluation is “a comparative analysis of alternative courses of actions in terms of both their costs and consequences” (Drummond et al. 2005). This encompasses decisions on resource allocation to the healthcare system within the economy as well as to different and competing activities within the healthcare system itself.

Health expenditure measures the final consumption of health goods and services (i.e. current health expenditure). This includes spending by both public and private sources on medical services and goods, public health and prevention programmes and administration, while spending on capital formation (investments) are excluded (see OECD 2015). The figure below shows that health expenditure as a percentage of GDP was on average 8.9% in the OECD countries during 2013.

Figure 1.1 Health expenditure as a share of GDP, 2013 (or nearest year).



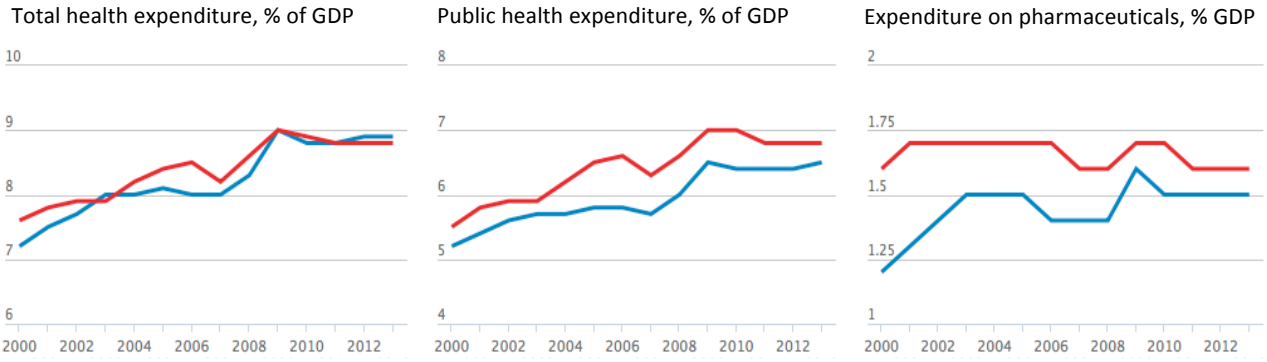
Note: Excluding investments unless otherwise stated.

1. Data refers to 2012
2. Including investments

Source: OECD.

As far as the trend in health expenditure is concerned, most of the second half of the 20th century and the beginning of the 2000s were characterized by a steady increase in health spending as a share of GDP. Things changed during the financial crisis around 2008 when, after an initial increase due to the sharp fall in GDP, health expenditure on GDP decreased gradually in many OECD countries until they reached a stabilisation, following economic growth (Figure 1.2).

Figure 1.2 Trends in health expenditure.



OECD countries: blue line. Italy: red line.

Source: OECD

The great variation in the amount of money that countries spend on health is related to

country-specific budget priorities but also to the determinants of health expenditure, which influence the choice of countries and may stimulate or limit the investments in health. One of the primary drivers of health expenditure is certainly the demographic structure of a population. In particular, the phenomenon of an ageing population, along with growing life expectancy and low fertility rates, characterizes most developed countries and has contributed to the increase in health care spending (see Dybczak and Przywara 2010). Indeed, although consumption of health services depends on health status, the elderly tend to use health care more than the younger cohorts. However, Oliveira Martins and de la Maisonneuve (2006) argue that the link between demographic factors and health expenditure may be weaker than one expects, with other non-demographic factors playing a relevant role. Firstly, income growth positively influences both private and public spending on health care, which are led by a greater awareness on health status and availability of new therapies. Nevertheless, the relationship between GDP and spending on health care depends substantially on income elasticity of demand for health services, whose value is still the central issue of a large debate (Dybczak and Przywara 2010). Also technological progress is viewed by many economists as one of the main drivers of health care costs, as it is "the most important supply factor affecting the entire process of development, production, delivery and financing of health care" (Dybczak and Przywara 2010, 6).

Keeping the costs of health care under control has become a core concern particularly in those countries where health is mainly publicly funded, through taxes and/or social insurance. Across all OECD countries, health care is financed through a mix of public and private funds (in the form of household's payments and/or private health insurance), and government spending usually represents the main source of funds (OECD 2015).

In this context, Health Technology Assessment (HTA) is the field of scientific research that informs policy and clinical decision-making on the introduction and use of health technologies. These include pharmaceuticals, devices, diagnostics, procedures and other clinical, public health and organizational interventions. Kobelt (2013, 1) defines HTA as a "multidisciplinary process that summarises information about the medical, social, economic and ethical issues related to the use of a health technology throughout its life span in a transparent, systematic, unbiased and robust manner". HTA thus examines short- and long-term impact of new health technology and provides policy makers with necessary knowledge about policy alternatives, supporting an informed decision (Banta 2009).

Technology assessment related to the healthcare system developed remarkably since the

last decades of the 20th century and was especially concerned with the safety, cost-effectiveness and efficacy aspects of technology (see Banta 2009). Later on, also the social and ethical issues became important, although cost-effectiveness remained a prominent focus that was exacerbated by the financial crisis of the late 2000s. Economic evaluation is part of HTA and it is a component of a more comprehensive assessment and decision process, where different disciplines must be integrated together with the purpose of advising in the allocation of scarce resources between available alternatives.

The patients –consumers of the new services and technologies – fuel the demand for HTA as they expect more and better information about alternative methods of treatment (Jennet 1992). The public's participation not only influences the priorities of research and provision, but it also helps defining broader outcomes for the assessment process itself, such as quality of life both during and after treatment.

Economic evaluation is often based on the relative efficacy observed in the controlled environment of a clinical trial. This may represent a limit of the appraisal process that would rather need to assess the effectiveness, hence cost-effectiveness, of the treatment in the uncontrolled environment of normal clinical practice (Kobelt 2013, and Drummond et al. 2005). However, the existing techniques for economic evaluation also try to overwhelm the limits related to the features of a controlled clinical trial and attempt to deliver general results with a broader application.

1.2 METHODS FOR ECONOMIC EVALUATION

There are several techniques for economic evaluation, all of them using similar approaches to estimate costs but different methods for measuring consequences, and they are "always comparative and applied to explicit alternatives" (Kobelt 2013, 12). The guiding principle consists in choosing the alternative that maximizes health outcome given a resource constraint. If two strategies deliver the same outcome, the one that costs less is always preferred. Although a technology that delivers a better outcome is normally preferable to existing ones, an outcome improvement often comes at higher costs. Thus, an incremental cost-effectiveness ratio (ICER) must be computed to compare alternative strategies. An ICER is the extra investment required for an additional health benefit and it is calculated as the ratio between the difference in cost and the difference in effect of two alternative interventions, A and B:

$$\text{ICER} = \frac{C_B - C_A}{E_B - E_A}$$

The ICER can be employed as a decision tool by defining a willingness-to-pay (WTP) threshold. If the incremental cost per unit of health benefit is below such threshold, then the (more costly) intervention should be adopted (Kobelt 2013).

The perspective assumed in a HE analysis (e.g. society, patient, healthcare provider, third-party payer) determines the scope and the extent of costs and benefits that are assessed. From a global standpoint, the usual "reference point" is the societal perspective and this is usually the most comprehensive of all approaches, as all relevant costs are accounted for over the specified time horizon, regardless of who incurs them. Adopting a societal perspective also avoids foregoing important benefits that would not be captured by a narrow, budget-specific viewpoint and that could lead to a sub-optimal resource allocation. However, other perspectives are often vital as they inform key decision makers and stakeholders about the impact of health interventions from their vantage point.

Regardless of perspective that is chosen, a transparent analysis must also precisely specify the time period considered in the study and the time span of interest in order to ensure that the relevant costs and outcomes attributable to the study's perspective and scope are accounted for and assigned (Jegers et al. 2002).

When two alternative strategies with the same goal are being evaluated, there is the possibility that some costs are the same for the alternatives, thus having no impact on the results of the evaluation process. Identification of the relevant resources consumed to be included in the analysis is a necessary step for cost estimation. Indeed, costs are computed by multiplying the units of resource used by their unit price, which should represent their opportunity cost – the value of forgone benefit that could be obtained from a resource in its next best alternative use. Typical costs associated to resource use can be classified as follows:

- direct medical costs: costs to the healthcare system (e.g. hospitalization, visits, drugs, tests and procedures, medical devices, nursing care);
- direct non-medical costs: expenses associated to the care of a patient, such as, transportation, related services, devices etc. Care provided by family members, also known as informal care, is included either in this category or in the following one;
- indirect costs: productivity losses that represent a cost to the society and are related to time spent off of work, on behalf of the patient (sick days, impaired productivity at work, early retirement or premature death).

In order to adjust for different timing at which resource use takes place and benefits accrue, another important aspect of HE assessment is discounting, or the determination of the present value of future costs and benefits. Although no consensus on the most appropriate discount rate exists, there are two main options for choosing the discount rate. The former is the real interest rate, which is the foregone return rate on alternative investments for the society. The latter is the rate that reflects social temporal preferences. In practice, however, it is difficult to select an appropriate discount rate; it is rather preferred to present costs in undiscounted form and perform a sensitivity analysis, considering discount rates of 0%, 3% and 5% (Drummond et al. 2005). Indeed, an inherent component that underlies almost all economic analyses is the concept uncertainty, due to the fact that several variables (e.g. discount rate, precise costs, mortality and morbidity rates, etc.) are often unknown. Uncertainty is related to many factors, including lack of precision in variables estimates, the absence of data and the reliance upon lower levels of scientific evidence, and also to inherent technical and technological limits of applied levels of scientific evidence (Heyland et al. 1999). To attenuate uncertainty, HE analyses employ comprehensive and sophisticated sensitivity analysis to illustrate how changes in key underlying variables used in studies and models influence changes in cost, outcomes and cost effectiveness.

In the following paragraphs, different types of economic evaluation will be described. As already noted, what distinguishes the different techniques is how outcomes are treated. In particular, a medical issue lies behind the economic analysis: whether the evaluation process provides support in choosing between alternative treatments for the same disease or informs about what interventions for different diseases should be prioritized.

1.2.1 COST-BENEFIT ANALYSIS

In a cost-benefit analysis (CBA), consequences of programs are defined in monetary units, so that the comparison between incremental costs and incremental benefits of the health program is straightforward. In particular, the discounted stream of future benefits (in monetary terms) is compared with the stream of costs, giving rise to the net benefit/cost of the policy. If the difference between the stream of benefits and costs is positive, there is a net benefit and the intervention is actually acceptable from the viewpoint of the society. The ultimate objective of the CBA is the efficient allocation of resources not only within the health sector, by means of allowing a comparison between alternative investments in different

sectors of the economy (Kobelt 2013). Indeed, it evaluates the health program on the basis of all the outcomes, not just in terms of health, thus representing a complete evaluation method.

In a CBA, costs are measured as opportunity costs, while benefits are expressed as the maximum willingness to pay (WTP) for the outcome of the program. WTP is a measurement technique to evaluate, in monetary terms, the health gains and other kinds of benefits arising from a program. In particular, three main categories of benefits can arise from a program: future costs that will be avoided, health gains and improvements (intangible benefits), productivity gains due to better health conditions. WTP usually accounts only for part of these benefits, in particular the health gains.

As far as the valuation of benefits in monetary terms is concerned, several approaches have been proposed (see Drummond et al. 2005).

The *human capital approach* is based on the idea that the time spent by a person in good health is "convertible" in terms of higher productivity on the labour market. The health outcome is thus evaluated on the basis of the market salaries, in relation to the time spent by a person in good health. However, this approach has some important limitations. Although the salary can be considered an approximation of the marginal productivity of the worker, a lot of market imperfections are likely to influence the analysis, such as inequalities on the labour market. Moreover, if the analysis assumes the perspective of the society, then the outcome is the "time gained in good health", which is not fully convertible in terms of salary. This is, for example, the case of a housewife whose occupation is not "traded" on the labour market, but it is nevertheless valuable to her family and, thus, to the society. Lastly, another criticism concerns the importance of accounting for the collective WTP for the program, which is likely to put this approach in contrast with the paretian principles of Welfare Economy.

An alternative approach (known as the *revealed preferences*) considers the trade-off between risk and salary, in order to analyse the relationship between the health risks associated with an occupation and the salary that the individual requires to accept that job. This is consistent with the Welfare Economy Principles, because it is based on the individuals' preferences towards the increase/reduction of the salary related to a health risk.

The third (and most widely used) method is that of *contingent valuation*: individuals participate surveys and are asked to imagine hypothetical (contingent) scenarios and to reveal their maximum WTP for the realization of a program – with the consequent benefits that it delivers. The main limit to the application of this method derives from the difficulty in

estimating, in monetary terms, the WTP of the individuals, through the questions that they are asked to answer.

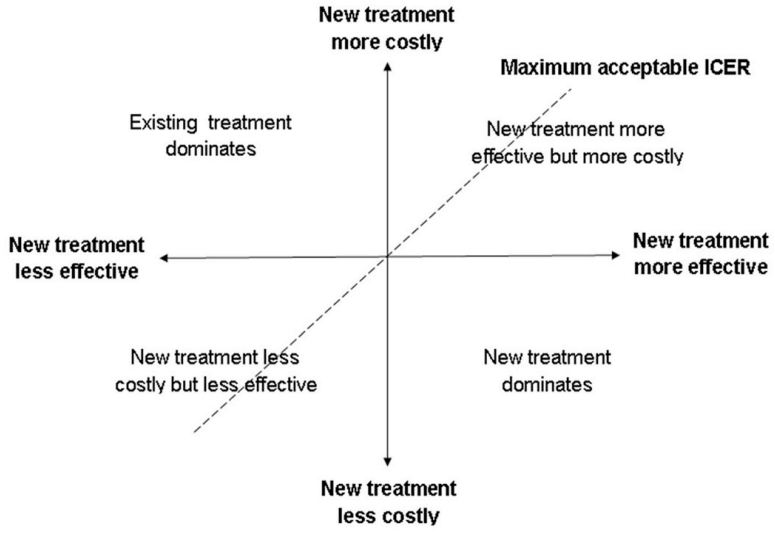
According to Drummond et al. (2005), CBA can be considered the most powerful valuation approach since it addresses the issue of allocative efficiency. However, the main drawback consists in the difficulties arising when trying to value (in monetary terms) the intangible benefits of a health program.

1.2.2 COST-EFFECTIVENESS ANALYSIS

The CEA (cost-effectiveness analysis) is a valuation method of health programs, which takes into account both the costs and the consequences of a policy intervention. The incremental cost of a program is, indeed, compared to the incremental gains in terms of health. Health consequences are measured in physical units correlated with the policy objective and are usually expressed in terms of reduced mortality and morbidity or life years gained. The economic evaluation should, thus, account for the final outcome of a policy over the long term; however, this is often not available or requires a long time (and consequently a lot of resources) to be measured. An alternative procedure consists in measuring the intermediate outcomes and finding a link with the final ones.

The crucial decision tool is the ICER that is used to compare interrelated interventions, in the sense that they usually address the same issue and deliver similar outcome (Tan-Torres Edejer et al. 2003). Alternatively, the estimated cost-effectiveness of the new project is compared with a fixed price cut-off point representing the assumed social WTP for an additional unit of health, expressed as one of the previously described measures. Figure 1.3 below represents the cost-effectiveness plane that provides a framework for the policy maker's choice. The plane is divided into four quadrants, where effectiveness lies on the abscissa and costs on the ordinate. Apart from situations with a clearly dominant intervention compared to alternatives – less effective and more costly or more effective and less costly, represented in the fourth and second quadrants, respectively– there is usually the need for a threshold value. According to this threshold, represented by the dotted line, the decision maker decides whether the additional cost is justified by the additional benefit obtained.

Figure 1.3 Cost-effectiveness plane.



Source: Hounton and Newlands (2012)

A crucial issue in CEA is the valuation of policy effectiveness. It is mainly a matter of medical evidence, which raises the problems of quality of the data and relevance of results. Indeed, the design of a clinical study at the base of an economic analysis influences these two aspects (see Drummond et al. 2005). The endpoint should be to evaluate the *effectiveness* of a program – i.e. the real world outcome – not its *efficacy* – i.e. its theoretical outcome, in a controlled environment (Kobelt 2013). However, some clinical trials are designed in such a way that results are delivered more in terms of efficacy than proper effectiveness and, unfortunately, data on effectiveness may not be available.

Another aspect to be considered in a CEA is the comprehensiveness, according to which clinical data should be representative of the data in the existing literature. A selective collection of clinical data may lead, indeed, to a more (or less) favourable economic evaluation.

1.2.3 COST-UTILITY ANALYSIS

This type of evaluation focuses on the quality of the policy outcome, in terms of health gains or diseases avoided due to the application of a health program. The incremental cost of a policy is compared to the health improvement attributable to the policy itself, measured in quality-adjusted life years (QALYs); results are expressed in terms of cost per QALYs gained. Other possible (but by far less used) measures of the outcome are disability-adjusted life years (DALYs) and healthy year equivalents (HYEs). These are all measures of health

over a medium/long term, which incorporate the impact of interventions on years of life lost due to premature mortality and years of life spent with a disease, weighted by the severity of that outcome.

It may appear similar to CEA, but there is a key difference concerning the measurement of the results. In a CUA (cost-utility analysis), the results may be composite and more complex with respect to the individual outcomes measured in a CEA, allowing the comparison of very different health programs (Kobelt 2013). QALYs is a measure of the outcome that incorporates two aspects: on the one side, the increase in life years (mortality), on the other side, the improvement in the quality of life (morbidity). QALYs is computed by weighting a time quantity, expressed by the increase in life years, by a quality adjustment measure representing the relative preference that individuals or society place on different health states (i.e. utility). The final long-term outcome is, thus, necessary to conduct a CUA because this must be converted into QALYs, which is not possible for intermediate outcomes (Drummond et al. 2005).

The crucial aspect of a CUA is to assess the utility valuation of health states, on behalf of the individuals. This means that a scale of preferences (ranking) towards different health outcomes must be defined, where a utility of one usually refers to full health and zero represents death. Two methods are commonly applied to define utilities associated to different health states (see Drummond et al. 2005, and Kobelt 2013).

The first one consists in interviewing patients directly and collecting their subjective evaluations. Individuals are asked to rate different scenarios associated to their health and treatment received, usually through standard gambles or time trade-offs. This approach is, however, quite time-consuming and requires trained interviewers.

The other method relies on preference-based generic quality-of-life questionnaires, the most widespread of which is the EuroQol Group's EQ-5D. In this case, health states are associated to a predefined set of weights that allow calculating QALYs directly. For this purpose, health is characterized by five dimensions – mobility, self-care, usual activities, pain/discomfort and anxiety/depression – each one measured on three levels of severity – coded as 1 (no problems), 2 (some problems), 3 (severe problems). A framework of 243 theoretically possible health states (plus death and unconsciousness) can be defined, on the basis of an individual's combination of attributes.

Although it has many similarities with CEA, CUA might be a preferable method when the program needs to be evaluated both in terms of quantity and quality of life affected.

Moreover, it is more appropriate when the policy maker must choose among alternative programs – in other words, in case of resource allocation – because QALYs is a more generic measure of health outcome.

1.3 DECISION ANALYTIC MODELS

The golden rule for HE is to use clinical trial data in order to extrapolate values for costs and consequences. However, this is not always possible and trial data might not reflect what would happen in real clinical practice, limiting the generalizability of results. An alternative approach is to extrapolate from existing data to different population groups, longer time frames, or disease endpoints (Mandelblatt et al. 1997). Models are also employed in conducting sensitivity analyses on the parameters included. There are two most commonly used types of model structure: decision tree models and Markov models. Both consider the consequences of clinical actions under uncertainty and represent the possible evolution of events, accounting for probability that a subsequent health state occurs and a final outcome is obtained over a limited time frame.

Decision tree models, in particular, represent events occurring either by chance or decisions over time. Each path in a decision tree depicts one possible sequence of events, associated with a probability and a final outcome (Mandelblatt et al. 1997).

Recurring events are better represented through state-transition models, such as Markov models. Markov models are useful to describe disease outcomes and effects of interventions, because patients can be classified into mutually exclusive "Markov states" and move to a different state as the disease progresses or the intervention succeeds. Transitions occur within a specified time frame, called "Markov cycle", and a whole model is made of different cycles until a final outcome is reached. These decision analytic models are particularly important in settings were risk of disease and disease progression may continuously change over time (see Kobelt 2013) and transition probabilities depend on the current state.

2 ACUTE KIDNEY INJURY AND BIOMARKERS FOR AN EARLY DIAGNOSIS

2.1 THE KIDNEY

The kidneys are two bean-shaped organs, each about the size of a fist. They are located against the back muscles in the upper abdominal cavity, one on each side of the spine. Each kidney is made up of about a million nephrons composed by a filter, the glomerulus, and a tubule; nephrons can be considered the filtering units allowing the kidney perform its functions. Indeed, the kidneys play a dominant role in regulating the composition and volume of the extracellular fluid by means of filtering about 200 litres of blood to produce about 2 litres of urine, composed of wastes and extra fluid. In particular, they prevent the build-up of wastes and extra fluid in the body, they keep levels of chemicals (i.e. electrolytes) stable, they produce hormones that help regulate blood pressure, make red blood cells, produce vitamin D which is necessary for bones health.

The filtration capacity of the kidneys can be measured by the Glomerular Filtration Rate (GFR), which estimates how much blood passes through the glomeruli each minute. In clinical practice, GFR is estimated using a biomarker, the serum creatinine (SCr) that is a waste molecule generated from muscle metabolism and mostly filtered out by the kidneys. Therefore, it is considered a fairly reliable indicator of kidney function. When the filtration capacity is reduced, the kidneys are no longer able to remove all the waste products. This means that the patient has a kidney or renal disease. In this setting, the GFR is important to define the grade and severity of the disease. Several kidney diseases, indeed, affect the filtering units of the nephrons compromising their capability to remove wastes and fluids.

Kidney disease is classified as acute or chronic, and these conditions may lead to what is called end-stage renal disease (ESRD).

Chronic Kidney Disease (CKD) is a slow, progressive and irreversible loss of kidney function (number of working nephrons) over a long period of time (months or years). Levin and Stevens (2014, p.52) define it as “abnormalities of kidney structure or function, present for more than 3 months, with implications for health”. CKD is classified according to cause, GFR category and albuminuria category. Based on GFR, the progression and stages of CKD can be defined as indicated in the following table.

Table 2.1 GFR categories in CKD

Stage	GFR (ML/MIN/1.73m^2)	Description
G1	≥90	Normal or high
G2	60-89	Mildly decreased
G3a	45-59	Mildly to moderately decreased
G3b	30-44	Moderately to severely decreased
G4	15-29	Severely decreased
G5	<15	Kidney failure

CKD: chronic kidney disease, GFR: glomerular filtration rate
Source: Levin and Stevens (2014)

Stage 5 is ESRD, which needs to be treated either through kidney transplantation or with dialysis.

Conversely, Acute Kidney Injury (AKI) is a syndrome that results in a sudden decrease in renal function or kidney damage. AKI may lead to CKD or even kidney failure requiring dialysis. AKI is treated in a more detailed way in the following paragraphs.

2.2 DEFINITION AND CLASSIFICATION OF AKI

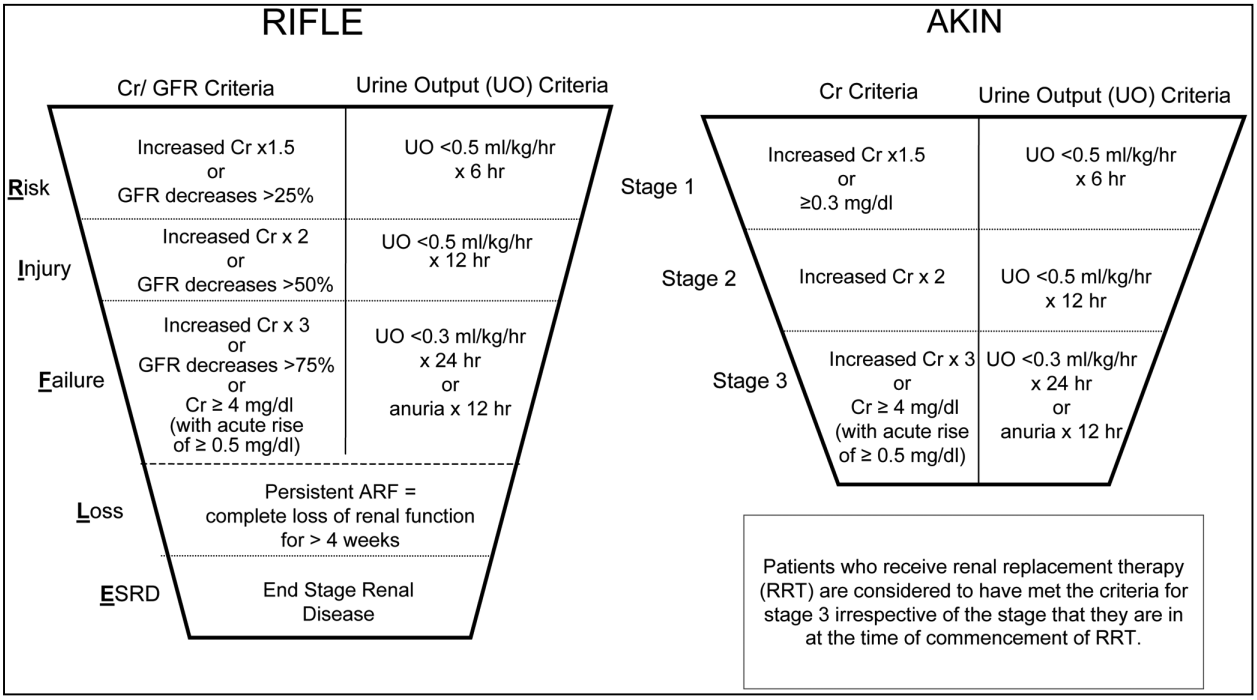
AKI is a common complication among hospitalized patients that affects kidney structure and function and it is associated with adverse clinical outcomes (KDIGO 2012 ch. 2.1). As reported by Mariscalco et al. (2011), in cardiac surgery AKI occurs up to 40% of patients, while in intensive care unit (ICU) one out of four patients is affected by AKI (see De Smedt et al., 2012). AKI is associated with increased morbidity, longer hospital stay, increased mortality (approximately 50% among ICU patients), the need for renal replacement therapy (RRT) and, consequently, increases in healthcare costs (Mariscalco et al. 2011, and De Smedt et al. 2012). Negative outcomes correlated with AKI may be improved in case of a timely detection of AKI and the initiation of an appropriate therapy, as well as the identification of the underlying causes, the monitoring of progression and responses to interventions (Shaw, Chalfin and Kleintjens 2011). However, in the past, the lack of consensus in AKI definition and classification did not allow to estimate the right incidence and to compare study results. Due to the relevance of this disease, awareness of AKI has risen and efforts have been made in order to formulate a commonly accepted definition of AKI. GFR is still considered the most useful index of kidney function; nevertheless, changes in SCr

and urine output (UO) are used as surrogates of changes in GFR in order to define criteria for diagnosis and staging of AKI (see KDIGO 2012, ch. 2.1).

In 2002, the first definition and classification of AKI stages was created in Vicenza during a 2-days conference by the Acute Dialysis Quality Initiative (ADQI): the RIFLE criteria (an acronym for Risk, Injury, Failure, Loss, End-stage kidney disease) were born. This classification is based on criteria for SCr and UO, with a patient that can fulfil either of the two criteria or both and is classified according to the one criterion leading to the worst classification (KDIGO 2012, ch. 2.1). Three classes of AKI severity (risk, injury and failure) and two outcome classes (loss of kidney function and end-stage renal disease) are defined.

In 2007, a modified version of this classification was developed in order to increase specificity and sensitivity of AKI diagnosis: the AKIN (Acute Kidney Injury Network) classification. It distinguishes three AKI stages and does not consider the two outcome classes; it is based on criteria for SCr and UO but not on GFR changes. The figure below shows and compares these two classification criteria.

Figure 2.1. RIFLE and AKIN classifications for acute kidney injury.



Cr: creatinine, GFR: Glomerular filtration rate, ARF: acute renal failure.
Source: Cruz, Ricci and Ronco (2009), (adapted from Bellomo et al. 2004 and Mehta et al. 2007)

In 2012, the Kidney Disease Improving Global Outcomes (KDIGO) released their clinical practice guidelines for AKI, which build off of the RIFLE criteria and the AKIN criteria[1]. AKI is defined as any of the following conditions (without grading):

- increase in SCr by ≥0.3 mg/dl (≥26.5 lmol/l) within 48 hours; or
- increase in SCr to ≥1.5 times baseline, which is known or presumed to have occurred within the prior 7 days; or
- urine volume <0.5 ml/kg/h for 6 hours.

AKI is then staged for severity according to the criteria in the following table.

Table 2.2 Staging of AKI

Stage	Serum creatinine	Urine Output
1	1.5-1.9 times baseline OR ≥0.3 mg/dl (≥26.5 μmol/l) increase	<0.5 ml/kg/h for 6-12 hours
2	2.0-2.9 times baseline	<0.5 ml/kg/h for ≥12 hours
3	3.0 times baseline OR increase in serum creatinine to ≥4.0 mg/dl (≥353.6 μmol/l) OR initiation of renal replacement therapy OR, in patients <18 years, eGFR < 35 ml/min per 1,73 m^2	0.3 ml/kg/h for ≥24 hours OR Anuria for ≥12 hours

AKI: acute kidney injury

Source: Adapted from KDIGO, ch. 2.1 (2012)

The SCr is still the most widely used marker for quantification of renal function: it is easily recorded as part of routine analysis and it has a low cost. However, there are some relevant drawbacks concerning this approach. First of all, SCr levels are influenced by many non-renal factors (sex, age, race, muscle mass, diet and by certain drugs and clinical conditions) and an increase in its level normally occurs when half of kidney function is already lost, thus with substantial time delay with respect to kidney injury. Moreover, it does not adequately account for structural renal tubular injury, which may be extensive in the setting of AKI (Shaw, Chalfin and Kleintjens, 2011). Finally, difficulties arise in case of patients with renal dysfunction whose urine output is scarce.

[1] See Acute Kidney Injury Work Group. Kidney Disease: Improving Global Outcomes (KDIGO) - Clinical Practice Guideline for Acute Kidney Injury. (2012) *Kidney International Supplements,* 2, 1-138.

All these issues related to the use of SCr as a marker for kidney function have raised the need of more efficient methods for AKI diagnosis, which has led to the discovery of other markers as possible alternatives.

2.3 BIOMARKERS FOR EARLY DETECTION OF ACUTE KIDNEY INJURY

Several studies have been conducted with the aim of identifying urinary and serum proteins as early *biomarkers* (acronym for biological markers) that can help in the early detection of kidney damage and possibly discern AKI subtypes. In particular, during the last years, efforts have been made in order to find and validate novel biomarkers for AKI, produced by genes as soon as few hours after the injury has occurred.

Various proteins and biochemical markers have emerged as possible and promising candidates for a timely diagnosis of AKI; however, few of them have reached an "acceptable level of suitability or precision" (Wetz et al 2015, p 1). These biomarkers include, among others, kidney injury molecule-1 (KIM-1), neutrophil gelatinase-associated lipocalin (NGAL), interleukin-18 (IL-18), Cystatine-C (CysC), N-acetyl-glucosaminidase (NAG) and fatty acid-binding protein (FABP). The most important features of a good biomarker are defined in the KDIGO Clinical Practice Guideline (2012, ch. 2.1), by Tsigou et al. (2012, 1-2) and by Zhou et al. (2006) as follows:

a. to diagnose renal dysfunction early;
b. to allow a discrimination between pre-renal, intrinsic and post-renal causes;
c. to provide information about the primary location of the injury (e.g. renal tubules, interstitium, vasculature);
d. to provide a specificity for kidney injury in the presence of coexistent injury affecting other organs;
e. to discern AKI from other forms of renal disease;
f. to classify AKI according to severity;
g. to monitor the progression and the response to therapy;
h. to be cost-effective compared to current practice;
i. to be analysed through relatively non-invasive and easy to perform tests;
j. to have high degree of sensitivity.

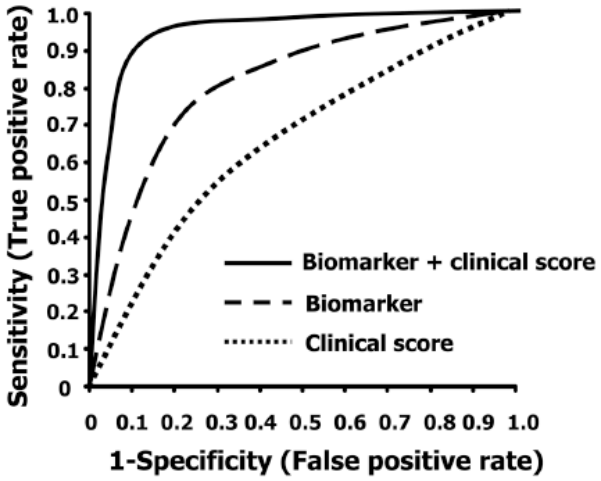
New biomarkers may also help to detect patients with a subclinical risk of developing kidney injury; in other words, those cases in which, even in the absence of a renal dysfunction, a tubular damage to the kidney exists (see Zhou et al. 2006, and Haase, Kellum and Ronco 2012). The expression "subclinical" AKI is used with reference to this condition (see Haase, Kellum and Ronco, 2012, 1-2), such that AKI is not "clinically manifest" in the form of a decreased GFR and increased SCr, yet a structural damage to the kidney has happened and this fact exposes the patient to a higher risk of renal function loss in the future. For this reason, subclinical AKI should be in fact considered AKI. The role of new tubular damage biomarkers would be that of detecting a structural damage, when creatinine shows no evidence of renal dysfunction.

A rule that defines how the biomarker is to be employed as a clinical diagnostic instrument must be specified. This means that a cut-off value distinguishing normal from abnormal levels of the biomarker needs to be identified, such that for every value of the marker that is higher than this cut-off there is a high risk for the patient to develop AKI.

The performance of a biomarker at a variety of cut-off values can be graphically displayed through a receiver operating characteristic (ROC) curve. ROC curves are often used to graph the accuracy of diagnostic instruments along the scale of possible values, showing the trade-off between sensitivity and specificity for each value[2]. This allows selecting the best cut-off level, the value of the diagnostic test (the biomarker, in this case) that maximizes the difference between the so-called "true positives" and "false positives" (see D'Arrigo et al. 2011). The area under ROC curve (AUC) evaluates, then, the overall quality (predictive ability) of the diagnostic instrument and a perfect biomarker should have an AUC of one (Zhou et al. 2006). Individual biomarkers usually have an AUC inferior to 1, but combining them with other biomarkers and/or other clinical information can increase the AUC of the whole diagnostic panel to approximate the desired value of 1 (see Figure 2.2).

[2] See the following chapter for more details on ROC curve analysis.

Figure 2.2



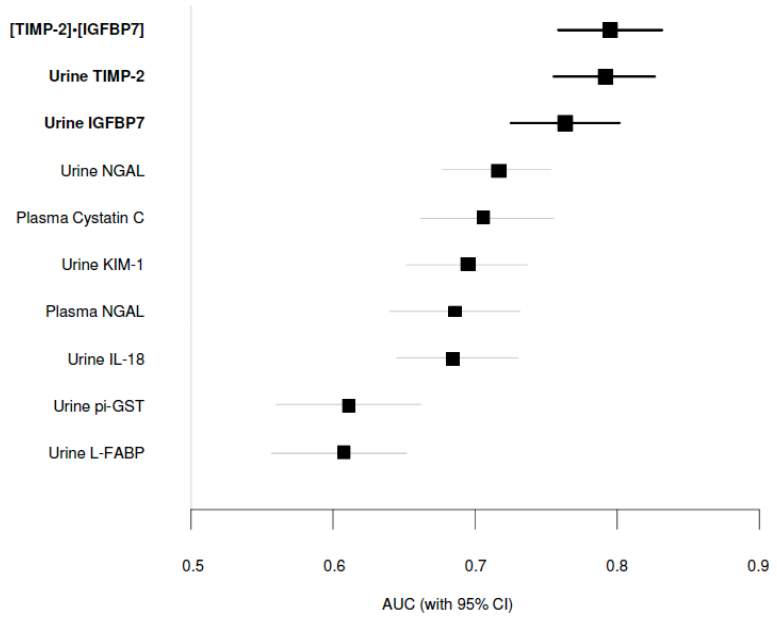
Example of how an area under the receiver-operating characteristic curve (AUC) may improve by means of combining a biomarker with other clinical information.
Source: Zhou et al. (2006)

2.4 THE NEPHROCHECK® TEST

Among all the biomarkers that clinical studies are trying to evaluate and validate, two of them, combined together, are outperforming the others according to the latest studies (see Kashani et al. 2013): urine insulin-like growth factor-binding protein 7 (IGFBP7) and tissue inhibitor of metalloproteinases-2 (TIMP-2). They are soluble proteins, expressed in kidney and other tissues, and are thought to be involved in G1 cell-cycle arrest during the earliest phases of injury. Thus, both TIMP-2 and IGFBP7 are markers of cellular stress in the early phase of tubular cell injury caused by a wide variety of insult (see Wetz et al. 2015, and Ronco 2014).

Kashani et al. (2013) conducted a multi-centre observational study in North America and Europe, enrolling 744 critically ill patients at risk of developing AKI, where IGFBP7 and TIMP-2 were identified, for the first time, as the top-performing biomarkers among a set of existing novel biomarkers. The assessment of the biomarkers' performance was primarily based on the value of the receiver operating AUC, which was 0.77 and 0.75 for IGFBP7 and TIMP-2, respectively, "for the identification of risk of developing AKI (RIFLE-I/F) within 12 to 36 hours" (Kashani et al. 2013, 4). In particular, IGFBP7 and TIMP-2 seemed to have additive predictive value when used together.

Figure 2.3 Area under the receiver-operating characteristic curve (AUC) for TIMP-2 and IGFBP7, as compared to other existing biomarkers of AKI.



IGFBP7: insulin-like growth factor-binding protein 7, IL-18: interleukin-18, KIM-1: kidney injury molecule-1, L-FABP: liver-type fatty acid-binding protein, NGAL: neutrophil gelatinase-associated lipocalin, pi-GST: glutathione S-transferase Pi, TIMP-2: tissue inhibitor of metalloproteinases-2.
Source: Kashani et al. (2013).

Consequently, in the further validation phase of the investigation, the so-called Sapphire study, the multiplication [TIMP-2]×[IGFBP7] of the two markers was evaluated: the AUC of this combination was 0.80, again, for the development of moderate to severe AKI (stage 2 to 3) within 12 hours from sample collection. A primary clinical cut-off value of 0.3 was also derived, within the Sapphire study, as having a high sensitivity for predicting the development of moderate to severe AKI in the following 12 hours (see Ronco 2014). Figure 2.3 shows

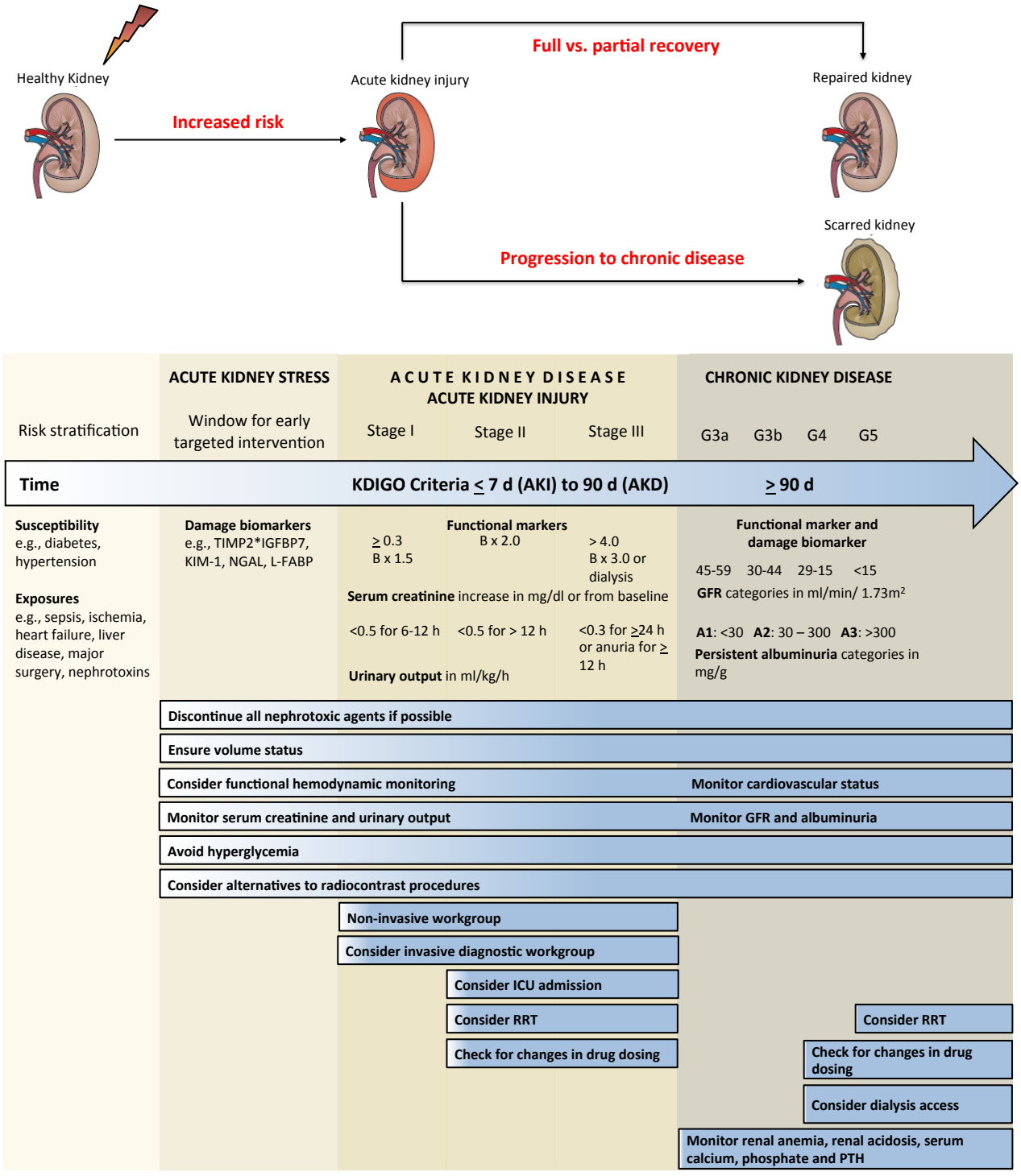
clearly how the two biomarkers (combined together but also alone) outperform other existing markers.

For the purpose of this study, the NephroCheck® Test (Astute Medical, San Diego, CA, USA), which was developed by Astute Medical and employed in the Sapphire validation study, was used. The NephroCheck® Test is an immunoassay measuring the concentration of TIMP-2 and IGFBP-7 in urine and combining them together to provide "a single numerical test result" in a range between 0.04 and 10.0. The detection process takes place through the insertment of the NephroCheck® Test cartridge into the associated diagnostic device, the Astute140® Meter, that exploits a "fluorecence detection technology to quantitatively measure biomarkers of acute kidney injury in human urine samples".

The NephroCheck® Test represents a promising diagnostic device that could potentially fulfil the following objectives. Firstly, it signals which patients have an increased risk of developing AKI. Secondly, it allows a stratification of patients, distinguishing those that have a lower risk form those having a greater risk of developing AKI in the subsequent 12 hours. For this purpose, two clinical cut-offs have been defined (see Kashani et al. 2013, and Hoste et al. 2014). A lower cut-off of 0.3 allows early recognition of most patients that are at risk of developing AKI. Therefore, 0.3 is intended to be used in routine clinical practice as a signal to start those measures and undertake mild management strategies, which promote adaptive renal recovery and may prevent progression to chronic disease, as outlined in the KDIGO guidelines[3] (see Figure 2.4). A higher cut-off of 2.0 identifies individuals that are at the highest risk of AKI, for whom more active kidney sparing interventions may be appropriate.

[3] For more details see kdigo.org/home/guidelines/acute-kidney-injury/ and kdigo.org/home/guidelines/ckd-evaluation-management

Figure 2.4. Conceptual model of AKI to CKD and stage-based management of AKI according to KDIGO guidelines.



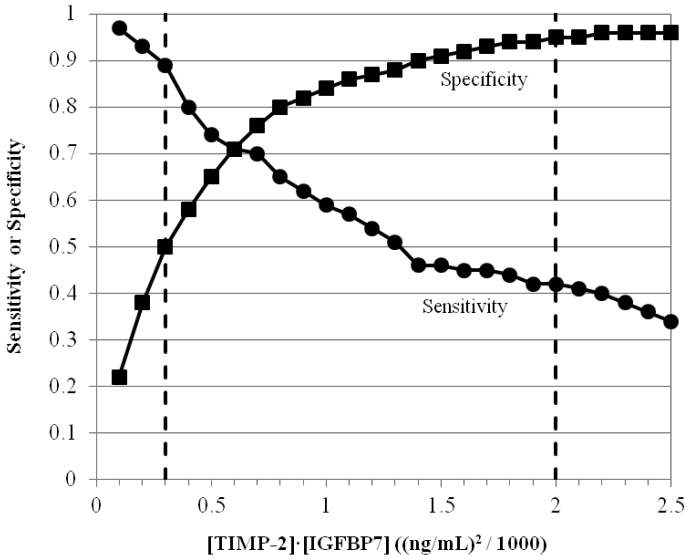
AKD: Acute Kidney Diseases and Disorders, AKI: Acute Kidney Injury, GFR: Glomerular Filtration Rate, PTH: parathyroid hormone, RRT: Renal Replacement Therapy.

Adapted from Ferenbach and Bonventre (2015)

Finally, as already anticipated, awareness of an increased risk enables clinicians to undertake nephro-protective measures, to make choices on management and timing of therapies, and to implement strategies that can help preserving the kidneys and potentially mitigate the adverse outcomes associated with AKI.

In conclusion, the opportunity to detect AKI early is a precious tool for clinicians, with positive impact on patient management and care. On the one hand, biomarkers cannot forecast future AKI development, if this event has not happened yet when biomarkers are sampled. On the other hand, biomarkers may represent a powerful and robust alert system (also in combination with the overall patient's clinical picture). Clinical cut-offs of 0.3 and 2.0 have, indeed, been selected as having, respectively, a high sensitivity and high specificity (see Figure 2.5).

Figure 2.5 Sensitivity and specificity for [TIMP-2] × [IGFBP7]



Values are calculated for the development of AKI (KDIGO Stage 2-3) within 12 hours of sample collection, as a function of [TIMP-2] × [IGFBP7] cut-off value in the Sapphire cohort.
Source: Hoste et al., (2014)

Even if a crucial issue remains on how physicians may realize this potential fully in clinical settings, since no specific therapy exists for AKI treatment, at the same time it is recognized that delay in AKI diagnosis contributes to ineffective care (Hoste et al. 2014). On the contrary, early recognition of the risk affords sufficient time to intervene actively or through more kidney-sparing measures, helps reducing the incidence and severity of AKI case and avoiding additional exposure to other AKI-causing events (see Hoste et al. 2014, Ronco 2014, Shaw, Chalfin and Kleintjens 2011).

3 LOGISTIC REGRESSION AND ROC CURVE ANALYSIS

3.1 THE LOGISTIC REGRESSION MODEL

Regression models are often employed in medical research to evaluate the contribution of several factors to an outcome of interest. The aim of any regression analysis is to find the best fitting, parsimonious but also clinically interpretable model to describe the relationship between the outcome (or dependent) variable and a set of independent predictors (called explanatory variables or covariates). In many clinical settings, the outcome variable is binary, meaning that it takes one of only two possible values representing that an event takes place or not, or more generally the presence or absence of an attribute of interest. A logistic regression model is the most frequently adopted in the analysis of this type of data (see Hosmer, Lemeshaw and Sturdivant 2013). For the purpose of this study, the "event" of interest consists of developing/not developing AKI after cardiac surgery and a multiple logistic regression model was chosen.

In a logistic regression, the outcome variable y is such that

$$y = \begin{cases} 1 & \text{if the attribute is present} \\ 0 & \text{otherwise.} \end{cases}$$

y is the realization of the random variable Y, that can assume value one and zero with probability $\pi(\boldsymbol{x})$ and 1- $\pi(\boldsymbol{x})$, respectively.

The key quantity to be estimated in the logistic regression, like in any other regression model, is the mean value of the outcome variable Y, given the value of a vector of p independent predictors $\boldsymbol{x}' = (x_1, x_2, \ldots x_p)$: the conditional mean $\mathrm{E}(Y|\boldsymbol{x})$, where $\boldsymbol{x}$ represents specific values assumed by the independent variables. This quantity – read as "the expected value of Y given $\boldsymbol{x}$" – is equal to the conditional probability $\Pr(Y = 1|\boldsymbol{x})$ and must be greater than or equal to zero and lower than or equal to one ($0 \leq \mathrm{E}(Y|\boldsymbol{x}) \leq 1$).

The regression model for Y is therefore $Y = \pi(\boldsymbol{x}) + \varepsilon$ and the error term ε can assume two values:

$$\varepsilon = \begin{cases} 1 - \pi(\boldsymbol{x}) & \text{with probability } \pi(\boldsymbol{x}) \\ -\pi(\boldsymbol{x}) & \text{with probability } 1 - \pi(\boldsymbol{x}). \end{cases}$$

In other words, Y follows a Bernoulli distribution with parameter $\pi(\boldsymbol{x})$:

$$\Pr\{y|\boldsymbol{x}\} = \pi(\boldsymbol{x})^y \big(1 - \pi(\boldsymbol{x})\big)^{1-y}$$

(note that the expression is equal to $\pi(\boldsymbol{x})$ if $y = 1$ and to $1 - \pi(\boldsymbol{x})$ if $y = 0$). The expected value and variance of Y conditional on x are $\mathrm{E}(Y|\boldsymbol{x}) = \pi(\boldsymbol{x})$ and $\mathrm{var}(Y|\boldsymbol{x}) = \pi(\boldsymbol{x})\big(1 -$

$\pi(\boldsymbol{x})\big)$, where $\mathrm{E}(Y|\boldsymbol{x}) = \Pr(Y = 1|\boldsymbol{x})$ is the probability of the event (or the probability that the attribute is present).

Considering that the model describes a conditional probability, a distribution that satisfies the constraint $0 \leq \mathrm{E}(Y|\boldsymbol{x}) \leq 1$ is necessary and the logistic distribution does it. The curve of this distribution (represented in Figure 3.1) is S-shaped, delimited from the lines y=1 and y=0 to which it asymptotically tends.

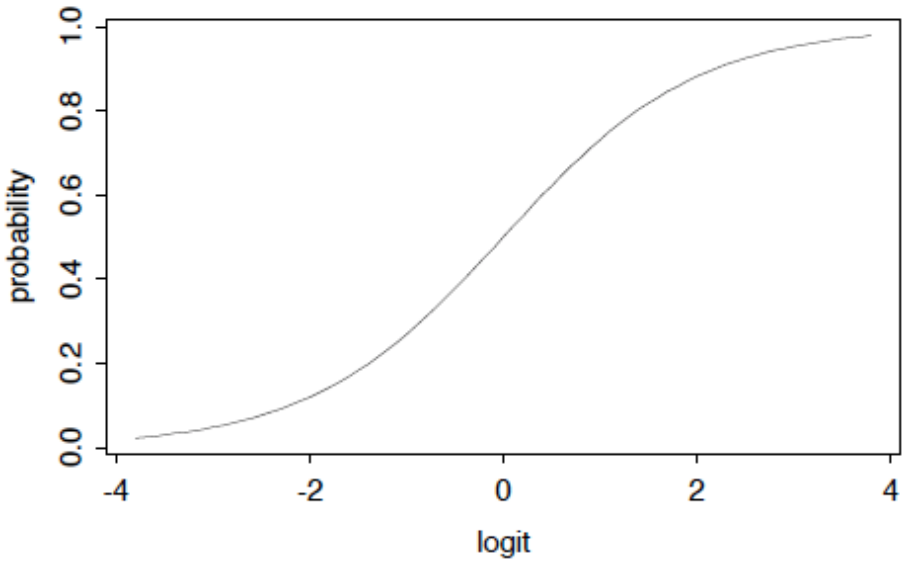


Figure 3.1. The Logit Transformation. (Source: Rodriguez)

In particular,

$$\pi(\boldsymbol{x}) = \frac{e^{g(\boldsymbol{x})}}{1 + e^{g(\boldsymbol{x})}}$$

is the specific form of the logistic regression model that satisfies the constraint $0 \leq \mathrm{E}(Y|\boldsymbol{x}) \leq 1$, where $g(\boldsymbol{x})$ is the *logit transformation* of $\pi(\boldsymbol{x})$:

$$g(\boldsymbol{x}) = \ln\left[\frac{\pi(\boldsymbol{x})}{1 - \pi(\boldsymbol{x})}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p.$$

The logit $g(\boldsymbol{x})$ is expressed as the natural logarithm of the odds, defined as "the ratio of the probability to its complement" (Rodríguez 2007). It has some interesting properties: it is linear in its parameter, may be continuous and may range from $-\infty$ to $+\infty$, depending on the $\boldsymbol{x}$.

3.2 ESTIMATION AND INTERPRETATION OF THE MULTIPLE LOGISTIC REGRESSION MODEL

Fitting a logistic regression model to a sample of *n* independent observations (x_i, y_i) for the *i*-th individual, with *i*=1,2,…,*n,* requires the estimation of the vector of parameters $\boldsymbol{\beta}' = (\beta_0, \beta_1, \ldots, \beta_p)$ through the method of *maximum likelihood*, that delivers the values of the unknown parameters that maximize the probability of obtaining the observed set of data. This method requires the specification of *the likelihood function*, which expresses the probability to observe the realized data as a function of the unknown coefficients. The likelihood function is obtained as the product of the contribution for each *i*-th independent observation (x_i, y_i),

$$\pi(\boldsymbol{x}_i)^{y_i}[1-\pi(\boldsymbol{x}_i)]^{1-y_i},$$

so we get

$$l(\boldsymbol{\beta}) = \prod_{i=1}^{n} \pi(\boldsymbol{x}_i)^{y_i}[1-\pi(\boldsymbol{x}_i)]^{1-y_i}.$$

The log of this expression is computed to obtain the log-likelihood function

$$L(\boldsymbol{\beta}) = \ln[l(\boldsymbol{\beta})] = \sum_{i=1}^{n}\{y_i \ln[\pi(\boldsymbol{x}_i)] + (1-y_i)\ln[1-\pi(\boldsymbol{x}_i)]\}.$$

The estimate of $\boldsymbol{\beta}$ that maximize the expression is obtained by means of differentiating $L(\boldsymbol{\beta})$ and computing the likelihood equations:

$$\sum_{i=1}^{n}[y_i - \pi(\boldsymbol{x}_i)] = 0$$
$$\sum_{i=1}^{n} x_{ij}[y_i - \pi(\boldsymbol{x}_i)] = 0$$

for *j*=1,2,..,*p*.

The value of $\boldsymbol{\beta}$ obtained by the solution to these equations is the *maximum likelihood estimate* $\widehat{\boldsymbol{\beta}}$.

A key issue is to understand the meaning of the regression coefficients estimated in a logistic regression. Apart from the intercept coefficient (in most cases of limited interest), the slope coefficients represent the rate of change of a function of the dependent variable per unit of change in the independent variable. In a logistic regression model, this function is the logit transformation $g(\boldsymbol{x}) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$, so that each β is a change in the logit, corresponding to a unitary change in the predictor:

$$\beta_1 = g(x_1 + 1) - g(x_1).$$

A meaningful interpretation of this association is offered in terms of the *odds ratio*.

In the simplest case where there is one dichotomous covariate (x can be either 0 or 1), the odds ratio OR is the ratio of the odds for $x = 1$ to the odds for $x = 0$:

$$\text{OR} = \frac{\frac{\pi(1)}{[1-\pi(1)]}}{\frac{\pi(0)}{[1-\pi(0)]}}.$$

By substituting the expression $\pi(x)$ for x = 0, 1 it becomes OR $= e^{\beta_1}$. The odds ratio indicates how much more likely/unlikely – in terms of odds – it is for the outcome to be present ($y = 1$) in individuals with $x = 1$ with respect to individuals with $x = 0$. In case of a continuous covariate, the interpretation of the odds ratio does not differ much; the crucial point is to define a meaningful change in the covariate x, such that it helps the interpretation of how the odds of the outcome changes with the x.

3.3 LOGISTIC REGRESSION DIAGNOSTICS

To assess the goodness of fit of the logistic regression model some statistics are normally computed, to assess whether the probabilities estimated by the model reflect accurately the true outcome pattern in the data. To do so, both overall measures of fit and measures of the difference between observed and fitted values $(y - \hat{y})$, also graphically, are important. Firstly, it is important to define that the expression *covariate pattern* describes a particular configuration of values assumed by the covariates in a model. The number of covariate patterns may be an issue when it comes to assess the fit of a model (see Hosmer, Lemeshaw and Sturdivant 2013).

In a model containing p explanatory variables $\boldsymbol{x}' = (x_1, x_2, \ldots x_p)$ and with J possible distinct covariate patterns observed for all the n subjects, if some individuals have the same value of $\boldsymbol{x}$ then $J < n$. Denoting the number of subjects with $\boldsymbol{x} = \boldsymbol{x_j}$ by m_j, with j=1,2,…J, it follows that $\sum m_j = n$. If y_j indicates the number of individuals such that $y = 1$, among the m_j subjects characterized by the pattern $\boldsymbol{x} = \boldsymbol{x_j}$, then $\sum y_j = n_1$ is the total number of subjects with $y = 1$.

A first measure of the difference $(y - \hat{y})$ is the *Pearson residual*, which measures the relative deviations between observed and fitted values. For a generic j-th covariate pattern, the Pearson residual is defined as the standardized difference between the observed frequency and the predicted frequency

$$r(y_j, \hat{\pi}_j) = \frac{(y_j - m_j\hat{\pi}_j)}{\sqrt{m_j\hat{\pi}_j(1-\hat{\pi}_j)}}$$

and the Pearson chi-square statistic is computed

$$\chi^2 = \sum_{j=1}^{J}\left[r\left(y_j, \hat{\pi}_j\right)\right]^2.$$

Another method is based on the contribution of each point to the likelihood. This is expressed by the *deviance residual*

$$d\left(y_j, \hat{\pi}_j\right) = \pm\left\{2\left[y_j \ln\left(\frac{y_j}{m_j\hat{\pi}_j}\right) + \left(m_j - y_j\right)\ln\left(\frac{m_j - y_j}{m_j\left(1-\hat{\pi}_j\right)}\right)\right]\right\}^{1/2},$$

where the sign is the same of $\left(y_j - m_j\hat{\pi}_j\right)$. The goal is to minimize the sum of the deviance residuals and the corresponding statistic is

$$D = \sum_{j=1}^{J} d\left(y_j, \hat{\pi}_j\right)^2.$$

With grouped data, under the assumption that the fitted model is correctly specified, both statistics are asymptotically distributed as a chi-square with $J - (p + 1)$ degrees of freedom.

A summary measure of goodness of fit is the Hosmer-Lemeshow Test, which is based on the idea that the predicted frequency and observed frequency should match closely, and that the model fits better as the frequencies match more closely. The test statistic is computed as the Pearson chi-square from the contingency table of observed and expected frequencies, and the Hosmer-Lemeshow Test yields a large p-value in case of good fit.

The Hosmer-Lemeshow statistic is computed as

$$\hat{C} = \sum_{k=1}^{g}\left[\frac{(o_{1k} - \hat{e}_{1k})^2}{\hat{e}_{1k}} + \frac{(o_{0k} - \hat{e}_{0k})^2}{\hat{e}_{0k}}\right],$$

with

$o_{1k} = \sum_{j=1}^{c_k} y_j$,

$o_{0k} = \sum_{j=1}^{c_k}\left(m_j - y_j\right)$,

$\hat{e}_{1k} = \sum_{j=1}^{c_k} m_j\hat{\pi}_j$,

$\hat{e}_{0k} = \sum_{j=1}^{c_k} m_j(1 - \hat{\pi}_j)$

and c_k is the number of covariate patterns in the *k*-th group.

When the model involves many continuous predictors, the contingency table might get too large with the risk of delivering significant results more than often. Therefore, a strategy usually adopted consists in splitting the sample in *g* groups according to the predicted probabilities (see Hosmer, Lemeshaw and Sturdivant 2013). For example, if *g*=10 is chosen, as usually suggested, for a sample with n observations, then the first group contains a number of subjects n/10 having the smallest estimated probabilities. The test statistic approximately

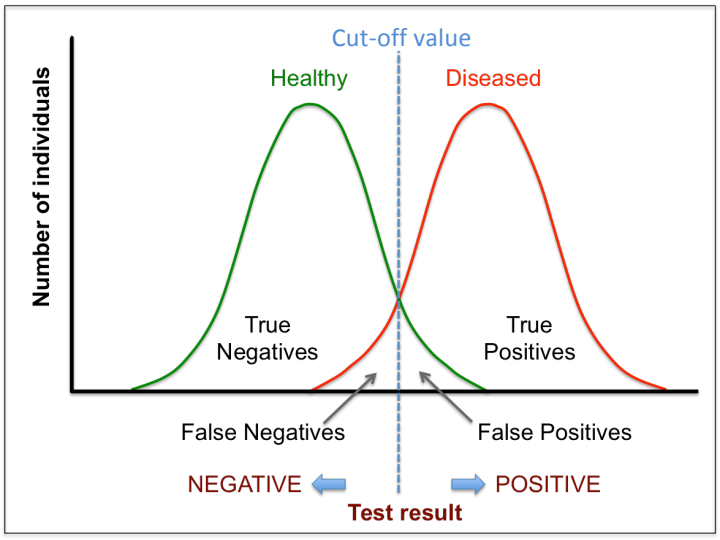
follows a chi-square distribution with $g-2$ degrees of freedom, when the model is correctly specified.

3.4 THE RECEIVER OPERATING CHARACTERISTIC CURVE FOR THE ANALYSIS OF A DIAGNOSTIC TEST'S PERFORMANCE

The receiver operating characteristic (ROC) curve is a graphical plot used to assess the performance of a classifier. This technique finds wide application in the evaluation of clinical diagnostic tests that provide a result in the form of a continuous variable. An ideal test is able to discriminate perfectly diseased cases from normal (i.e. healthy) cases, but when this is not the case ROC curve analysis measures the accuracy of a test along the range of all possible values. In addition to that, ROC analysis allows selecting the best cut-off, the value of the test that maximizes the difference between true positives and false positives (D'Arrigo et al. 2011).

All these concepts can be better explained with the help of a graphical representation. Figure 3.2 illustrates that the distribution of a hypothetical test results for the two groups (healthy and diseased) partially overlap, being not able to discriminate perfectly and misclassifying some individuals.

Figure 3.2 Graphical representation of the proportions of TP, TN, FP and FN



Individuals that are incorrectly classified can be either "false negatives" (FN, people that the test identifies as negative, thus healthy, when in fact they are positive, thus diseased)

or “false positives”(FP, people classified as diseased that in fact are healthy) (see also Table 3.1).

On the basis of the proportion of individuals classified in the categories illustrated in Table 3.1, the power of a diagnostic test is evaluated by computing a series of statistics. Let “a” be the number of true positives, “b” of false negatives, “c” false positives and “d” true negatives, then:

- *sensitivity* = $\frac{a}{a+b}$ is the probability that the test result will be positive when the disease is present. Also known as the *true positive rate*, it is expressed as a percentage;
- *specificity* = $\frac{d}{c+d}$ is the probability that the test result will be negative when the disease is not present. Also known as the *true negative rate*, it is expressed as a percentage. Specificity is related to the *false positive rate* ($\frac{c}{c+d}$ the proportion of FP individuals on all the healthy) through the equation *specificity* = 1 – *fp rate;*
- *positive predictive value* (PPV) = $\frac{a}{a+c}$ is the probability that the disease is present when the test is positive;
- *negative predictive value* (NPV) = $\frac{d}{b+d}$ is the probability that the disease is not present when the test is negative;
- *accuracy* = $\frac{a+d}{a+b+c+d}$ is the proportion of patients correctly classified.

Table 3.1. Contingency table for test results against the disease state.

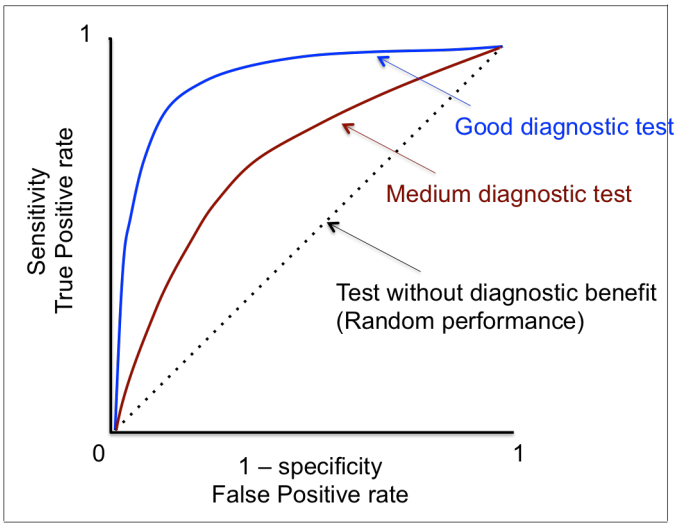
		Disease	
		Present	**Absent**
Test Result	**Positive**	True Positive (TP) a	False Positive (FP) c
	Negative	False Negative (FN) b	True Negative (TN) d

Differently from the PPV and the NPV, sensitivity and specificity are independent of the disease prevalence, in other words, they do not depend on the number of disease cases in the population of interest (D’Arrigo et al. 2011).

The ROC curve plots, in a two-dimensional plane, the sensitivity (on the *y axis*) as a function of the false positive rate (1 – specificity, on the *x axis*) for different cut-off points. This graphical representation illustrates the trade-offs between the TP (benefits of the test)

and the FP (costs of the test) for every point on the curve, each one representing a possible decision threshold. The more to the northwest of the graph a point lies, the higher is the accuracy of the associated threshold value because it is further from a random classification (represented by the dotted line in Figure 3.3). The area under the curve (AUC) measures the overall accuracy of the test: the greater the AUC, the better the discriminating performance of the test. An ideal test would have au AUC of 1, in practice this is very difficult to obtain and a good test should have AUC≥0.8.

Figure 3.3. Basic ROC curve representation on the *xy* plane.



4 ADVERSE OUTCOMES AND COSTS ASSOCIATED WITH ACUTE KIDNEY INJURY

Acute kidney injury (AKI) is a serious and common problem among hospitalized patients and it is a recognized complication particularly following cardiac surgery, where incidence can reach 40-50% (see Mariscalco et al. 2011 and Shaw, Chalfin and Kleintjens 2011), and in intensive care unit (ICU) where approximately one out of four patients is diagnosed AKI (De Smedt et al. 2012). On the basis of single-centre reports, Chertow et al. (2005) report that 5% to 7% of all hospitalized patients develop AKI. However, many studies conduced so far recorded a variable AKI incidence, because different AKI definitions were adopted in the past. Wide ranges for AKI incidence are, thus, due to the lack of unique and widely accepted classification criteria based on serum creatinine (SCr), until the KDIGO criteria were released in 2012 with the aim of building clinical practice guidelines for AKI and a benchmark for AKI definition and classification[4]. Adoption of a uniform definition is necessary also to make results from different studies comparable.

Failure to identify, diagnose and classify AKI timely is a serious issue because of the adverse outcomes that arise even in case of mild AKI cases (Kerr et al. 2014); in more severe cases, where kidney fails and dialysis treatment is necessary, 30% to 50% of patients die (see Taylor 2011). Allowing AKI to progress may lead to negative short- and long-term outcomes for the patient, such as increased morbidity, both in-hospital and post-discharge increased mortality, need for renal replacement therapy (RRT) and, eventually, possible progression to chronic kidney disease (CKD). Identification of risk factors, careful patient monitoring, timely and appropriate start of required therapies are necessary to reduce AKI risk and subsequent deterioration of patient conditions.

The lack of a consensus definition and classification criteria, so far, has not only hampered clinical judgement and prompt intervention, but has also made it difficult to estimate resource use associated with AKI and its impact on healthcare costs (see Dasta et al. 2008). Nevertheless, in this chapter some findings will be presented concerning the estimation of costs related to AKI, not only in terms of resource use and costs during hospitalization, but also long-term costs and adverse clinical outcomes. In doing so, the aim is to delineate a

[4] For more details on KDIGO guidelines for AKI definition and classification and on previously adopted criteria see Chapter 2.

framework to help understanding the burden of AKI, as well as the urgent need for effective AKI detection methods.

Direct assessment of AKI-related costs and outcomes has taken place mostly at single-centre level. Hereafter are presented some findings drawn from the work of Kerr et al. (2014), who estimated the economic impact of AKI, using data on hospital episode statistics (HES) and reference costs in England for the period 2010-11, as well as data from academic studies. Previous attempts to estimate consequences and AKI associated costs were made by Chertow et al. (2005) at an academic medical centre in Boston and by Dasta et al. (2008), who quantified the total postoperative costs associated with AKI following CABG surgery at Pittsburgh Medical Centre.

A relatively extensive literature exists, as far as costs of dialysis is concerned, also for patients who need dialysis treatment as result of developing AKI during hospitalization. Costs and outcomes associated with dialysis will also be briefly reported in the following paragraphs.

4.1 ADVERSE CLINICAL OUTCOMES ASSOCIATED WITH ACUTE KIDNEY INJURY

4.1.1 SHORT-TERM OUTCOMES

All AKI patients, even those who experience milder cases, have longer hospital stays, a fact that makes AKI an expensive condition. In their study, Kerr et al. (2014) extrapolated some data from academic studies (more precisely, from East Kent Hospitals University Foundation Trust EKHUFT), in addition to data collected through HES. The reason why these additional analysis were deemed necessary is that AKI has been traditionally under recognised and under coded in hospital routine datasets, a fact that might be due to the already cited lack of uniform definition and classification of AKI. In particular, Kerr et al. (2014) suggested that hospital data are more likely to record the severe cases, thus underestimating the actual incidence of AKI. Indeed, approximately 70% of the AKI cases recorded in EKHUFT data were AKI stage 1 (according to AKIN classification criteria) and 15.33% of all hospital admissions were characterized by an AKI case (any severity). Conversely, HES recorded AKI in only 2.43% of all hospital admissions, confirming that HES data underestimate actual AKI incidence and prevalence.

In terms of hospital length of stay, HES data reported length of stay for AKI patients

which is 2.57 times as long as for other patients, on average. In EKHUFT data, conversely, individuals who developed AKI experienced a length of stay 1.62 times that of non-AKI patients; additionally, it was shown that length of stay increases along with AKI severity (see Table 4.1).

Besides longer hospital stays, AKI inpatient care is associated with more intense resource use, in particular in the form of excess critical care bed days. The cost of a critical care bed day is higher as compared to other departments bed days and, for the AKI population, the analysis pointed out a critical care bed day usage 4.32 times the usage for patients without AKI (Kerr et al., 2014).

As previously mentioned, AKI is associated also with increased in-hospital mortality. Data from HES reported that 28.11% of AKI patients died before discharge and the relative risk[5] with respect to non-AKI patients is 4.69. The percentage drops to 13.93% according to EKHUFT data (relative risk is 3.50), which might due to the fact that academic studies recognised also less severe AKI cases. The findings extrapolated from EKHUFT data are summarized in Table 4.1 below.

Table 4.1. LOS, critical care days and in-hospital mortality by AKI status, EKHUFT data

	No AKI	**All AKI**	**AKIN 1**	**AKIN 2**	**AKIN 3**	**Unknown**
LOS						
Mean (SD)	4.5 (10.5)	10.72 (15.44)	9.7 (14.6)	12.3 (16.0)	14.9 (18.5)	2.3 (9.8)
Ratio	1	1.62	1.52	1.88	2.16	0.45
Critical care						
Mean (SD)	0.05 (1.02)	0.35 (2.35)	0.17 (1.74)	0.31 (1.80)	1.57 (4.78)	0.02 (0.51)
Ratio	1	4.32	2.60	5.61	18.2	0.02
In-hospital mortality						
% mortality	2.00	13.93	8.10	25.60	33.30	1.97
Relative Risk	1	3.50	2.11	5.79	8.94	1.31

All values are statistically different (p-value<0.001).
AKI: acute kidney injury; LOS: length of stay.
(Source: adapted from Kerr et al., 2014)

[5] Relative risk is defined as the ratio between cumulative incidence in the exposed group (AKI patients) and cumulative incidence in the unexposed group (non-AKI patients).

Taylor (2011) reported some numbers from the Manchester University Foundation Trust that are in line with the framework drawn by Kerr et al. (2014) based on EKHUFT data. Incidence of AKI was 23.5%, with a median length of stay of 17.8 days in the AKI group and 6.8 days in the non-AKI group. Additionally, 11.5% AKI patients required critical care, compared to only 2% of the remaining group of patients. As far as in-hospital mortality is concerned, 5.6% of AKI patients and 1.37% of non-AKI patients died before discharge. These findings suggest that if AKI incidence were reduced by 10%, then Manchester University Foundation Trust could save as many as 5,000 bed days.

4.1.2 LONG-TERM OUTCOMES

AKI is associated not only with higher in-hospital mortality and need for additional treatments during hospital stay, but also with longer-term adverse outcomes, such as increased mortality risk and higher risk of progression to CKD.

Kerr et al. (2014) found out from EKHUFT data that 90 days following discharge 0.56% of AKI patients were on RRT. After excluding those patients that might have anyway progressed to CKD requiring dialysis even in the absence of AKI, because they already presented CKD stages 4-5, then 0.26% of patients who developed AKI required RRT 90 days after discharge.

Pannu et al. (2013) analysed renal outcomes among patients who survived AKI (KDIGO stage 2 or 3), defining renal recovery as a post-AKI SCr within 25% of the baseline (pre-hospitalization). 3,231 patients were followed after discharge over a 34-month period: 30.8% of AKI survivors died and 2.1% progressed to kidney failure (CKD stage 5). Additionally, participants who did not recover renal function had higher risk of mortality and adverse renal outcomes (such as progression to CKD), as measured by hazard ratios (HR)[6] in comparison with patients who recovered kidney function: mortality HR was estimated at 1.26 and adverse renal outcomes HR at 4.13.

4.1 AKI-ASSOCIATED COSTS

When it comes to evaluating the economic impact and financial burden of AKI, the main problem is certainly the absence of a uniform definition of AKI and, consequently, an

[6] Hazard Ratios measure the instantaneous risk at a particular time. They differ from risk ratio in the sense that the latter describe a cumulative risk.

accurate estimate of AKI incidence (see Chertow et al. 2005). Nevertheless, some single-centre studies have been conducted, so far, with the aim of evaluating AKI-associated costs. Before summarizing the main findings that can be drawn from these researches, it is worth noting that many of them actually focused on rather tightly defined patient groups, such as patients developing AKI after cardiac surgery or CABG (Coronary Artery Bypass Graft), and that additional research should be done to extend findings to a more heterogeneous population (see Lewington and Hall 2014). Besides this fact, Lewington and Hall (2014) also argue that one limitation of these studies is that AKI-related cost may not be fully AKI-attributable cost: to capture the actual burden of AKI, it would be important to discern those costs that would have occurred also in the absence of AKI. The same reasoning applies also in case of lifetime post-discharge costs of care for patients who developed AKI during hospitalization.

Dasta et al. (2008) quantified the total postoperative costs and resource usage associated with AKI following CABG surgery, classified according to RIFLE criteria[7]; this study provided also an insight on clinical outcomes deriving from AKI, such as length of stay, need for RRT and hospital mortality. Costs and outcomes of 258 AKI patients were compared to a control group of patients, matched on the basis of preoperative (baseline) characteristics. Estimation of costs focused on the postoperative period, so preoperative treatments were not included.

Table 4.2. Costs and outcomes of patients with and without AKI after CABG (2002 $)

	Postoperative costs	**ICU costs**	**Postop. LOS (days)**	**ICU LOS (days)**	**RRT (%)**	**Mortality (%)**
AKI patients	$37,674	$25,949	11.0	3.2	3.5	11.2
n=258	(23,654–68,433)	(15,47–59,175)	(7–18)	(1.5–6.5)		
Controls	$18,463	$13,836	5.0	1.4	0.4	2.3
n=258	(14,704–23,822)	(11,165–19,923)	(4–7)	(1.1–2.0)		

Medians (interquantile ranges). All values are statistically different (p-value<0.001).
AKI: acute kidney injury; ICU: intensive care unit; LOS: length of stay; RRT: renal replacement therapy.
(Source: adapted from Dasta et al., 2008)

As Table 4.2 displays, costs and adverse clinical outcomes for AKI patients were significantly higher than those for control patients. The 258 patients with AKI generated a cumulative cost of $18.3 million. Findings concerning length of stay, requirement for RRT

[7] See Chapter 2 for a description of existent classifications for AKI.

and mortality show that these negative clinical outcomes statistically increase in AKI patients; all these results are in line with those of previously cited studies. Moreover, Table 4.3 reports resource usage and clinical outcomes stratified by RIFLE categories, showing how costs significantly rise as AKI gets more severe (from RIFLE-R to RIFLE-F category); in a similar way, clinical outcomes get worse. It appears clearly that the main driver in postoperative costs is the prolonged length of stay and the need for critical care assistance: ICU costs represent between 73.3% and 93.7% of total postoperative costs, depending on severity of AKI.

Table 4.3. Resource usage and outcomes of three RIFLE categories in patients with AKI after CABG (2002 $)

	RIFLE-R	**RIFLE-I**	**RIFLE-F**
Incidence (%)	53.5	27.1	19.4
Total postoperative costs	$29,697 (20,041–52,351)	$38,924 (25,092–70,424)	$52,618 (35,250–91,954)
ICU costs	$21,775 (13,444–41,427)	$28,872 (17,961–63,322)	$49,328 (21,454–83,687)
ICU costs (% of total costs)	73.3	74.2	93.7
Total postoperative LOS days	9 (6-17)	11 (7-19)	16 (12-25)
ICU LOS days	2.29 (1.34–5.2)	3.45 (1.9–7.7)	5.42 (2.8–12.3)
ICU LOS days (% of total LOS)	25.4	31.3	33.8

Medians (interquantile ranges). All values are statistically different (p-value<0.001).
AKI, acute kidney injury; ICU, intensive care unit; LOS, length of stay.
(Source: adapted from Dasta et al., 2008)

Chertow et al. (2005) analysed data from Brigham and Women's hospital, a medical centre in Boston, to show that even small changes in Serum Creatinine (namely, changes of SCr > 0.3 mg/dl) were associated with increased hospital length of stay, mortality and costs, with respect to a group of controls. Main study results, as regarding costs estimates, are displayed in Table 4.4.

Total annualized hospital costs for the recorded AKI patients represent 5% of total hospital costs and, more interestingly, also milder cases of AKI (non requiring dialysis) caused a substantial financial burden that cannot be ignored from a public health perspective (Chertow et al., 2005).

Table 4.4. Costs associated with selected changes in SCr

Criterion for SCr change (n)	Mean Increase in Total Cost ($)
↑ SCr ≥ 0.3 mg/dl (2892)	$8,902
↑ SCr ≥ 0.5 mg/dl (1236)	$12,656
↑ SCr ≥ 1.0 mg/dl (351)	$21,475
↑ SCr ≥ 2.0 mg/dl (105)	$33,161
↑ SCr by 25% (4060)	$7,469
↑ SCr by 50% (1967)	$10,125
↑ SCr by 100% (714)	$15,192
↑ SCr by 50% to a minimum peak of 2.0 mg/dl (352)	$19,517
↑ SCr≥0.5mg/dl with baseline SCr<2.0mg/dl or ↑ SCr ≥1.0 mg/dl with baseline SCr≥2.0 and <5.0 mg/dl (1160)	$13,451

Results are relative to those without the change indicated, on a total sample of n=9205. SCr: serum creatinine.
(Source: adapted from Chertow et al., 2005)

In the already cited study of Kerr et al. (2014), the economic impact of AKI to the National Health System (NHS) in England for the year 2010-11 was estimated, exploiting data from HES and academic studies (EKHUFT). Cost estimation deriving from HES data were based on Healthcare Resource Group (HRG) tariffs – HRG are groups of health care activities to which a patient admission is assigned according to diagnosis and interventions. A first important finding is that AKI related costs deriving from HES data tend to underestimate the actual AKI economic burden, not only because of the under-recognition of AKI cases typical in HES data, but also because only 16.22% of AKI recorded cases were grouped to a HRG representing AKI, leading to an estimated cost of £75 million. The remaining patients who developed AKI were attributed to a different HRG, reflecting the fact that AKI often comes along with other illnesses and interventions. Costs related to these admissions were calculated according to the excess days of hospitalization related to AKI, for a total additional cost of £304 million. Total inpatient expenditure associated to admissions recording an AKI event amounted, thus, to £380 million. Nevertheless, this estimate is still likely to undervalue the actual AKI-associated cost, because it is based on average cost per bed day for a "standard" patient and excludes critical care use. In-patients care for an individual who

develops AKI, however, might be more expensive due to medications, intensive care and additional therapies required, such as RRT in worst cases. Moreover, this estimate excludes the cost of post-discharge care.

According to Kerr et al. (2014) records from EKHUFT are probably more representative of actual AKI incidence in England (15.33% of total hospital admissions, as compared to 2.43% in HES data). If this were actually the case, then inpatient expenditure related to AKI would amount to £1.02 billion in England, plus post-discharge care costs of £179 million, for a total cost of care of £1.2 billion. Additional long-term costs arising after discharge are due to a higher incidence of CKD and RRT in patients who developed AKI, as compared to non-AKI patients. Table 4.5 below reports how total care costs are computed, based on HES data and on extrapolations from EKHUFT academic studies.

Table 4.5. Estimated expenditure related to AKI, England 2010–11

	HES data	**Extrapolation from EKHUFT**
Admissions to AKI HRG	£75,186,389	£75,186,389
Excess LOS in other HRGs	£304,364,710	£750,463,603
Critical care	No data available	£198,232,502
Total inpatient care	£379,551,099	£1,023,882,494
Post-discharge care	No data available	£179,345,543
Total Care	£379,551,099	£1,203,228,037

LOS: length of stay; HRG: Health Resource Group
Source: adapted from Kerr et al. (2014)

Based on these findings, the financial burden of AKI amounted to 1% of total NHS budget in England during 2010-11. The scope for cost savings is substantial in this framework. According to NCEPOD report in the UK, Kerr et al. (2014) argued that 20% of AKI cases are avoidable. If so many AKI cases were actually prevented, then savings to the NHS would amount to £200 million (0.2% of the total NHS budget in England).

4.2 COSTS OF RENAL REPLACEMENT THERAPY AND LONG-TERM DIALYSIS DEPENDENCE FOLLOWING ACUTE KIDNEY INJURY

As previously mentioned, in the most severe cases AKI may require the provision of a RRT. Indeed, as many as 4-5% of critically ill patients require such therapy during their ICU stay, following AKI (see Manns et al. 2003). Two major modalities of RRT can be

distinguished: continuous RRT (CRRT) and intermittent RRT (IRRT). Each of the two modalities has a set of advantages and disadvantages; each patient may be a candidate for either therapy, depending on his/her characteristics and disease, and the way in which the two therapies are provided can vary widely across regions, due to availability, costs, physicians expertise, hemodynamic stability and the primary purpose of the procedure (fluid removal or solute clearance). So far, there seems to be no clear evidence about which, if any, of the two therapies delivers the higher survival rate (see Srisawat et al. 2010, and Mazairac et al. 2013). Assessment of the outcomes and costs of different RRT is deemed necessary due to high mortality, morbidity and costs associated with AKI and its treatment, in order to allow an efficient allocation of resources.

Hereafter are presented the findings of some studies that focused on the costs of dialysis. Most of them compared different dialysis treatments and are useful to get an insight into the costs arising from AKI that necessitates RRT.

RRT is associated with significant higher hospital costs, longer hospitalization and increased mortality rates as compared to AKI patients not requiring RRT, as Table 4.6 below shows. From these results provided by the research of Dasta et al. (2008) it is clear that costs, length of stay and days in ICU are more than double, on average, for patients who need RRT after AKI, as compared to AKI patients that do not need such treatment.

Table 4.6. Effects of RRT on costs and outcomes in patients with AKI after CABG (2002 $)

	Total postop. costs	**ICU costs**	**Total postop. LOS (days)**	**ICU LOS (days)**	**Mortality (%)**
Cases with RRT n=27	$74,040	$71,511	21.0	6.7	37.0%
Cases without RRT n=231	$34,953	$23,880	10.0	2.7	8.2%
Controls n=258	$18,463	$13,836	5.0	1.4	2.3%

Median values. All values are statistically different (p-value<0.001).
AKI: acute kidney injury; ICU: intensive care unit; LOS: length of stay; RRT: renal replacement therapy.
(Source: adapted from Dasta et al., 2008)

De Smedt et al. (2010) looked at treatment modalities for AKI with the purpose of determining the preferable method, accounting for the associated health costs. The authors considered not only CRRT and IRRT, but also a conservative (CONS) therapy that did not encompass dialysis. The authors found that CRRT was the most expensive treatment with a

mean total cost (including hospitalization costs and follow-up costs over a two-year period) of €51,365, and that both RRT lead to additional expenditure between €9,643 and €17,563 compared with CONS (mean total cost €33,802). However, the additional cost of a RRT in comparison with CONS (or of one type of RRT with respect to the other) should be evaluated together with the benefits associated with each therapy (improvement in clinical outcomes, reduction in downstream costs etc.).

Many patients recover renal function after developing AKI but not all of them. Given the high number of critically ill patients who require RRT, failure to recover kidney function and progression to end-stage kidney disease (ESKD) that leads to dialysis dependence represents a relevant medical and economic issue.

Recently, Ethgen et al. (2015) hypothesized that the choice of initial treatment of AKI – namely, CRRT or IRRT – could influence long-term outcomes in terms of dialysis dependence. In particular, based on a recent meta-analysis and systematic review from Schneider et al. (2013), the authors investigated the possibility that initial treatment with CRRT could be associated with a lower rate of dialysis dependence among survivors and, consequently, could be economically superior to IRRT. First of all, it should be noted that the upfront cost of CRRT is higher than IRRT – daily cost is respectively $858 and $226, as estimated by Manns et al. (2003). Ethgen et al. (2015) reported results of other studies that estimated, among patients affected by AKI, a rate of survival of 37% at 180 days from discharge and a rate of dialysis dependence at 3 years of 21.7 with CRRT and 26.6 with IRRT. Consequently, a five-year cost-effectiveness analysis showed that, despite the lower upfront average cost of IRRT, the total cost over a five-year time horizon was higher for IRRT patients as a result of the greater rate of dialysis dependence – meaning that CRRT is cost-effective as compared to IRRT (five-year ICER of CRRT with respect to IRRT is – $116,121).

However, these results appear in contradiction with previous studies that found no health or economic advantage of CRRT over IRRT, so there still seems to be no clear-cut evidence in favour of either treatment modality.

4.3 IS ACUTE KIDNEY INJURY AN AVOIDABLE COMPLICATION?

From the results presented in the previous paragraphs, it appears clearly that AKI has a substantial economic burden, deriving from longer hospital stays, need for critical care and RRT, as well as from long-term adverse clinical outcomes.

However, recent debates have focused on the idea that AKI is, in many cases, an avoidable complication: even 20% to 30% of AKI cases are predictable and avoidable, making the high incidence of this disease unacceptable (Kerr et al. 2014, and Taylor 2011). Of course this requires accurate monitoring of patient's physiological parameters (including those related to kidney functions) and possibly calculation of risk scores; clinicians should be trained so that they are capable of implementing the appropriate measures, such as reviewing medications if necessary, suspending nephrotoxic drugs, ensuring proper patient's hydration, treating infections and other illnesses promptly. Consulting a specialist should also be highly recommended in case a patient develops AKI, in order to define promptly a therapy and ensure he/she is treated properly. Many AKI cases, indeed, are mild and associated with the temporary loss of 20-30% of kidney function and could be treated more effectively when identified and addressed early.

According to Lewington and Hall (2014), attributing causality to AKI for mortality and additional costs is fundamental to target interventions aimed at reducing the burden of AKI. Anyway, early detection and timely intervention may lead to substantial savings: as Kerr et al. (2014) estimated, if 20% of AKI cases where prevented, these savings to the NHS in England could amount to £200 million (0.2% of the total NHS budget).

On the one side, implementing prevention principles, adopting basic measures of care and ability to recognize risk of kidney injury on behalf of non-specialists are certainly crucial steps in AKI prevention. On the other side, diagnostic innovations and screening methods offer a massive opportunity for the early detection of AKI. In this setting, new biomarkers and diagnostic tests based on them, such as the NephroCheck® Test, may play a crucial role, in that they may allow quick recognition of AKI. Indeed, the NephroCheck® Test may signal increased risk of AKI as soon as 12 hours after cardiac surgery and 24 to 48 hours before SCr increases.

The following chapter will present the results arising from the retrospective study that took place AT the San Bortolo Hospital of Vicenza, with the aim of evaluating the NephroCheck® Test as a diagnostic tool for AKI on a population of patients undergoing cardiac surgery.

5 PREDICTING THE PROBABILITY OF DEVELOPING ACUTE KIDNEY INJURY AFTER CARDIAC SURGERY

5.1 STUDY DESIGN AND AIM OF THE STUDY

The aim of this dissertation is to evaluate the predictive ability of the NephroCheck® Test for an early diagnosis of AKI in patients undergoing cardiac surgery (CSA-AKI: Cardiac Surgery-Associated AKI). A necessary step consists in comparing the efficacy and timeliness of the diagnosis of CSA-AKI with the NephroCheck® Test to the current methods based on Serum Creatinine (SCr). To do so, we analysed a population of patients at risk of developing AKI for whom the NephroCheck® Test was not available and another sample of patients with comparable epidemiological and clinical characteristics, on whom the NephroCheck® Test was used. The accuracy of the diagnostic test was assessed by means of computing the area under the receiver operating characteristic curve (ROC curve) and other measures of performance.

Similarly to most innovative technologies, the NephroCheck® Test is an expensive tool, especially if compared to current diagnostic methods that are already part of clinical routine analysis based on SCr and urine output. Therefore, even if it is effective, it might be excessively resource consuming to be used on a routine basis, especially considering the current trend characterized by cuts in healthcare spending. This innovation could consequently be employed more parsimoniously, in all those cases where there is a higher risk of developing AKI and a timely diagnosis could allow interventions that prevent or limit progression of the disease. The identification of such cases, within a broader population of patients presenting some risk of developing AKI, can take place by building a model for AKI prediction and defining a rule for selecting those cases where the NephroCheck® Test should be used. Alternatively, introducing the NephroCheck® Test in the clinical routine may deliver substantial advantages, in terms of early AKI diagnosis and detection, that could allow more careful management of all patients at risk. This means that some AKI cases might be avoided, disease severity might be lowered and intervention might prevent disease progression in the worse situations, thus ultimately leading to cost savings. Investments required to employ the NephroCheck® Test on a routine basis could be then offset by the lower resource consumption in patient care related to AKI.

5.2 STUDY SAMPLE AND DATA COLLECTION

For the purpose of this dissertation two cohorts of patients undergoing cardiac surgery in the San Bortolo Hospital of Vicenza were considered. First of all, the base model was developed on a sample of cardiac surgery patients during the second semester of 2014; then, another group of patients that underwent elective cardiac surgery along 2015 and for whom the NephroCheck® Test was available was considered. The protocol of the study was presented to and approved by the local Ethical Committee at the San Bortolo Hospital of Vicenza.

The selection of patients to be included in the study followed the criteria that are described here. Inclusion criteria consisted of selecting (a) only 18 years or older individuals (b) who underwent cardiac surgery procedures. Patients were, however, excluded from the study if (c) they underwent a surgery procedure of TAVI type (Transcatheter Aortic Valve Implantation) (d) they were already receiving a dialysis treatment at the moment of hospitalization.

For the 2014 cohort, the data of 332 patients were collected. For the 2015 cohort, the patients had to give their prior informed consent to undergo an additional medical examination (the NephroCheck® Test), consisting of a urine test at several time points before and after surgery. Samples for the NephroCheck® Test assessment were collected at hospital admission (baseline), 30 minutes after the initiation of ECC (extra-corporeal circulation, a type of surgical procedure), at CS-ICU (cardiac surgery intensive care unit) admission, 4 hours, 12 hours and 24 hours after surgery, at hospital discharge and, finally, during a follow-up that took place after three months. Therefore, patients that underwent urgency or emergency surgery could not be included: only patients operated on an elective list and giving their informed consent to the additional test were enrolled in the 2015 cohort, for a total of 110 patients.

All data were collected according to a case report form in Excel format. The case report form was drawn up by a team of physicians and biologists, on the basis of experience, knowledge and existing literature on the topic, in order to include the parameters that had already been shown to or are expected to correlate with AKI. These parameters include anamnestic characteristics of the patients, pre-operative treatments (within 7 days before operation), surgery procedure, post-operative treatments and clinical data (up to 24h after surgery).

The 2014 cohort was then used to develop a model for predicting AKI after cardiac surgery. The model included, as explanatory variables, anamnestic and preoperative information as well as data on intra- and postoperative therapies. The aim was to build a logistic regression model for predicting the probability that a patient develops AKI after cardiac surgery, based on the aforementioned features. This model was then applied to the 2015 cohort and a variable containing the results from the NephroCheck® Test was added to it. This allowed to assess if the NephroCheck® Test improves predictive ability of the model compared to the base model, which does not account for the NephroCheck® Test results.

The outcome and explanatory variables considered in the regression models to be estimated are listed and described in the following paragraphs.

5.3 OUCOME VARIABLE AND INDEPENDENT PREDICTORS

The aim of this analysis is to estimate the probability $\Pr(Y = 1|\boldsymbol{x})$ that an event Y occurs, depending on p explanatory variables $\boldsymbol{x}' = (x_1, x_2, \ldots x_p)$ that might be either categorical or continuous predictors.

Hereafter, the outcome variable Y and the covariates are described in details.

- The dependent variable *Y* is *akipost*: AKI development after cardiac surgery (within 10 days). It is a binary variable, coded as 1 if the patient develops AKI and 0 if he/she does not. For the purpose of this dissertation, development of AKI was calculated according to the serum creatinine (SCr) criteria described in Table 2.2, Chapter 2: baseline creatinine is the lowest SCr value recorded during the whole hospitalization period and an increment of ≥0.3 mg/dl, occurring within 10 days after cardiac surgery, was considered for recording an AKI event. However, the first post-operative SCr measurement, which takes place at intensive care unit (ICU) admission, was not considered for the purposes of detecting all AKI cases. The reason is that the blood (more specifically, plasma) concentration of creatinine could be affected downward by the hemodiluition resulting from the intraoperative infusion therapy. Considering this SCr value may, therefore, lead to "over-identification" of AKI events, in other words, to the identification of more events than actual AKI cases according to KDIGO criteria.

The covariates included in the final model are the following:

- *akipre* is a dichotomous predictor that assumes value of 1 if the patient develops AKI before undergoing cardiac surgery, during hospitalization. A patient that has just

developed AKI may be susceptible to subsequent renal injuries, since his/her kidneys have recently experienced stressful conditions. As described later on in this chapter, very few patients developed preoperative AKI, as it is more likely to occur in most severe cases.

- *female* is another binary covariate that describes the gender category (1 if the patient is female, 0 if male).
- *agecut2*: patient's age. Firstly considered as a continuous variable, age was then converted into age groups because the risk of developing AKI is likely to increase with age in a non-linear way. In particular, it was decided to divide the cohorts of patients into three age groups, since AKI risk might increase substantially after the age of 65 and even more around the age of 75. Age groups are defined for individuals younger than 65 (reference group), individuals with an age between 65 and 74 (extremes included), and patients who are 75 or older.
- *obese:* a binary variable that takes value 1 if the patient is affected by obesity, as calculated through the body mass index (BMI) formula at the moment of hospitalization. The BMI is defined as the body mass divided by the square of the height (expressed in units of kg/m^2). If BMI assumes a value above 30, then the subject is considered obese. This covariate might be considered a proxy for a more general anamnestic health status, since other dysfunctions or metabolic disease are likely to be present along with obesity (such as dyslipidaemia, hypertension, diabetes etc.). Other variables denoting an individual's health status were also considered at the beginning of the analysis, but they were excluded from the final model either because uninformative – sometimes due to the high rates of missing values or misreporting – or because they turned out to be statistically not significant in the univariate and/or multivariate regression model. However, the basic idea that lies behind the inclusion of this variable is that more severe health conditions are likely to be positively correlated with AKI.
- *ckdanamnesi* indicates if the patient, at hospitalization, has a moderate to severe form of chronic kidney disease (CKD)[8]. As mentioned previously in this chapter, dialysed patients (CKD Stage 5) were excluded from the study sample. This variable is binary and, consequently, is coded as 1 if the patient presents a CKD Stage 3 or 4, 0 otherwise.

[8] See Chapter 2, Table 2.1, for CKD (chronic kidney disease) classification criteria.

- *surtype:* type of surgery. This is a categorical variable, obtained from several binary variables indicating which surgical procedure was employed. In particular, it tells if the patient underwent only a "single heart valve (aortic, mitral or tricuspid) surgery procedure" (0, *svalve*), only a "multiple heart valve surgery" (1, *multvalve*), "CABG (Coronary Artery Bypass Graft) surgery" (2, *cabg*) and "replacement of the ascending aorta" (3, *aorta*). Different surgery procedures may affect in a different way the severity of patient conditions, the treatments required after surgery and the risk of renal complications, which might be particularly high in the case of bypass procedures (see Mariscalco et al. 2011, and Mao et al. 2014). Indeed, patients that undergo CABG surgery have severe coronary heart disease, which can reduce blood flow to the kidney, eventually causing kidney failure if untreated.
- *eccdur*: is a continuous variable that describes the duration of the ECC (Extra-Corporeal Circulation), expressed in minutes. The ECC is a procedure that consists of diverting a patient's blood flow through an artificial circuit and pumping the blood back into the patient's body. Cardiopulmonary bypass, extracorporeal membrane oxygenation and hemodialysis are some of the most common forms of ECC. It is a well-known risk factor for postoperative AKI (Jorge-Monjas et al. 2009, and Mao et al. 2014) since its duration is likely to correlate positively with the risk of developing AKI.
- *postcrea*: this is the value of postoperative SCr that was excluded for the purpose of defining the outcome variable, as described earlier in this chapter. The value of this variable is influenced by many factors, not only those that commonly affect SCr value[9], but also by the surgical therapies, such as the amount of fluids that are administered to the patient. In developing their AKI risk score, Jorge-Monjas et al. (2009) showed that postoperative SCr is positively correlated with subsequent AKI development.
- *lasix20* expresses the number of phials of a diuretic administered during the first 24 hours following surgery, each phial containing 20 mg. The reason for the inclusion of this variable is that it is believed that intraoperative and early postoperative variables impact the clinical course of the patient (see Jorge-Monjas et al. 2009) and, in turn, affect the postoperative treatments undertaken by clinicians. Administration of post-operative diuretics usually takes place in case of excess fluid retention. Indeed, the

[9] Chapter 2 describes in more details the features of SC as a biomarker.

patient may be in a situation of fluid overload, due to the high volume of fluids that he/she has been administered during surgery. When this excess volume is very high, it might be threatening for the heart and the kidneys; therefore excretion of fluids is stimulated through diuretics. Moreover, in case the kidneys are not working properly, diuretics help kidney functions by means of stimulating excretion of fluids, therefore administration of diuretics might signal a higher risk of renal injury.

- *nc12h* and *nc12hcut03:* these two variables refer to the results of the NephroCheck® Test, measured 12 hours after surgery. Therefore, they are available only for the second cohort of patients (year 2015). The first variable (*nc12h*) is continuous, while the other one (*nc12hcut03*) is a categorical variable that is defined according to the cut-off value already derived in previous studies (Ronco 2014). Recalling what explained in Chapter 2, the NephroCheck® Test delivers a number in a range between 0.04 and 10.0 and a lower cut-off of 0.3 has been identified as it allows early recognition of most patients that are at risk of developing AKI within the following 12 hours (see Kashani et al. 2013, and Hoste et al. 2014). The binary variable *nc12hcut03* is, thus, coded as 1 if the NephroCheck® Test delivers a value which is higher than 0.3, it is coded as 0 in all the other cases.

In the beginning, other variables were considered to be included in the model as covariates because they are considered risk factors for AKI:

- *nonelective:* as explained before, in the 2014 cohort there are some patients who underwent urgency or emergency surgery procedures, which are likely to increase the risk of developing more adverse clinical conditions thereafter. For this group of patients, *nonelective* assumes the value of 1. The 2015 cohort, conversely, does not include any urgent/emergent patient by design. In the beginning *nonelective* was entered also in the multivariate model but then it was dropped, as explained in details further in this chapter.

The following variables, differently from *nonelective*, were not significant in the univariate analysis (see paragraph 5.6) but were included in the multivariate model at the beginning of the analysis. They were excluded, later on, because non significant also in the multivariate model:

- *lactate:* is the value of post-operative lactate. Jorge-Monjas et al. (2009) included lactate in their post-operative AKI risk score, so this variable was accounted for at

the beginning of this analysis.

- *clampdur* is the duration (in minutes) of the aortic cross-clamping, a particular surgical procedure that may be harmful for kidneys, especially when prolonged.
- *diabetes*: a dummy for the presence of diabetes, a disease which is a risk factor for AKI.
- *hypertens* takes value 1 if the patient is affected by hypertension – another risk factor for AKI.
- *mv12h*: a dummy that tells if the patient is still mechanically ventilated after 12h from surgery, it should signal that patient is recovering slowly as he/she is unable to breathe autonomously.
- *my,* stands for myocardial infarction. If the patient had a myocardial infarction episode, this binary variable assumes value 1. Patients hospitalized with a myocardial infarction or that had a previous episode are at higher risk of developing AKI (Kosiborod 2012).

5.4 DESCRIPTIVE STATISTICS

The descriptive statistics of the variables listed above for the two study groups considered for the purpose of this dissertation are presented in Table 5.1. It can be noticed that some of these variables present missing values, thus leading to a reduction in the sample size for the estimation of the regression model.

Table 5.1 Descriptive Statistics

	2014				2015			
Continuous Variables	NON-AKI		AKI		NON-AKI		AKI	
	Mean±SD	Obs	Mean± SD	Obs	Mean± SD	Obs	Mean±SD	Obs
age	64.74±12.38	253	72.38±9.02	79	59.37±12.73	95	66.93±11.01	15
eccdur	109.69±37.83	243	120.67±51.0	76	118.67±33.8	93	148.77±59.16	13
postcrea	.84±.201	250	1.10±.42	79	.77±.15	92	1.03±.29	13
lasix20	1.35±2.07	253	4.92±7.23	79	.84±.97	95	2.03±4.13	15
lactate	2.15±1.38	246	2.49±1.65	76	2.09±1.18	91	2.75±2.20	13
clampdur	74.53±30.61	230	77.06±33.66	69	84.02±31.25	92	92.75±32.13	12

Categorical Variables	NON-AKI		AKI		NON-AKI		AKI	
	n	%	n	%	n	%	n	%
akipre	4	1.58	7	8.86	0	0	0	0
female	79	31.23	27	34.18	28	29.47	4	26.67
obese	32	12.65	23	29.11	9	9.47	5	33.33
nonelective	46	18.18	23	29.11	0	0	0	0
ckdanamnesi	10	3.95	9	11.39	0	0	0	0
svalve	143	56.52	23	29.11	70	73.68	7	46.47
multvalve	22	8.7	6	7.59	7	7.37	1	6.67
cabg	73	28.85	42	53.16	8	8.42	4	26.67
aorta	15	5.23	8	10.13	10	10.53	3	20
mv12h	46	18.18	35	44.30	11	11.58	5	33.33
diabetes	36	14.23	16	20.25	7	7.37	3	20.00
my	33	13.04	14	17.27	1	1.05	0	0
hypertens	177	69.96	63	79.75	53	55.79	11	73.33

To check whether the two cohorts of patients – 2014, "no NephroCheck® Test cohort", and 2015, the "NephroCheck® Test cohort" – were comparable and to assess if there existed any significant difference in the distribution of the selected covariates between the samples, a *two sample t-test* was run, for continuous variables, or a two sample test of proportions, for categorical variables. A *t-test* robust to unequal variance was run when equal variances could not be assumed. Results are presented in Table 5.2 (*p*-values in the last column).

Table 5.2 Patient population profile by cohort:
2014 no- NephroCheck® Test and 2015 NephroCheck® Test

Variable	2014			2015			P
	Obs.	Mean	Std Dev	Obs.	Mean	Std Dev	
age	332	66.55559	12.10608	110	60.4	12.72662	0.0000
eccdur	319	112.3072	41.52865	106	122.3585	38.72644	0.0287
postcrea	329	.9039514	.2899141	105	.8047619	.1925392	0.0001
lasix20	332	2.196551	4.230015	110	1	1.780372	0.0000

	Obs	n	Mean	Obs	n	Mean	P
akipost	332	79	0.238	110	15	0.136	0.0240
akipre	332	11	0.033	110	0	0	0.0532
female	332	106	0.319	110	32	0.291	0.5779
obese	332	55	0.166	110	14	0.127	0.3363
nonelective	332	69	0.208	110	0	0	0.0000
ckdanamnesi	332	19	0.572	110	0	0	0.0103
svalve	332	166	0.50	110	77	0.70	0.0003
multvalve	332	28	0.843	110	8	0.727	0.6996
cabg	332	115	0.346	110	12	0.109	0.0000
aorta	332	23	6.93	110	13	11.82	0.1041

P-values (two-sided) are obtained from two sample *t-test*, in case of continuous variables, and a two sample test of proportions, for categorical variables.

There are, indeed, some variables that present statistically different (p-value<.05) – even if the difference is not dramatically large – averages between the two cohorts. On average, the 2014 cohort is older, undergoes surgical procedures that tend to last less, shows higher values of postoperative creatinine (*postcrea*) and is administered more diuretics (*lasix20*) in the early post-surgical period. As far as the difference in administration of diuretics is concerned, it is worth reminding that the variable *lasix20* is expressed in terms of number of phials administered to the patient; from a clinical point of view, a difference of one phial is not substantial. Moreover, within the 2014 cohort, a higher proportion of patients underwent CABG procedures or was already affected by CKD (stage 3 or 4) at the moment of hospitalization.

At this point it is important to recall what has been briefly explained in the previous paragraphs, concerning the selection of the two samples and how this might influence patient population profile. Differently from the no-NephroCheck® Test cohort (2014), the NephroCheck® Test group (2015) does not include patients that underwent urgency or emergency surgery. This not only affects the distribution of the variable *nonelective*, which is by definition 0 for all the NephroCheck® Test cohort individuals, but also other features. Among the 2014 cohort, there are some patients showing a higher risk profile, from a medical point of view, just because their clinical conditions at the moment of surgery are urgent. For example, this might explain why, in this cohort, there is a significantly higher proportion of patients that develop AKI after surgery (*akipost*) and more CABG cases (*cabg*), a surgical procedure that is common in urgency cases.

Basically, these findings suggest that some differences exist between the two cohorts, as far as the distribution of the covariates of interest is concerned, and this could interfere with the aim of the study – estimating the ability of the NephroCheck® Test in predicting the probability of developing AKI after cardiac surgery. Indeed, in estimating the effect of the NephroCheck® Test on the 2015 cohort, we might find a result that is actually due to underlying existing differences between the two cohorts, if these variables correlate with the outcome *akipost*. However, this issue is addressed by means of controlling for these confounders in the multiple regression model, illustrated in the following paragraphs, thus estimating the effect of the NephroCheck® Test other things being equal.

5.5 MODEL ESTIMATION AND ESTIMATION RESULTS

Recalling the equations described in Chapter 3, the model to be estimated is a multiple logistic regression model:

$$Y = \pi(\boldsymbol{x}) + \varepsilon$$

where the probability $\pi(\boldsymbol{x})$ is expressed as

$$\pi(\boldsymbol{x}) = \frac{e^{g(\boldsymbol{x})}}{1 + e^{g(\boldsymbol{x})}}$$

and $g(\boldsymbol{x})$ is the logit transformation

$$g(\boldsymbol{x}) = \ln\left[\frac{\pi(\boldsymbol{x})}{1 - \pi(\boldsymbol{x})}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p.$$ [10]

[10] See Chapter 3 for the theoretical framework on multiple logistic regression models.

The vector of parameters $\boldsymbol{\beta}' = (\beta_0, \beta_1, \dots, \beta_p)$ is estimated through the method of *maximum likelihood*, by means of maximizing the *log-likelihood function*

$$L(\boldsymbol{\beta}) = \ln[l(\boldsymbol{\beta})] = \sum_{i=1}^{n}\{y_i \ln[\pi(\boldsymbol{x}_i)] + (1 - y_i)\ln[1 - \pi(\boldsymbol{x}_i)]\}.$$

In order to identify the covariates to include in the multivariable logistic regression model, the method of purposeful selection of covariates has been applied, as described by Hosmer, Lemeshaw and Sturdivant (2013). In the beginning, a wide group of variables of interest was selected because they were thought to be risk factors or to correlate with AKI after cardiac surgery, according the literature and/or physicians' experience. The association of these potential predictors with the outcome was analysed through univariate logistic regression: regression coefficients, *p*-values and odds ratio were computed.

The odds ratio (OR) for a dichotomous predictor x_1 is defined as

$$\text{OR} = \frac{\dfrac{\pi(1)}{[1 - \pi(1)]}}{\dfrac{\pi(0)}{[1 - \pi(0)]}} = e^{\beta_1}$$

and it represents the odds that the outcome is present ($y = 1$), given that $x_1 = 1$, as compared to the odds that the outcome occurs in case $x_1 = 0$. For continuous predictors, the odds-ratio is defined for a specified change "*c*" in the covariate, while for categorical predictors it is calculated with respect to the reference group.

In the beginning, a multiple regression model was fit with all the predictors described in paragraph 5.3. Table 5.3 displays results from univariate logistic regressions only for the covariates that were statistically significant in the univariate analysis and were selected to fit a successive multivariate model. The other variables described in the previous paragraph were not statistically significant either in univariate and multivariate analysis, so they were not further considered in the model.

Table 5.3. Univariate Logit Regressions

akipost	Coeff	Std. Err.	z	P>z	[95% Conf. Interval]		Odds Ratio
akipre	1.800403	.6409072	2.81	0.005	.5442474	3.056558	6.052083
_cons	-1.240787	.1338094	-9.27	0.000	-1.503048	-.9785252	.2891566
female	.1342006	.2732639	0.49	0.623	-.4013868	.669788	1.143622
_cons	-1.207812	.158044	-7.64	0.000	-1.517572	-.8980511	.2988506
obese	1.042569	.3116252	3.35	0.001	.4317952	1.653344	2.836496
_cons	-1.372811	.1496062	-9.18	0.000	-1.666034	-1.079588	.2533937
nonelective	.61422	.2964885	2.07	0.038	.0331131	1.195327	1.848214
_cons	-1.307367	.1506256	-8.68	0.000	-1.602588	-1.012146	.2705314
ckdanamnesi	1.139206	.4790742	2.38	0.017	.2002376	2.078174	3.124286
_cons	-1.244566	.1356501	-9.17	0.000	-1.510435	-.978697	.2880658
surtype							
1	1.695652	.8689162	1.03	0.303	.621082	4.629399	1.695652
2	3.577129	1.061024	4.30	0.000	2.000124	6.397529	3.577129
3	3.315942	1.631693	2.44	0.015	1.264014	8.698854	3.315942
_cons	.1608392	.0361338	-8.13	0.000	.1035529	.2498165	.1608392
agecut2							
65-	.9519779	.3620571	2.63	0.009	.242359	1.661597	2.590829
75-	1.674355	.3505258	4.78	0.000	.9873368	2.361373	5.335351
_cons	-2.08833	.283334	-7.37	0.000	-2.643655	-1.533006	.1238938
eccdur	.0060719	.0030491	1.99	0.046	.0000956	.0120481	1.00609
_cons	-1.860188	.3816048	-4.87	0.000	-2.608119	-1.112256	.1556434
postcrea	3.44328	.5897006	5.84	0.000	2.287488	4.599072	31.28941
_cons	-4.392156	.5820989	-7.55	0.000	-5.533049	-3.251263	.012374
lasix20	.2695997	.064497	4.18	0.000	.1431878	.3960116	1.30944
_cons	-1.755123	.1828338	-9.60	0.000	-2.113471	-1.396776	.1728859

In Table 5.4 are displayed the estimated results from the multiple logistic regression model on the cohort of 2014 – no NephroCheck® Test - including all the covariates present in Table 5.3.

Table 5.4. Logit Regression (1)

Logit Regression				**Number of obs = 316**		
				LR chi2(12) = 107.73		
				Prob > chi2 = 0.0000		
Log likelihood = -120.4593				Pseudo R2 = 0.3090		
akipost	**Coeff.**	**Std. Err.**	**z**	**P>z**	**[95% Conf. Interval]**	
akipre	.7007932	.7985516	0.88	0.380	-.8643391	2.265926
female	.811344	.3872266	2.10	0.036	.0523937	1.570294
agecut2						
65-	.7084455	.4342146	1.63	0.103	-.1425995	1.55949
75-	1.245737	.4352451	2.86	0.004	.3926721	2.098802
nonelective	.264335	.405631	0.65	0.515	-.5306872	1.059357
ckdanamnesi	-1.380507	.8135363	-1.70	0.090	-2.975009	.2139944
obese	.9229286	.4019929	2.30	0.022	.135037	1.71082
eccdur	.0000795	.0041153	0.02	0.985	-.0079863	.0081453
surtype						
1 multvalve	.2055498	.6978997	0.29	0.768	-1.162308	1.573408
2 cabg	1.437388	.3961625	3.63	0.000	.6609241	2.213852
3 aorta	1.33304	.6573812	2.03	0.043	.0445966	2.621483
postcrea	3.239962	.7773735	4.17	0.000	1.716338	4.763586
lasix20	.1996874	.0647919	3.08	0.002	.0726976	.3266771
_cons	-6.425256	.9638716	-6.67	0.0000	-8.314409	-4.536102

After fitting a multiple regression model containing the identified potential predictors, some turned out to be statistically not significant in the multivariable model; therefore it was considered whether dropping these variables from the model.

In particular, some of the findings presented in the table above were different from what could be expected, based on the literature. For example, the duration of ECC (*eccdur*) turns out to be not significant at all. However, this could partially be explained by the fact that this sample cohort includes a lot of valve surgical procedures (*svalve* and *multivalve* categories of *surtype*). Specifically, at the San Bortolo Hospital, most of these valve surgeries are performed through the so-called minimally invasive procedure, which involves tiny incisions and should lead to a quicker and less discomfort patient's recovery. Indeed, these types of surgery are usually not as dangerous for the kidney as a CABG procedure (*cabg*) and tend to last less. Secondly, many of the studies conducted so far with the aim of studying and identifying the risk factors of AKI after cardiac surgery have considered CABG patients only. These patients are characterized by a reduced heart function that, in turn, affects negatively the kidneys in the sense that kidneys suffer from reduced blood pumping and the risk of AKI increases. Therefore it is reasonable to assume that the duration of ECC may be a particularly relevant factor in this subpopulation.

Also the effect of a preoperative AKI event is not significant in this sample (*akipre*), which might be due to the fact that a very small proportion of patients (3.31%) reported this feature. These covariates were retained in the multivariable model, even if statistically not significant according to the p-value of their Wald statistic, because they improved the model goodness of fit and/or because of their importance from a clinical viewpoint. In particular, *akipre* was deemed necessary in the multiple regression model for clinical reasons: an individual who experiences AKI before surgery is, in fact, more susceptible to further insults to the kidney.

On the contrary, it was considered whether *nonelective* could be dropped. A likelihood ratio test was computed before taking the final decision. The test is performed by means of calculating the likelihood ratio:

$$lr = -2 \ln \left(\frac{L(m_1)}{L(m_2)} \right)$$

where $L(m_1)$, $L(m_2)$ are the likelihood of the more restrictive and less restrictive model, respectively. Under the null hypothesis the resulting statistic follows a chi-squared distribution, with degrees of freedom corresponding to the number of parameters that are constrained.

Results from the multiple logistic regression model, excluding *nonelective* from the covariates, are presented in Table 5.5.

Table 5.5. Logit Regression (2)

Logit Regression				**Number of obs = 316**		
				LR chi2(12) = 107.31		
				Prob > chi2 = 0.0000		
Log likelihood = -120.66996				Pseudo R2 = 0.3078		
akipost	**Coeff.**	**Std. Err.**	**z**	**P>z**	**[95% Conf. Interval]**	
akipre	.7257285	.8008885	0.91	0.365	-.843984	2.295441
female	.7806902	.3843724	2.03	0.042	.0273341	1.534046
agecut2						
65-	.7234092	.4338026	1.67	0.095	-.1268284	1.573647
75-	1.255784	.4355459	2.88	0.004	.4021301	2.109438
ckdanamnesi	-1.328244	.8052198	-1.65	0.099	-2.906446	.2499578
obese	.894472	.3993891	2.24	0.025	.1116838	1.67726
eccdur	-.0001824	.0041145	-0.04	0.965	-.0082467	.007882
surtype						
1 multvalve	.1726546	.6958329	0.25	0.804	-1.191153	1.536462
2 cabg	1.504977	.3818915	3.94	0.000	.756483	2.25347
3 aorta	1.373589	.6523991	2.11	0.035	.0949103	2.652268
postcrea	3.232229	.7758255	4.17	0.000	1.711639	4.752819
lasix20	.1961158	.0641255	3.06	0.002	.0704321	.3217995
_cons	-6.425256	.9638716	-6.67	0.000	-8.314409	-4.536102

The likelihood ratio test delivers a statistic of 0.42, which is distributed as a chi-squared with one degree of freedom. This corresponds to a p-value of 0.5163, therefore model (2) was chosen as the base model for further analysis. It is important to underline, once more, that this model still controls for all the relevant and statistically significant covariates that are distributed differently between the two cohorts, as showed in previous Table 5.2.

5.6 ASSESSING THE GOODNESS OF FIT OF THE MODEL

Hereafter are presented some regression diagnostics, performed with the aim of checking that model (2) fits sufficiently well, before applying it to the 2015 cohort.

Goodness of fit was, first of all, assessed by means of computing the Hosmer-Lemeshaw statistic, splitting the sample into 10 groups of covariate patterns according to the strategy suggested by Hosmer, Lemeshaw and Sturdivant (2013). The statistic is 6.61 and the associated p-value is 0.5790, which indicates that the model fits sufficiently well.

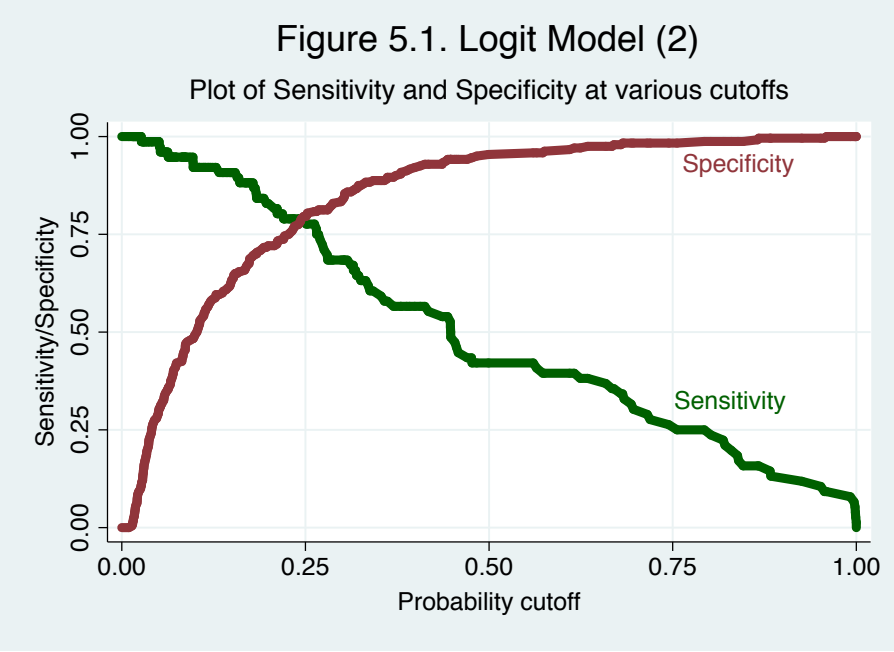
Secondly, it was necessary to ensure that no important variables have been omitted from the model and that the correct link function (namely, the logit function) has been chosen. Consequently, an additional regression was run, where the explanatory variables were the linear predicted values from the original model (*_hat*) and their squares (*_hatsq*). This strategy allows detecting a specification error because, if the variable *_hatsq* is found to be statistically significant, it means that the model is not properly specified. On the contrary, if the model is correctly specified, *_hat* is significant while *_hatsq* should not have much predictive power except by chance. Results from this test for specification error are displayed in Table 5.6.

Table 5.6 Specification Error Test – Logit Regression (2)

Specification Error Test				**Number of obs = 316**		
				LR chi2(2 = 108.60		
				Prob > chi2 = 0.0000		
Log likelihood = -120.02563				Pseudo R2 = 0.3115		
akipost	**Coef.**	**Std. Err.**	**z**	**P>z**	**[95% Conf. Interval]**	
_hat	.9308983	.1256512	7.41	0.000	.6846265	1.17717
_hatsq	-.0569775	.038459	-1.48	0.138	-.1323557	.0184007
_cons	.0719111	.2107614	0.34	0.733	-.3411737	.4849959

The variable *_hat* is statistically significant, which confirms that the relevant predictors have been included in the model, and *_hatsq* is not significant (p-value=0.138), so according to the test we are not in the presence of a specification error.

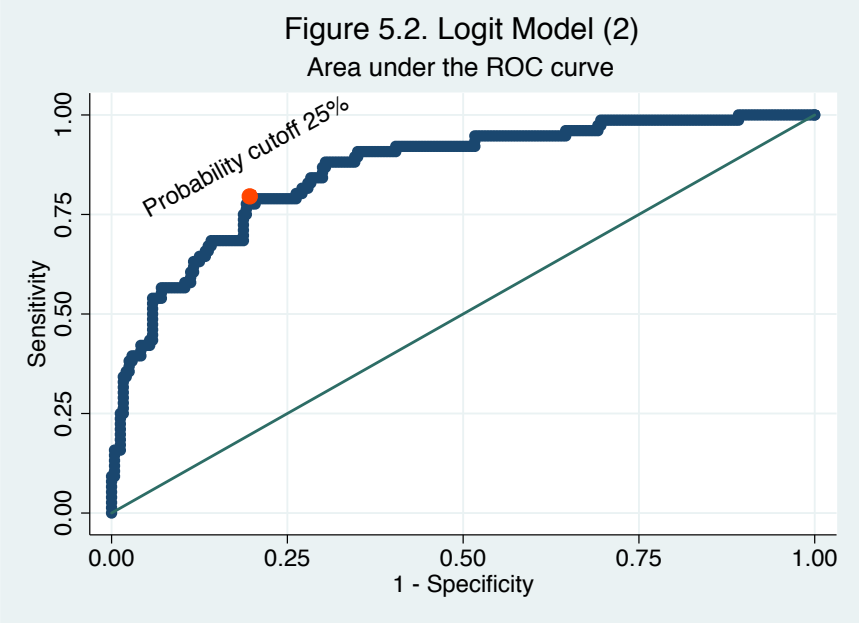
The results of the fitted regression model and, in particular, its performance in terms of classification can be then assessed by means of constructing a classification table. This approach uses the estimated probability to predict group membership and, in particular, a classification table shows the results deriving from cross-classifying the observed values and the predicted values for the outcome variable. In this way, it is possible to compare the actual number of events with the number of events predicted by the logistic regression model. To construct this table, it is necessary to define a cut-point and the estimated probability must be compared to it: if the estimated probability is higher than this specified cut-off, then the predicted value is classified as positive (in this context, AKI); if the is lower than the cut-off, it is classified as negative (non-AKI). Therefore, accurate prediction of group membership provides evidence that the model fits (see Hosmer, Lemeshaw and Sturdivant 2013).



In Figure 5.1, sensitivity and specificity are plotted against all possible probability cut-off points. This kind of graphical representation is useful in order to decide an "optimal" probability cut-off for the purpose of classification. This choice depends also on the objective of classification, such as minimizing the false positives, maximizing the proportion of cases correctly classified, maximizing the negative predictive value and so on. However, according to this plot, an "optimal" choice for a cut-off point could be 0.25, where the sensitivity and the specificity curves approximately cross. Table 5.7 reports the classification table and other summary statistics at the cut-off value of 0.25.

Table 5.7 Classification Table at 25%.
2014 cohort

Classified	Actual		
	AKI	non-AKI	Total
AKI	59	49	108
non-AKI	17	191	208
Total	76	240	316
Sensitivity	77.63%		
Specificity	79.58%		
PPV	54.63%		
NPV	91.83%		
Correctly classified	79.11%		AUC 85.85%



The Receiver Operating Characteristic (ROC) curve is also informative about the ability of the model to discriminate between subjects that develop and do not develop the disease. Figure 5.2 shows the ROC curve for logit model (2), where sensitivity is plotted in function of 1-specificity for each cut-off value. The area under the ROC curve is 85.85%, which suggests that the model performs relatively well in discriminating between AKI and non-AKI patients.

5.7 ESTIMATING THE EFFECT OF THE NEPHROCHECK® TEST

After assessing the model's goodness of fit and finding no evidence of model's misspecification, the same model was applied on the 2015 cohort, at first without including the results of the NephroCheck® Test as a covariate (model 3), then a separate regression was run to assess the improvements in prediction arising from the inclusion of the NephroCheck® Test variable to the base model (model 4).

Table 5.8. Logit Regression (3)

Logit Regression				**Number of obs = 105**		
				LR chi2(12) = 32.23		
				Prob > chi2 = 0.0000		
Log likelihood = -23.202789				Pseudo R2 = 0.4099		
akipost	**Coeff.**	**Std. Err.**	**z**	**P>z**	**[95% Conf. Interval]**	
akipre	(omitted)					
female	1.811228	1.252554	1.45	0.148	-.6437317	4.266189
agecut2						
65-	.3003301	.9742612	0.31	0.758	-1.609187	2.209847
75-	.9525558	1.245209	0.76	0.444	-1.48801	3.393121
ckdanamnesi	0	(omitted)				
obese	-.2083682	1.220265	-0.17	0.864	-2.600043	2.183307
eccdur	.0202696	.0100347	2.02	0.043	.0006019	.0399374
surtype						
1 multvalve	-.5377587	1.501564	-0.36	0.720	-3.480769	2.405252
2 cabg	.8253511	1.053822	0.78	0.434	-1.240103	2.890805
3 aorta	-.9411625	1.379194	-0.68	0.495	-3.644333	1.762008
postcrea	9.48359	3.682992	2.57	0.010	2.265058	16.70212
lasix20	.3802842	.4347449	0.87	0.382	-.4718002	1.232369
_cons	-14.0992	4.091784	-3.45	0.001	-22.11895	-6.079455

Since the two cohorts presented some differences as far as the distribution of the variables of interest is concerned, a test on the equality of regression coefficients was run.
It consists of a cross-model hypotheses testing: a Wald test to assess whether the effect of $x_1, x_2, \dots, x_p$ on the binary outcome is the same across two models – in this setting, model (2) and (3) – that are fit on independent samples, so that the estimators are stochastically independent.
Let the estimated coefficient vector be $\boldsymbol{b}$ and its estimated variance-covariance matrix be $\mathbf{V}$. Let $\boldsymbol{\beta}$ be the true value of the parameter and $\mathbf{R}\,\boldsymbol{\beta} = \mathbf{r}$ the set of p linear hypotheses to be tested jointly, then the Wald test statistic is

$$W = (\mathbf{R}\,\boldsymbol{b} - \mathbf{r})'(\mathbf{R}\mathbf{V}\mathbf{R}')^{-1}(\mathbf{R}\,\boldsymbol{b} - \mathbf{r}),$$

which under the null hypothesis follows a chi-squared distribution with p degrees of freedom. More specifically, the hypotheses to be tested in this case is the equality of coefficients across the model fit on the 2014 sample and on the 2015 sample:

$$H_0: \boldsymbol{b}_{2014} = \boldsymbol{b}_{2015}$$

and the Wald statistic is then

$$W = (\boldsymbol{b}_{2014} - \boldsymbol{b}_{2015})'[\mathrm{Var}(\boldsymbol{b}_{2014}) + \mathrm{Var}(\boldsymbol{b}_{2015})]^{-1}(\boldsymbol{b}_{2014} - \boldsymbol{b}_{2015}),$$

where $\mathrm{Var}(.)$ is the estimated variance-covariance matrix for the coefficients.
Hypotheses testing delivers a Wald statistic of $W = 14.87$, which under the null hypothesis is a value from a chi-squared with 10 degrees of freedom (*cons* excluded). With a *p*-value of 0.1368 the null hypothesis that the coefficients across the two models are the same cannot be rejected.

Given this result, the following step consisted of adding to the model the covariate containing the results of the NephroCheck® Test. First of all, the binary coding of the NephroCheck® Test variable was considered (*nc12hcut03*) and results of the regression model are shown in Table 5.9.

Table 5.9. Logit Regression (4)

Logit Regression				**Number of obs = 105**		
				LR chi2(12) = 39.21		
				Prob > chi = 0.0000		
Log likelihood = -19.712646				Pseudo R2 = 0.4986		
akipost	**Coeff.**	**Std. Err.**	**z**	**P>z**	**[95% Conf. Interval]**	
akipre	(omitted)					
female	1.513295	1.24011	1.22	0.222	-.9172759	3.943866
agecut2						
65-	-.5566859	1.104002	-0.50	0.614	-2.72049	1.607119
75-	.8689912	1.382606	0.63	0.530	-1.840866	3.578848
ckdanamnesi	(omitted)					
obese	-1.21116	1.465763	-0.83	0.409	-4.084002	1.661683
eccdur	.0314445	.0135114	2.33	0.020	.0049627	.0579263
surtype						
1 multvalve	-1.094783	1.538816	-0.71	0.477	-4.110806	1.921241
2 cabg	.4871458	1.182684	0.41	0.680	-1.830872	2.805164
3 aorta	-1.592183	1.571634	-1.01	0.311	-4.672529	1.488163
postcrea	11.96972	4.54354	2.63	0.008	3.064548	20.8749
lasix20	.6151161	.515955	1.19	0.233	-.3961372	1.626369
nc12hcut03	2.967106	1.204397	2.46	0.014	.6065308	5.32768
_cons	-18.18748	5.402325	-3.37	0.001	-28.77584	-7.599119

As expected, the coefficient of *nc12hcut03* is positive and statistically significant (*p*-value 0.014). However, the most interesting and important results can be better understood by comparing the two models (with and without *nc12hcut03*) applied on the same cohort of patients, in particular by looking at the respective classification tables. This helps better understanding how the NephroCheck® Test improves the performance of the model in discriminating patients who will develop AKI after cardiac surgery from those who will not.

Table 5.10. Classification Table at 25%. Without NephroCheck® Test – 2015 cohort

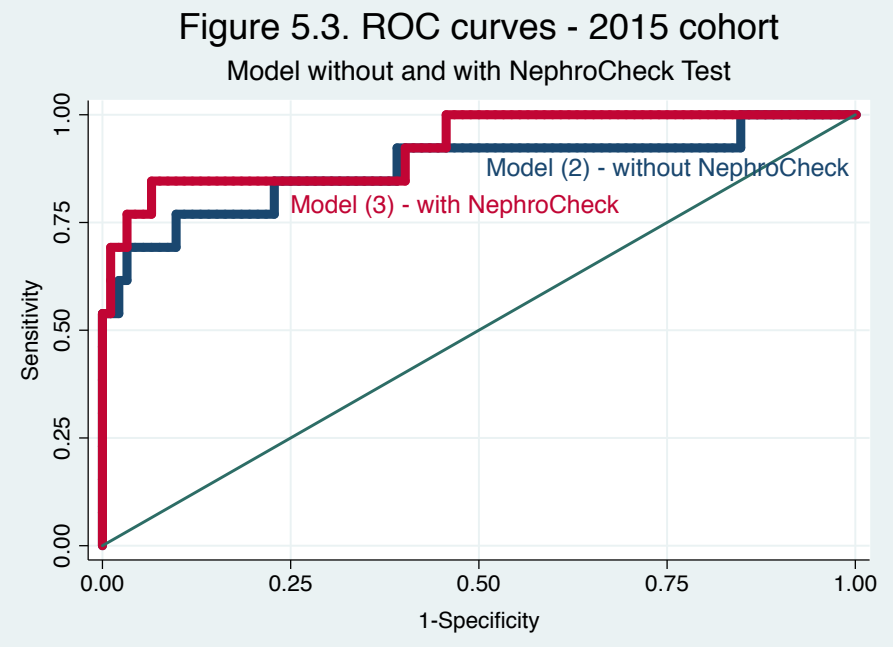
Classified	Actual		Total
	AKI	non-AKI	
AKI	9	6	15
non-AKI	4	86	90
Total	13	92	105
Sensitivity	69.23%		
Specificity	93.48%		
PPV	60.00%		
NPV	95.56%		
Correctly classified	**90.48%**	**AUC**	**87.54%**

Table 5.11. Classification Table at 25%. With NephroCheck® Test (*nc12hcut03)* – 2015 cohort

Classified	Actual		Total
	AKI	non-AKI	
AKI	11	6	17
non-AKI	2	86	88
Total	13	92	105
Sensitivity	84.62%		
Specificity	93.48%		
PPV	64.71%		
NPV	97.73%		
Correctly classified	**92.38%**	**AUC**	**92.38%**

The model without *nc12hcut03* already did a relatively good job at classifying the patients into AKI and non-AKI groups (90.48% of patients correctly classified); however, once *nc12hcut03* is included, the model improves substantially in terms of sensitivity (from 69.23% to 84.62%) without losing anything in terms of specificity. In other words, the true positives rate increases and, at the same time, the false negative rate decreases.

Figure 5.3 also displays the improvement in the area under the ROC curve, due to the inclusion of *nc12hcut03* in the model.



Afterwards, the continuous variable *nc12h* replaced *nc12hcut03* in the logit model and the two models were compared. Results from this regression are displayed in Table 5.12 below.

Table 5.12. Logit Regression (5)

Logit Regression				Number of obs = 105		
				LR chi2(12) = 36.95		
				Prob > chi2 = 0.0001		
Log likelihood = -20.840295				Pseudo R2 = 0.4699		
akipost	**Coeff.**	**Std. Err.**	**z**	**P>z**	**[95% Conf. Interval]**	
akipre	(omitted)					
female	2.201251	1.357064	1.62	0.105	-.4585449	4.861047
agecut						
65-	-.2109562	1.095458	-0.19	0.847	-2.358015	1.936103
75-	.8274323	1.351367	0.61	0.540	-1.821198	3.476062
ckdanamnesi	(omitted)					
obese	-.8781113	1.459336	-0.60	0.547	-3.738356	1.982134
eccdur	.024244	.0120807	2.01	0.045	.0005663	.0479217
surtype						
1 multvalve	-.6103087	1.544984	-0.40	0.693	-3.638423	2.417805
2 cabg	.15481	1.205525	0.13	0.898	-2.207976	2.517596
3 aorta	-1.321832	1.514433	-0.87	0.383	-4.290066	1.646402
postcrea	11.73584	4.51603	2.60	0.009	2.884579	20.58709
lasix20	.7005834	.4972069	1.41	0.159	-.2739241	1.675091
nc12h	4.161741	1.853187	2.25	0.025	.5295625	7.79392
_cons	-17.52182	5.320362	-3.29	0.001	-27.94954	-7.094101

Also the continuous variable *nc12h* is statistically significant and results in terms of coefficient value and significance of the whole model do not seem to differ a lot. In terms of classification performance at 25% probability cut-off, Table 5.13 shows that the model with the continuous variable *nc12h* is less sensitive (76.92%, as compared to 84.62% of the model with the binary variable *nc12hcut03*) and the false number of false negatives slightly increases, but the false positive rate decreases (specificity increases to 95.65%, as compared 93.48%). Overall, the proportion of patients correctly classified mildly increases to 93.33%.

Table 5.13. Classification Table at 25%.
With NephroCheck® Test (*nc12h*)
– 2015 cohort

Classified	Actual		Total
	AKI	non-AKI	
AKI	10	4	14
non-AKI	3	88	91
Total	13	92	105
Sensitivity	76.92%		
Specificity	95.65%		
PPV	71.43%		
NPV	96.70%		
Correctly classified	**93.33%**		**AUC 89.63%**

A criterion that can be used in order to decide between the two alternative model specifications is the Akaike Information Criterion (AIC) (Hosmer, Lemeshaw and Sturdivant 2013). This measure is defined in the following way:

$$\text{AIC} = -2\times L + 2\times(p+1),$$

where L is the log-likelihood of the fitted model and p is the number of regression coefficients excluding constant covariates. The rule is, in general, that models with a lower value of AIC should be preferred with respect to models with a higher value.
In this context, model (4) has an AIC=63.425 and model (5) reports AIC=65.681. Difference in AIC is not substantial, however, according to this criterion the model with the binary variable *nc12hcut03* may be preferred to the model with the continuous one, *nc12h*.

Even if these results mildly support the choice of the binary variable *nc12hcut03*, additional analysis could be done to check whether the variable *nc12h* is linear in the logit, by means of fitting a lowess smooth curve.

"Lowess" stands for locally weighted regression and lowess smoothing is a non-parametric procedure that provides a graphical representation of the relationship between the outcome variable and one (or more) independent variable (Jacoby 2000). This technique

consists of fitting a regression line to the data around an observation i (x_i, y_i) and, from this regression of y on x, predict the smoothed value $\widehat{y_i}$ for that particular point. This regression is "local" in the sense that only the subset of observations falling close to observation i (along the x axis) are used to predict the smoothed value. The weighting procedure is such that the central point (x_i, y_i) receives the highest weight, while points that are farther receive less weight. The procedure is repeated for each observation (a separate regression equation is estimated for each i) to draw the plot of the locally weighted regression, by means of connecting adjacent fitted points.

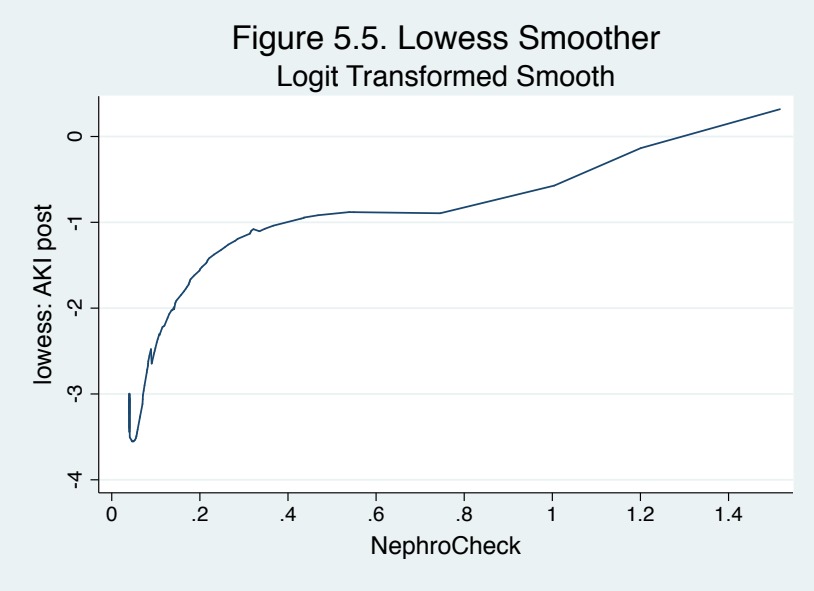


Figure 5.5 shows the lowess smoothed curve in terms of the log of the odds ratio and it actually shows some departure from a linear relationship.

However, this issue can be further examined through Fractional Polynomials modelling. This method can be used when there is suspect of non-linear relationship between the outcome and a continuous predictor, *nc12h* in this setting, in ordee to identify the optimum power transformation for the predictor. A fractional polynomial model is fit choosing powers from the subset $S = \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$ and allowing for logarithms, then substituting the continuous covariate with the best fitting fractional polynomial, defined as

$$\beta_0 + \sum_{j=1}^{m} \beta_j H_j(x)$$

where $H_1(x) = x^{(p_1)}$ and for $j = 2, \ldots, m$:

$$H_j(x) = \begin{cases} x^{(p_j)} & \text{if } p_j \neq p_{j-1} \\ H_{j-1}(x)\ln(x) & \text{if } p_j = p_{j-1} \end{cases}.$$

The "best" model is the one with the lowest deviance, m=2 in Table 5.14, which is then compared to lower degree models through partial *F* test. The column of *p*-values shows that the linear model cannot be rejected in favour of the fractional polynomial model with m=2 (*p*-value=0.172), meaning that there is no need for adding power transformations of the covariate *nc12h* to the model.

Table 5.14. Fractional Polynomial Analysis of *nc12h*

nc12h	df	Deviance	Dev. dif.	P	Powers
Not in model	0	46.406	9.718	0.045	
Linear	1	41.681	4.993	0.172	1
m = 1	2	37.421	0.733	0.693	-.5
m = 2	4	36.688	--	--	-2 -2

Overall, the comparison of models (4) and (5) can lead to the conclusion that the two models do not exhibit substantial differences and the choice between the two model could be based on the objective of classification (such minimization of the false positive rate or minimization of the false negative rate). Moreover, according to Fractional Polynomials analysis, it can be concluded that the relationship between the continuous covariate of interest (*nc12h*) and the outcome does not show substantial departures from linearity, thus suggesting that there is no need for adding other additional terms to the model.

CONCLUSIONS

This dissertation has investigated the use of the NephroCheck® Test as a diagnostic tool for the early identification of AKI (Acute Kidney Injury). AKI is a relevant complication affecting hospitalized patients and it is particularly frequent after cardiac surgery and in intensive care units. Any AKI case, from its milder forms to the most severe ones, is associated with adverse clinical outcome, such as increased morbidity and mortality, longer hospital stay, need for additional treatments including renal replacement therapy (RRT). These negative impacts are not limited to the hospitalization period but might extend, especially in more severe cases, over a long run. As a result, AKI is associated with a substantial increase in health care cost. Kerr et al. (2014) estimated that the annual AKI-associated costs, including inpatient care and post-discharge care, is around £1.2 billions in England and represents 1% of the National Health System budget. Besides that, AKI is likely to have adverse long-term clinical effects, even when patients recover renal function. Pannu et al. (2013) found that, among the patients who had developed AKI (KDIGO stage 2 or 3), 30.8% of them died and 2.1% progressed to kidney failure (CKD stage 5) requiring dialysis within a 34-month period.

Unfortunately, there is no specific therapy for AKI but several clinical measures can be adopted to avoid progression of the injury and patients with a high risk of developing AKI can be monitored and treated carefully to prevent AKI. This means that the primary goal for dealing effectively with AKI is to recognise the onset of the injury early, in order to allow timely identification of appropriate treatments and interventions. An entire branch of research, based on this principle, has been directed towards the discovery and validation of early *biomarkers* for the diagnosis of AKI. Many biomarkers have actually been investigated and a combination of two of them has provided the best performance, so far: urine insulin-like growth factor-binding protein 7 (IGFBP7) and tissue inhibitor of metalloproteinases-2 (TIMP-2). The NephroCheck® Test (Astute Medical, San Diego, CA, USA) is a diagnostic device that has been developed to allow measurement of the concentration of these two markers and, eventually, support AKI detection.

The NephroCheck® Test was employed on a cohort of patients undergoing cardiac surgery in the San Bortolo Hospital of Vicenza during 2015. The results of testing were available for an empirical evaluation of the performance of the NephroCheck® Test as a diagnostic method for detecting AKI early. The basic idea consisted of comparing a method

for AKI recognition, based on information available in the clinical routine (thus, without additional examinations and costs), to one that accounted for the results provided by the NephroCheck® Test.

To do so, a multiple logistic regression model for predicting the probability that a patient develops AKI after cardiac surgery was constructed on a sample of patients who underwent cardiac surgery in 2014. The individuals in the two samples (2014-no NephroCheck® Test and 2015-NephroCheck® Test) had comparable epidemiological and clinical characteristics. The multiple regression model included variables describing patients' clinical background, surgical procedures and early post-surgical features; the covariates of the model were selected based on clinical considerations as well as goodness of fit of the logistic model. The selected model was then applied to the 2015 cohort and, subsequently, the results of the NephroCheck® Test were included as a covariate, in order to assess improvement in AKI risk prediction through the model.

Findings can be summarized as follows. The NephroCheck® Test is statistically significant as a covariate in the multiple logistic regression model (p-value < 0.05) and an increase in test results signals a higher probability of developing AKI after cardiac surgery, *ceteris paribus*. Once the NephroCheck® Test is included, the performance of the model improves: the area under the curve (AUC) goes from 87.54% to 92.38%, which is a very positive result, and the main contribution is given in terms of sensitivity (increasing from 69.23% to 84.62%, at the selected probability cut-off of 25%).

The test is structured in a way that it provides a numerical result over a continuous scale, with higher numbers indicating an increasing risk of developing AKI. Previously published validation studies had identified a 0.3 cut-off value for the test result as a threshold (see Ronco 2014), such that patients with test values higher than 0.3 are identified as having an increased risk of developing moderate to severe AKI (KDIGO stages 2-3). In the present study, the NephroCheck® Test was evaluated for predicting the probability of developing AKI (any stage) according to KDIGO guidelines. A binary covariate assuming value 1 for NephroCheck® Test results higher than 0.3 as well as a continuous covariate were generated and considered in two separate models. Comparing the two models, no substantial difference emerged in terms of overall model performance for the classification of patients into AKI and non-AKI groups. Additionally, no significant discontinuities or need for adding power transformations of the continuous NephroCheck® Test covariate were found.

The results presented in this dissertation must be evaluated in light of some considerations. First of all, the NephroCheck® Test was employed on a specific and rather homogeneous population, namely that of patients undergoing cardiac surgery. Therefore, further research would be necessary to validate the use of the test on a more heterogeneous population, such as including critical care patients.

Analysis and comparison of the classification tables of the models described here show that the NephroCheck® Test improves early AKI detection with respect to a model that accounts only for other clinical parameters. Nevertheless, an integrated approach could probably represent the most successful diagnostic strategy, combining existing diagnostics (SCr and urine output) and clinical risk scores with biomarkers-based diagnostics (the NephroCheck® Test) in relation to patient conditions. In particular, in a more heterogeneous patient context it might not be optimal to use biomarker testing routinely, but it could rather be more appropriate to identify a subset of patients in whom testing would be utilized more efficiently (Basu, Gist and Wheeler 2015).

A crucial consideration concerns also the timing of the test. On the one hand, the primary purpose is to detect AKI as early as possible. On the other hand, if the test is performed too early, the biomarkers might have not increased enough to signal the risk of kidney injury. This trade-off must be solved analysing the responsiveness of biomarkers' concentration to kidney damage, without forgetting that correct timing might be dependent on the population studied, besides the characteristics of the biomarker being analysed.

The model constructed in this dissertation exploiting the results of the NephroCheck® Test actually allows early detection of AKI after cardiac surgery. Test assessment took place 12h after surgery, which is much earlier than what current diagnostic methods allow (in the samples considered here, SCr usually does not increase until the third post-operative day). The scope for AKI prevention is, therefore, substantial even if it does not depend only on recognizing that a high risk of developing the injury exists, but also on the clinical measures that need to be started in order to prevent and treat the injury.

REFERENCES

Acute Kidney Injury Work Group, (2012). Kidney Disease: Improving Global Outcomes (KDIGO). KDIGO Clinical Practice Guideline for Acute Kidney Injury. *Kidney International,* Suppl. 2012, 2, 1-138.

Banta, D. (2009). What is technology assessment? *International Journal of Technology Assessment in Health Care*, 25(S1), 7–9.

Basu, R. K., Gist, K., and Wheeler, D. S. (2015). Improving acute kidney injury diagnostics using predictive analytics. *Current Opinion in Critical Care*, 21(6), 473–478

Bihorac, A., Chawla, L. S., Shaw, A. D., Al-Khafaji, A., Davison, D. L. et al. (2014). Validation of cell-cycle arrest biomarkers for acute kidney injury using clinical adjudication. *American Journal of Respiratory and Critical Care Medicine*, 189(8), 932-939.

Chertow, G. M., Burdick, E., Honour, M., Bonventre, J. V., and Bates D. W. (2005). Acute Kidney Injury, mortality, Length of Stay, and Costs in Hospitalized Patients. *Journal of American Society Nephrology,* 16, 3365–3370.

Cruz, D. N., Ricci, Z., and Ronco, C. (2009). Clinical review: RIFLE and AKIN – time for reappraisal. *Critical Care*, 13(3), 211.

D'Arrigo, G., Provenzano, F., Torino, C., Zoccali, C., and Tripepi, G. (2011). I test diagnostici e l'analisi della curva roc. *Giornale Italiano di Nefrologia,* 28(6), 642-647.

Dasta, J. F., Kane-Gill, S. L., Durtschi, A. J., Pathak, D. S., and Kellum, J. A. (2008). Costs and outcomes of acute kidney injury (AKI) following cardiac surgery. *Nephrology, Dialysis, Transplantation*, 23(6), 1970–1974.

De Smedt, D. M., Elseviers, M. M., Lins, R. L., and Annemans, L. (2012). Economic evaluation of different treatment modalities in acute kidney injury. *Nephrology, Dialysis, Transplantation*, 27(11), 4095-101.

Drummond, M. F., Sculpher, M. J., Torrance, G. W., O'Brien, B. J., and Stoddart, G. L. (2005). Methods for the Economic Evaluation of Health Care Programme. 3rd ed., Oxford: Oxford University Press (trad. it. Mennini, F. S. (2010). Metodi per la valutazione economica dei programmi sanitari. Roma: Il pensiero scientifico)

Dybczak, K., and Przywara, B. (2010). The role of technology in health care expenditure in

the EU. *Economic Papers 400*, European Economy.

Ethgen, O., Schneider, A. G., Bagshaw, S. M., Bellomo, R., and Kellum, J. A. (2015). Economics of dialysis dependence following renal replacement therapy for critically ill acute kidney injury patients. *Nephrology, Dialysis, Transplantation,* 30(1), 54-61.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874.

Ferenbach, D. A., and Bonventre, J. V. (2015). Mechanisms of maladaptive repair after AKI leading to accelerated kidney ageing and CKD. *Nature Reviews Nephrology*, 11, 264–276.

Haase, M., Kellum, J. A., and Ronco, C. (2012). Subclinical AKI—an emerging syndrome with important consequences. *Nature Reviews Nephrology*, 8, 735–739.

Heyland, D. K., Gafni, A., Kernerman, P., Keenan, S., and Chalfin, D. (1999). How to use the results of an economic evaluation. *Critical Care Medicine,* 27(6), 1195-1202.

Hosmer, D. W., Jr., Lemeshaw, S., and Sturdivant, R. X. (2013). *Applied Logistic Regression*. 3rd ed., Hoboken, NJ: John Wiley and Sons.

Hoste, E. A. J., McCullough, P. A., Kashani, K., Chawla, L. S., Joannidis, M., et al.. on behalf of the Sapphire Investigators (2014). Derivation and validation of cutoffs for clinical use of cell cycle arrest biomarkers. *Nephrology, Dialysis, Transplantation,* 29, 2054–2061.

Hounton, S., and Newlands, D. (2012). Applying the net-benefit framework for assessing cost-effectiveness of interventions towards universal health coverage. *Cost Effectiveness and Resource Allocation*, 10(1), 8.

Jacoby, W. J. (2000). Loess: a non-parametric, graphical tool for depicting relationships between variables. *Electoral Studies* 19, 577-613.

Jennet, B. (1992). Health technology assessment. The rule should be "no evaluation – no technology". *British Medical Journal*, 305, 67-68.

Jegers, M., Edbrooke, D. L., Hibbert, C. L., Chalfin, D. B., and Burchardi, H. (2002). Definitions and methods of cost assessment: an intensivist's guide. *Intensive Care Medicine*, 28(6), 680–685.

Jorge-Monjas, P., Bustamante-Munguira, J., Mario Lorenzo, M., Heredia-Rodríguez, M., Fierro, I., et al. (2016). Predicting cardiac surgery–associated acute kidney injury: The CRATE score. *Journal of Critical Care* 31(1), 130–138.

Kashani, K., Al-Khafaji, A., Ardiles, T., Artigas, A., Bagshaw, S. M., et al. (2013). Discovery and validation of cell cycle arrest biomarkers in human acute kidney injury. *Critical Care*, 17(1), R25

Kerr, M., Bedford, M., Matthews, B., and O'Donoghue, D. (2014). The economic impact of acute kidney injury in England. *Nephrology, Dialysis, Transplantation*, 29(79), 1362–1368.

Kobelt, G. (2013). Health Economics: An Introduction to Economic Evaluation. 3rd ed., London: Office of Health Economics.

Kosiborod, M. N. (2012). Trends in acute kidney injury in patients with acute myocardial infarction. *Archives of Internal Medicine*, 172(3), 246-253.

Levin, A., and Stevens, P.E. (2014). Summary of KDIGO 2012 CKD Guideline. Behind the scenes, need for guidance, and a framework for moving forward. *Kidney International,* 85(1), 49-61.

Lewington, A., and Hall, P. (2014). The cost of ignoring acute kidney injury. *Nephrology, Dialysis, Transplantation*, 0, 1-2

Mandelblatt, J. S., Fryback, D. G., Weinstein, M. C., Russell, L. B., Gold, M. R., and the Members of the Panel on Cost-Effectiveness in Health and Medicine (1997). Assessing the effectiveness of health interventions for cost-effectiveness analysis. *Journal of General Internal Medicin*e, 12(9), 551-558.

Manns, B., Doig, C. J., Lee, H., Dean, S., Tonelli, M., et al. (2003). Cost of acute renal failure requiring dialysis in the intensive care unit: Clinical and resource implications of renal recovery. *Critical Care Medicine*, 31(2), 449-55.

Mao, H., Katz, N., Ariyanon, W., Blanca-Martos, L., Adybell, Z., et al. (2014). Cardiac Surgery-Associated Acute Kidney Injury. *Blood Purification*; 37(S2), 34-50.

Mariscalco, G., Lorusso, R., Dominici, C., Renzulli, A., and Sala, A. (2011). Acute kidney injury: A relevant complication after cardiac surgery. *The Annals of Thoracic Surgery*, 92(4), 1539-1547.

Mazairac, A. H. A., Blankestijn, P. J., Penne, E. L., van der Weerd, N. C., den Hoedt, C. H., et al. on behalf of the CONTRAST investigators (2013). The cost-utility of haemodiafiltration versus haemodialysis in the Convective Transport Study. *Nephrology, Dialysis, Transplantation*, 28(7), 1865-1873.

Mishra, J., Dent, C., Tarabishiv, R., Mitsnefes, M. M., Ma, Q., et al. (2005). Neutrophil gelatinase-associated lipocalin (NGAL) as a biomarker for acute renal injury after cardiac surgery. *Lancet*, 365(9466), 1231–1238.

OECD (2015). Health at a Glance 2015: OECD Indicators. Paris: OECD Publishing.

Oliveira Martins J., and de la Maisonneuve, C. (2006). The Drivers of Public Expenditure on Health and Long-Term Care: an Integrated Approach. *OECD Economic Studies*, 2006(2), 11-11

Pannu, N., James, M., Hemmelgarn, B., Klarenbach, S. for the Alberta Kidney Disease Network (2013). Association between AKI, Recovery of Renal Function, and Long-Term Outcomes after Hospital Discharge. *Clinical Journal of American Society Nephrology*, 8(2), 194–202.

Rodríguez, G. (2007). *Lecture Notes on Generalized Linear Models*. [online] Available at: http://data.princeton.edu/wws509/notes [Accessed in July 2016].

Ronco, C. (2014). Cell cycle arrest biomarkers: New weapons for a new battle. *Blood purification*, 38, I-III.

Schneider, A. G., Bellomo, R., Bagshaw, S. M., Glassford, N. J., Lo, S., et al. (2013). Choice of renal replacement therapy modality and dialysis dependence after acute kidney injury : A systematic review and meta-analysis. *Intensive Care Medicine*, 39(6), 987-997.

Shaw, A. D., Chalfin, D. B., and Kleintjens, J. (2011). The economic impact and cost-effectiveness of urinary Neutrophil Gelatinase-Associated Lipocalin after cardiac surgery. *Clinical Therapeutics*, 33(11), 1713-1725.

Srisawat, N., Lawsin, L., Uchino, S., Bellomo, R., and Kellum, J. A. (2010). Cost of acute renal replacement therapy in the intensive care unit: results from The Beginning and Ending Supportive Therapy for the Kidney (BEST Kidney) Study. *Critical Care*, 14(2), R46.

Tan-Torres Edejer, T., Baltussen, R., Adam, T., Hutubessy, R., Achary,a A., et al. (edited by) (2003). Making choices in health: WHO guide to cost-effectiveness analysis. Geneva: World Health Organization.

Taylor, J. (2011). Calculating the Cost, Acute Kidney Injury. A Health Services Journal Supplement (Internet). London: Health Service Journal. Available from http://www.hsj.co.uk/Journals/2/Files/2011/6/23/HSJ_Supplement_acutekidneyinjur_23june.pdf

Tsigou, E., Psallida, V., Demponeras, C., Boutzouka, E., and Baltopoulos, G. (2013). Role of new biomarkers: Functional and structural damage. *Critical Care Research and Practice,* 2013, 361078

Wetz, A. J., Richardt, E. M., Wand, S., Kunze, N., Schotola, H., et al. (2015). Quantification of urinary TIMP-2 and IGFBP-7: an adequate diagnostic test to predict acute kidney injury after cardiac surgery? *Critical Care*, 19, 3.

Zhou, H., Hewitt, S. M., Yuen, P. S. T., and Star, R. A. (2006). Acute Kidney Injury Biomarkers - Needs, Present Status, and Future Promise. *Nephrology Self Assessment Program*, 5(2), 63–71.

WEB SITES CONSULTED:

http://www.astutemedical.com
http://www.ats.ucla.edu/stat/
http://www.htai.org/htai/what-is-hta.html
http://www.niddk.nih.gov/health-information/health-topics/Anatomy/kidneys-how-they-work/Pages/anatomy.aspx
https://www.medcalc.org/manual/introduction_statistics.php
www.stata.com