

Indice

Introduzione	1
Capitolo 1 Pianificazione, gestione e controllo delle sale operatorie	4
1.1 Analisi generale del contesto.....	4
1.2 Architettura organizzativa e gestionale	6
1.3 Problemi di Pianificazione, Gestione e Controllo	12
1.3.1 Esigenze di efficienza ed efficacia.....	15
1.3.2 Esigenze di sicurezza: il concetto di rischio clinico	17
1.3.3 Il ruolo dell'innovazione ed integrazione	22
1.3.4 Qualità nei processi clinici.....	26
1.3.5 Fattori di incertezza	28
Capitolo 2 Modelli e metodi quantitativi per la pianificazione e gestione delle sale operatorie	35
2.1 Generalità, livelli e processi decisionali, requisiti e specifiche funzionali.....	35
2.2 Approcci deterministici, generalità.....	39
2.3 Approcci stocastici	47
Capitolo 3 Modelli di Programmazione Stocastica a due stadi con ricorso per la gestione operativa delle sale operatorie	58
3.1 Architettura dei modelli.....	58
3.2 Sviluppo dei modelli.....	66
3.3 Analisi della struttura e delle proprietà dei modelli.....	71
Capitolo 4 Metodi risolutivi	76
4.1 Generalità su approcci esatti.....	76
4.2 Generalità su approcci euristici	80
4.3 Disegno e sviluppo dell'euristica proposta.....	82
Capitolo 5 Esperimenti computazionali	89
5.1 Strategia sperimentale.....	89
5.2 Costruzione delle istanze per la fase di testing.....	90
5.3 Risultati ottenuti nella fase di testing	94
5.4 Analisi dei risultati ottenuti	98
5.5 Approccio stocastico vs approccio deterministico: un “toy example”	101
Conclusioni e sviluppi futuri	110
Bibliografia	113
Ringraziamenti	118

Introduzione

Il presente lavoro di ricerca è il risultato del complessivo percorso di formazione triennale del Dottorato di Ricerca in “Ricerca Operativa” presso l’Università della Calabria.

Le motivazioni che hanno contribuito a sviluppare e presentare in questo elaborato strumenti quantitativi per la gestione operativa dei blocchi operatori, si inquadrano negli attuali scenari che caratterizzano i sistemi sanitari e vanno riferite alle loro specifiche peculiarità.

Il panorama attuale è infatti caratterizzato da un crescente livello di domanda di tutela della salute, a cui sono annesse richieste di prestazioni sanitarie sempre maggiori e di migliore qualità. Evidentemente, questo impatta fortemente sul lato dell’offerta dei servizi sanitari, considerata soprattutto la problematica dell’allocazione ottimale delle risorse necessarie, sempre più scarse in termini di disponibilità.

Per le suddette ragioni chi, in particolare, si occupa della pianificazione, della gestione e del controllo delle sale operatorie deve costantemente affrontare il problema del controllo e di una riduzione dei costi, tenendo conto, ovviamente, del fatto che ciò non deve influire negativamente sul livello di efficacia e qualità del sistema stesso.

Nella situazione attuale sarebbe, pertanto, fortemente auspicabile un rilancio improntato su nuovi approcci alla pianificazione e gestione sistemica della sanità che implica, al fine di un adeguato controllo dei costi, lo sviluppo di efficaci ed efficienti strumenti quantitativi.

Nello specifico, tali strumenti possono essere implementati come Sistemi di Supporto alle Decisioni, con l’obiettivo di coadiuvare il decisore nelle varie fasi di pianificazione e gestione, al fine di aumentare l’efficacia e l’efficienza del sistema e di migliorare contemporaneamente la sua risposta in relazione ad eventuali “parametri critici”.

Per tutti i suddetti motivi, l’obiettivo di questo lavoro è lo sviluppo di approcci modellistici e metodologici propri della “Programmazione Stocastica” per migliorare l’efficacia e l’efficienza nei processi decisionali di gestione operativa delle sale operatorie.

Tali approcci sono stati progettati e sviluppati dopo un’accurata analisi della letteratura scientifica di riferimento, al fine di portare un contributo innovativo nel settore.

Occorre osservare che la “Programmazione Stocastica” per la gestione delle sale operatorie è comunque un campo d’applicazione piuttosto inesplorato, dove solo in anni molto recenti la letteratura comincia ad offrire contributi di una certa rilevanza.

Vale la pena osservare che la “Programmazione Stocastica” può risultare uno strumento particolarmente efficace in tale contesto, in grado di affrontare una delle caratteristiche principali connaturata alla gestione dei blocchi operatori, ovvero l’incertezza.

L’elaborato si articola in cinque capitoli, a cui è stata aggiunta anche una sezione dedicata ad alcune considerazioni conclusive.

Il **Capitolo 1** contiene un’introduzione generale alla realtà di un blocco operatorio, andando nella fattispecie ad effettuare una disamina dei flussi di attività che avvengono in un tale sistema e ad illustrare la svariata natura delle problematiche che insorgono in questo contesto di riferimento (legate ad esigenze di efficacia e di efficienza, di sicurezza, di qualità nonché alla presenza del fattore di incertezza).

Nel paragrafo finale di questo primo capitolo viene dunque motivata, alla luce delle problematiche illustrate, la necessità dell’introduzione di un adeguato strumento quantitativo di supporto alle decisioni.

Nel **Capitolo 2** è motivato il delicato ruolo che l’Ottimizzazione può assumere in questo contesto e, dopo ciò, si prosegue presentando per sommi capi come essa può intervenire in questo settore.

A tal uopo si è passati ad esaminare quanto presente in letteratura circa l’ottimizzazione dei flussi di attività interni ad una sala operatoria, discernendo i modelli stessi in due grandi categorie.

La prima categoria racchiude tutti i modelli *deterministici*, ovvero quelli in cui il fattore di incertezza è trascurato o comunque trattato mediante approssimazioni ed assunzioni non troppo plausibili nella situazione reale.

Un’analisi di tali modelli inerenti a tale categoria illustra infatti che per un sistema complesso quale un blocco operatorio è davvero poco adatta una tipologia d’approccio siffatta, manifestando come il fattore d’incertezza deve essere trattato in maniera differente.

Proprio per questo motivo la seconda categoria racchiude invece tutti i modelli *stocastici*, in cui invece tale fattore di incertezza viene trattato in maniera specifica e ben più articolata.

A valle della presentazione dei concetti fondamentali legati alla *Programmazione Stocastica*, sono state pertanto illustrate le caratteristiche dei modelli più salienti inerenti a questa seconda categoria.

Un’analisi critica conclusiva ha illustrato le caratteristiche principali di tali modelli, nonché i limiti generali che ci si propone di andare a superare.

Il **Capitolo 3** è finalizzato alla presentazione dettagliata del modello di Programmazione Stocastica oggetto di questo lavoro.

Tale modello, la cui struttura di base rientra nell'ampia categoria dei problemi di Programmazione Intera Mista, è del tipo "a due stadi con ricorso", con l'ulteriore possibilità di poter essere configurato in tre specifiche varianti, a seconda di come viene definita la specifica funzione di ricorso.

Tale caratteristica, come sarà possibile notare nel corso del capitolo, offre al decisore la possibilità di avere a disposizione una vasta gamma di opzioni con le quali decidere di progettare la pianificazione e la gestione più efficiente per il blocco operatorio.

All'interno di questo capitolo viene comunque enfatizzata la complessità computazionale di una generica istanza di ciascuna variante del modello stesso, la quale potrebbe presentare in diversi casi un limite insormontabile per il decisore allorché i tempi computazionali potrebbero essere inaccettabili con i tradizionali metodi di risoluzione.

A tal uopo nel **Capitolo 4** è stata così presentata una metodologia euristica, che si propone di essere un mezzo alternativo per risolvere il problema diversamente dai tradizionali metodi esatti.

Nel **Capitolo 5** sono stati pertanto riportati i risultati di un'opportuna fase di testing mediante il sistema software *AIMMS 3.9*.

In tale fase vengono messi a confronto i risultati ottenuti risolvendo generiche istanze di ogni variante del modello sia con il risolutore **CPLEX 12.1** che con il metodo euristico presentato nel capitolo precedente.

I risultati dimostreranno che tale euristica si conferma essere allo stesso tempo sia efficace che efficiente, centrando dunque l'obiettivo di essere una valida alternativa dei tradizionali metodi di risoluzione allorché questi richiederebbero tempi computazionali inaccettabili per il decisore.

Infine nel capitolo vengono messi a confronto l'approccio stocastico e l'approccio deterministico equivalente sulla base dei risultati ottenuti con CPLEX per una sperimentazione di dimensioni estremamente ridotte.

Nella sezione **Conclusioni e sviluppi futuri** sono riportate le considerazioni finali circa il lavoro presentato, enfatizzando sull'originalità dei contributi apportati ed indicando linee di ricerca e sviluppi ancora aperti a possibili studi ed estensioni dello stesso modello oggetto dell'elaborato.

Il presente tomo è corredato da un CD in cui sono contenuti i codici con i quali ogni variante del modello è stata dichiarata nell'ambito del software *AIMMS 3.9* al fine di permetterne le adeguate sperimentazioni con l'opportuno linguaggio.

Capitolo 1

Pianificazione, gestione e controllo delle sale operatorie

1.1 Analisi generale del contesto

Nel corso degli ultimi anni è facile notare come il settore sanitario italiano stia soffrendo sempre più la carenza di politiche adeguate in grado di ottimizzarne l'utilizzo delle risorse e l'architettura dei cosiddetti "processi produttivi" che avvengono all'interno.

Tale carenza di politiche è solo in parte imputabile all'arricchimento dovuto all'introduzione di nuove componenti tecnologiche di alto profilo, dato che a tal uopo sono richieste figure tecniche specifiche e con determinate competenze: un esempio di tali componenti può essere l'introduzione di nuovi sistemi informatici con telecontrollo per la visione remota di sale operatorie, tanto per restare nell'ambito in cui questo lavoro di tesi si va ad incentrare.

Per svariate ragioni, comunque, un sistema sanitario necessita di figure professionali dalle competenze variegata e che siano capaci di progettare in maniera ottimale i processi di pianificazione, di gestione e di controllo ricorrendo alla loro esperienza inerente ad aspetti quali la progettualità, l'integrazione dei processi, la logistica...

Intercalandosi maggiormente in tale contesto di riferimento c'è innanzitutto da dire che nell'ambito di un sistema sanitario di qualsiasi dimensione, sono proprio le sale operatorie a rivestire un ruolo di cruciale importanza dato che rappresentano uno dei punti-cardine di tutto il sistema stesso, a causa dell'elevata complessità che ne caratterizza i vari processi interni e di tutte le criticità annesse.

Per rendersene ulteriormente conto, basta infatti pensare solamente che proprio le sale operatorie rappresentano i centri di costo ospedalieri di maggiore incidenza (le statistiche informano infatti che ad esse è imputabile circa il 40% dei costi totali ospedalieri [1] e che una percentuale oscillante tra il 60%-70% dei ricoveri ospedalieri è dovuta proprio alla necessità da parte del paziente di subire un intervento chirurgico).

Ci sono inoltre da considerare aspetti quali l'ampio flusso di attività che avviene all'interno di ciascuna sala operatoria ed il massiccio utilizzo delle risorse che vi entrano in gioco nell'ambito del "processo produttivo", che è altresì piuttosto complesso di suo e, come detto, richiede la presenza di diversi "attori" e figure professionali.

E questo perché, come suddetto e come sarà maggiormente dettagliato nel prosieguo di questo capitolo, negli ultimi anni il blocco operatorio è proprio il reparto ospedaliero che ha subito maggiori trasformazioni a causa dei miglioramenti delle tecniche chirurgiche.

Sempre in merito all'utilizzo delle risorse c'è poi da sottolineare che esso nella maggior parte dei casi avviene in maniera non efficiente, incidendo così in maniera significativa sui costi di natura gestionale-operativa del sistema stesso, e talvolta minando purtroppo anche con conseguenze letali il perseguimento degli standard previsti di prestazione sanitaria (in termini non solo di contenimento dei costi, ma anche di efficienza del servizio e di soddisfazione del paziente che deve usufruirne).

Va comunque sottolineato che quando si parla del concetto di "risorse" non vanno intesi solo gli strumenti, le modalità e le metodiche utilizzate nell'ambito di una sala operatoria per eseguire un determinato intervento chirurgico, ma anche le conoscenze, le possibilità e le opportunità di sviluppo futuro (tutte componenti che dovrebbero rivestire un ruolo di una certa importanza nell'ambito di un sistema sanitario).

A causa delle suddette motivazioni è chiaramente comprensibile che le fasi di pianificazione, di gestione e di controllo della sala operatoria prevedono quindi diversi gradi di complessità e coinvolgono competenze diverse e variegate.

A rendere poi queste fasi maggiormente delicate e di fondamentale importanza si aggiunge il fatto che sono proprio loro il punto di partenza dell'impostazione di tutte le altre strutture e parti che vanno a comporre il sistema sanitario nel suo complesso.

Ciò traduce quindi tali fasi in progetti di una certa complessità e che hanno l'obiettivo di creare un ambiente operativo capace di agevolare e di controllare il processo operatorio, nonché quello di favorire il miglioramento delle condizioni operative degli attori coinvolti.

L'analisi di un contesto caratterizzato da tale complessità non può pertanto limitarsi soltanto a considerarne gli aspetti generali, ma deve anche considerare gli aspetti organizzativi del "ciclo di lavoro" che si svolge al suo interno.

Sintetizzando si può dunque dire che soltanto con un adeguato piano strategico ed un'efficiente progettazione dei suddetti processi si possono risolvere i delicati e pressanti problemi legati a tali aspetti: fallire nell'obiettivo di pianificare, di gestire e di controllare adeguatamente quanto avviene in sala operatoria va a comportare conseguenze di non

indifferente entità sia dal punto di vista dei costi che chiaramente dal punto di vista della soddisfazione del paziente.

Tutto questo perché nell'ambito di queste fasi vi sono altresì importanti aspetti come le esigenze di sicurezza e di qualità nello svolgimento dei processi, aspetti che verranno approfonditi maggiormente nei successivi paragrafi di questo capitolo.

C'è comunque da evidenziare quanto sia difficile anche mantenere un adeguato controllo della qualità negli interventi in un ambiente così complesso e denso di criticità, ovvero far sì che gli "elementi variabili" del processo (le modalità operative, insomma) ed il grado di addestramento/competenza del personale che entra in gioco in questo processo rimangano pressoché costanti e che nel tempo garantiscano che il servizio erogato si mantenga conforme agli standard prefissati.

1.2 Architettura organizzativa e gestionale

In generale, una sala operatoria la si può definire come una parte del cosiddetto "blocco operatorio", ovvero l'insieme dei reparti operatori delle diverse unità operative.

Proprio al suo interno il paziente è sottoposto all'intervento chirurgico con metodiche e con modalità routinarie programmate e/o d'urgenza, a seconda dei casi.

Una sala operatoria si differenzia dalle altre unità operative per la presenza temporanea del paziente, strettamente correlata alla prestazione chirurgica da erogare e da intendere come fase del processo integrata con la degenza, nell'ambito del processo di guarigione del paziente medesimo.

Il tutto va così a caratterizzare un'elevata complessità assistenziale dovuta alla più o meno grave instabilità delle funzioni vitali del paziente, nonché all'intensità, alla complessità ed alla continuità delle prestazioni erogate e delle interazioni che si svolgono al suo interno.

In generale, comunque, per ogni struttura sanitaria bisogna pertanto definire in maniera sensata a priori il numero complessivo di sale operatorie in funzione della tipologia e della complessità delle prestazioni che la specialità chirurgica in questione si presuppone di erogare, tenendo altresì conto anche della complessità legata agli aspetti di natura squisitamente tecnologica-strumentale.

Il blocco operatorio, inoltre, deve rispondere in modo evidente ad un'ottimizzazione dell'impiego del personale con la graduale integrazione delle competenze fra operatori di

specialità differenti e diverse mansioni nell'ambito della stessa specialità, sempre seguendo alcuni dettami al fine di garantire una certa qualità nella prestazione.

Per quanto riguarda la dotazione organica del personale medico ed infermieristico, si può dire che essa va rapportata alla tipologia ed al livello di complessità degli interventi chirurgici.

L'attivazione di una sala operatoria deve difatti prevedere mediamente la presenza di almeno un medico anestesista, due chirurghi (di cui in genere uno maggiormente esperto detto "chirurgo senior" e l'altro meno esperto e detto per questo motivo "chirurgo junior") e due infermieri professionali, andando quindi a definire l'organigramma delle responsabilità e delle conseguenti funzioni da espletare.

Il tutto va naturalmente svolto senza trascurare aspetti fondamentali quali i collegamenti funzionali con le altre unità operative, i regolamenti interni e/o le procedure per lo svolgimento delle attività di gestione, nonché la presenza della documentazione inerente sia alla valutazione anestesiologicala che alla procedura interventistica (ovvero il cosiddetto "registro operatorio").

L'architettura organizzativa deve quindi considerare e prevedere soprattutto le procedure per fornire risposte adeguate sia alle richieste routinarie, sia alle richieste in emergenza/urgenza.

Proprio per tale motivo deve esistere quindi una lista operatoria, basata su determinati criteri che definiscano la tipologia d'intervento e dell'anestesia richiesta, nonché la gestione della lista stessa in caso d'urgenza.

Tale lista deve inoltre prevedere un documento informativo per i pazienti che descrive le modalità d'assistenza ed i diritti degli stessi, tra cui un programma di verifica dei risultati e di miglioramento della qualità.

Dal punto di vista invece del vero e proprio "processo" che avviene all'interno di una sala operatoria (che si può comunemente definire con semplicità "*processo clinico*" o "*workflow clinico*"), essa la si può considerare come un sistema input/output che va ad adoperare una trasformazione.

Tale trasformazione avviene in quanto il sistema riceve in entrata un determinato input (i pazienti che necessitano dell'intervento chirurgico) ed al suo interno il medesimo input viene trasformato subendo il suddetto "processo clinico" per mezzo delle risorse che vi entrano in gioco: quando tale processo sarà terminato il sistema avrà in uscita il suo output (ovvero il paziente dopo l'operazione).

Per riassumere quanto appena detto, si può vedere uno schema che sintetizzi l'idea di considerare una sala operatoria come un sistema input/output:

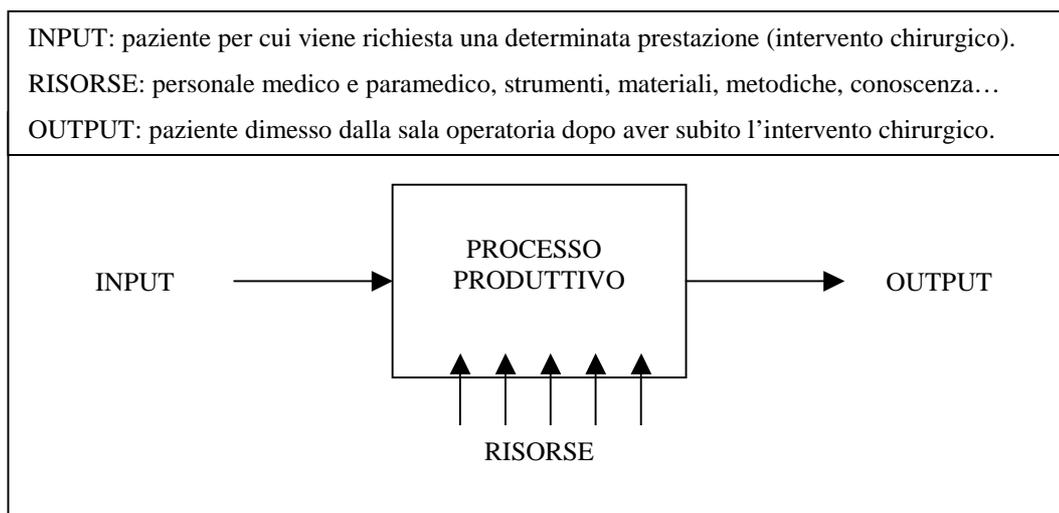


Figura 1.1: schematizzazione del “processo clinico” di una sala operatoria.

Esaminando maggiormente nel dettaglio il suddetto “*processo clinico*”, c’è da dire che il paziente non giunge direttamente dal reparto ospedaliero in cui era ricoverato alla sala operatoria per subire l’intervento chirurgico che gli è necessario, ma che vi sono anche delle operazioni a monte così come chiaramente lo stesso paziente dopo l’operazione può non essere direttamente rimandato nel medesimo reparto ospedaliero, ma che vi possono essere contestualmente anche delle operazioni a valle dell’intervento chirurgico stesso.

A tal uopo è necessario quindi segmentare il cosiddetto “processo produttivo” in tre fasi distinte e sequenziali:

1. fase pre-operatoria;
2. fase operatoria;
3. fase post-operatoria.

Nella fase “**pre-operatoria**”, che è per l’appunto precedente alla decisione di intervenire chirurgicamente sul paziente ed al trasferimento dal reparto di degenza alla sala operatoria, il paziente viene preparato a livello fisico e psichico (in base ovviamente al tipo di intervento chirurgico in questione), tenendo conto della valutazione delle sue caratteristiche individuali e considerando inoltre che una corretta valutazione può permettere di individuare interventi specifici e di definire il grado di rischio che il paziente può affrontare nei postumi dell’intervento stesso.

In quest’ambito dunque assume importanza l’educazione sanitaria, poiché al paziente vengono fornite informazioni inerenti alla descrizione della procedura a cui verrà

sottoposto, al ruolo che dovrà rivestire in questo processo ed agli effetti di eventuali farmaci che potrebbero essergli somministrati (in particolar modo gli analgesici).

Per casi particolari, inoltre, in questa fase il paziente può anche ricevere ragguagli sull'eventuale fase post-operatoria o dimostrazioni sugli eventuali esercizi riabilitativi che deve imparare ad effettuare, essendone chiaramente consapevole della loro utilità. Per espletare tutte le suddette funzioni vi possono essere all'interno della sistema sanitario delle strutture apposite che vengono comunemente chiamate "PHU" (acronimo di "Preoperative Holding Unit", ovvero, per l'appunto, "unità di possesso pre-operatorio"). Tuttavia vi sono in questa fase anche aspetti che possono essere definiti di natura burocratica, ovvero quelli inerenti all'accettazione del paziente nella sala operatoria e che comprendono le fasi di rilevazione e di registrazione dei dati relativi alle condizioni generali del paziente, al fine di individuarne come suddetto i bisogni e la definizione del rischio operatorio per lo stesso.

Alla suddetta fase a monte segue quindi la vera e propria fase "**operatoria**", che ha inizio con il trasferimento del paziente nella sala operatoria (in inglese "*operating room*" o più semplicemente nota con l'acronimo "OR") ed ha termine quando il paziente viene dimesso da essa. Chiaramente è comprensibile che si tratta della fase più delicata in cui sono considerati in primo piano i bisogni del paziente, la sua sicurezza ed il suo benessere. Gli attori che ne entrano in gioco rivestono pertanto un ruolo di cruciale importanza al fine di garantire un esito positivo dell'intervento e di evitare la possibilità di complicanze che possano eventualmente incidere in maniera negativa sul decorso post-operatorio. L'equipaggiamento di tali sale operatorie prevede invece la presenza di tutti gli strumenti ed i macchinari volti a consentire di gestire l'attività con il massimo grado di sicurezza per il paziente e di flessibilità operativa.

A valle della fase operatoria vi è quindi la fase "**post-operatoria**", la quale va ad includere il trasferimento del paziente dalla sala operatoria, appena terminato l'intervento chirurgico, alla cosiddetta "sala di recupero" o "sala di risveglio". Ovviamente non in tutti i casi il paziente dopo aver subito l'operazione deve essere obbligatoriamente trasferito qui: dipende dal tipo di intervento chirurgico in questione, poiché vi sono anche casi in cui il paziente dopo essere stato operato può essere senza alcun problema direttamente rimandato al reparto di degenza da cui era provenuto.

In questa fase gli obiettivi principali sono comunque quelli di garantire al paziente il ripristino dell'equilibrio fisiologico e l'adeguato sostegno a livello psichico, nonché eventualmente il sollievo dal dolore e la cura di ferite derivanti dall'intervento e/o la prevenzione o la cura di possibili complicazioni successive all'intervento stesso.

Anche in questo caso vi possono essere delle strutture dedite esclusivamente a questa fase e che sono denominate “PACU” ed “ICU” (che sono rispettivamente gli acronimi di “*Post-Anesthesia Care Unit*”, ovvero “unità di cura post-anestetica”, e di “*Intensive Care Unit*”, ovvero “unità di cura intensiva”). In una “PACU” si vanno prevalentemente ad affrontare gli effetti dell’anestesia che il paziente ha dovuto subire preliminarmente all’intervento chirurgico, mentre in una “ICU” vengono invece trasferiti i pazienti che hanno subito interventi chirurgici complessi e che hanno perciò maggiormente bisogno di un adeguato controllo circa l’immediato decorso post-operatorio. Tali strutture in genere devono essere munite della strumentazione necessaria al monitoraggio delle funzioni vitali del paziente, di respiratori automatici per eventuale assistenza respiratoria e di tutti i presidi necessari per il trattamento di ulteriori complicanze (defibrillatori, farmaci antishock, flebo per infusioni endovenose...) o per l’erogazione dell’assistenza al paziente stesso (siringhe, materiale per le medicazioni, disinfettanti...). Andando a sintetizzare quanto suddetto circa la dettagliata analisi del “processo clinico” possiamo schematizzare il tutto con la seguente configurazione:

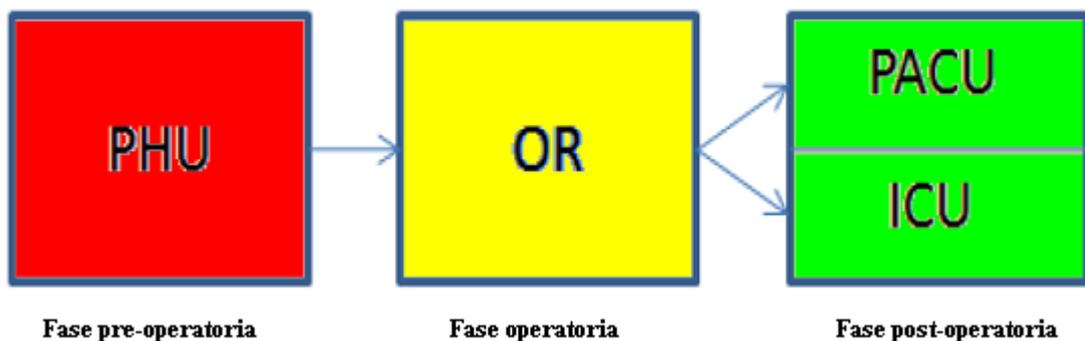


Figura 1.2: schema del “workflow clinico” relativo agli interventi chirurgici.

Passando invece a descrivere i modelli organizzativi e gestionali di riferimento in Italia c’è da dire che spesso negli ultimi anni sono state affrontate tematiche rilevanti al fine di migliorare la gestione dei blocchi operatori. Questo anche a causa dei suddetti fattori di insuccesso che troppe volte hanno condizionato la credibilità e la sicurezza del sistema sanitario nazionale e che sono stati inevitabilmente oggetto di discussione e di spunti critici.

A tal uopo, anche su un panel di testimonianze di esperti in questa fase, tali modelli si sono focalizzati su vari aspetti cardine legati a questo contesto proprio a rimarcare il fatto

che le sale operatorie rappresentano un punto nevralgico di tutto sistema sanitario, dato che in esse si svolgono quotidianamente operazioni ad alto livello di rischio e le cui attività vanno programmate dunque con la massima precisione possibile.

Sulla base di un'analisi completa del tema gestionale delle sale operatorie, in grado di non discernere da alcun aspetto né in termini organizzativi né in termini di sicurezza e che centralizza importanti concetti quali l'informatizzazione ed il controllo dei costi, si può dire che tali modelli per sommi capi propongono di prestare attenzione ad alcuni aspetti di svariata natura.

Andandoli ad elencare si hanno:

1. gli *“aspetti organizzativi e di programmazione”*;
2. gli *“aspetti di gestione del rischio e della sicurezza”*;
3. gli *“aspetti logistici”*;
4. gli *“aspetti tecnologici”*;
5. gli *“aspetti di controllo”*.

Gli *“aspetti organizzativi e di programmazione”* comprendono tutti quegli aspetti che si propongono di garantire la gestione ottimale del paziente nella sala operatoria ricorrendo all'apporto di modelli innovativi di costruzione ed organizzazione, quali alcuni dettami circa la programmazione delle attività, circa il monitoraggio dell'attività chirurgica e l'applicazione di misure necessarie a garantire la qualità.

Per quanto riguarda gli *“aspetti di gestione del rischio e della sicurezza”* tali modelli propongono una gamma di azioni di *“risk management”* da mettere in atto per evitare gli errori, per migliorare la qualità delle prestazioni e per garantire dunque la sicurezza del paziente.

A tal uopo le linee guida ed i database di *“best practices”* propongono una mappatura degli errori organizzativi ed operativi nonché principi di tracciabilità del percorso chirurgico che, unitamente all'apporto dell'informatica ed al coordinamento della documentazione sanitaria, possono permettere una migliore gestione di tale fase.

Gli *“aspetti logistici”* non sono invece da intendere soltanto come modalità di organizzazione dei flussi dei materiali, ma anche come gestione dei tempi nella sala operatoria.

Tra gli *“aspetti tecnologici”* non sono invece considerati soltanto concetti quali il ruolo dell'informatizzazione come strumento di coordinazione tra le sale operatorie ed i vari reparti ospedalieri e di gestione dell'attività operatoria nel suo complesso, ma anche

concetti quali gli indicatori per valutare gli aspetti di efficienza e di economicità di una sala operatoria.

A tal uopo si può dire che l'ANMDO, ovvero "*l'Associazione Nazionale dei Medici delle Direzioni Ospedaliere*", nell'arco dei suoi congressi ha avviato negli ultimi anni un gruppo di lavoro volto a concordare gli standard di misurazione delle attività dei reparti operatori al fine di renderli confrontabili con i meccanismi intrinseci di una sala operatoria, prestando così attenzione agli indicatori utili a monitorarne il buon utilizzo delle varie risorse.

Gli "*aspetti di controllo*" sono legati invece al controllo di gestione con particolare rilevanza sui centri di costo, definendo le modalità specifiche circa come individuarli e circa come tenere sotto controllo il loro processo di analisi, mirando altresì a stabilire driver per l'allocazione dei costi comuni. In questo ambito si introduce inoltre il concetto di "*balanced scorecard*", ovvero quello strumento di supporto nella gestione strategica che ne traduce il tutto in un insieme coerente di performance al fine di facilitarne la misurabilità, anche in questo contesto.

Tali modelli organizzativi e gestionali hanno il vantaggio di affrontare in maniera piuttosto dettagliata ogni aspetto inerente ad un sistema così ricco di criticità e di puntare a collocarsi come uno strumento di revisione in un approccio di tipo funzionale. Ma di contro essi hanno svantaggi in quanto sembrano essere caratterizzati da un'eccessiva burocratizzazione nonché da una standardizzazione di vari aspetti intrinseci che troppe volte portano a semplificazioni e ad assunzioni che non risultano essere sempre plausibili. Questo perché una peculiarità di un sistema sanitario consiste proprio nel fatto che le attività ed i conseguenti risultati dipendono strettamente dall'interdipendenza di una gamma di figure multi-professionali e che necessitano di una coerente ed accurata articolazione dei suddetti "*processi clinici*", che naturalmente dipendono strettamente dalle caratteristiche correlate al sistema sanitario di riferimento stesso.

1.3 Problemi di Pianificazione, Gestione e Controllo

Dopo aver analizzato in generale il contesto di riferimento ed aver descritto in maniera maggiormente dettagliata il "*processo clinico*" che avviene all'interno di una sala operatoria, si può passare finalmente a descriverne in maniera approfondita le problematiche legate alla fase di pianificazione, di gestione e di controllo.

Per prima cosa è importante introdurre la classificazione dei pazienti: in questo lavoro di tesi, come sarà possibile vedere nei capitoli successivi, i pazienti (che si possono denominare comunemente anche come “operazioni”, “interventi chirurgici” o “casi chirurgici”) vengono distinti in “elettivi” ed “urgenti”.

Per casi “elettivi” si intendono quelli che si vengono a verificare con maggior frequenza e che sono inerenti ai pazienti già in lista di attesa e per i quali è stato dunque possibile stabilire a priori nella fase di pianificazione il momento in cui deve avvenire la fase “operatoria”.

Per casi “urgenti” si intendono invece quelli che possono giungere da un momento all’altro nel sistema sanitario ed in cui il paziente presenta condizioni fisiche e funzioni vitali più o meno delicate, ragion per cui in ogni caso necessita con immediatezza dell’erogazione di un intervento chirurgico da parte del sistema stesso.

Proprio l’eventuale presenza di questi casi urgenti va naturalmente ad incidere in maniera significativa su quant’era stato pianificato in precedenza, perché a tali casi bisogna chiaramente dare precedenza causando così in parecchi casi un posticipo sull’inizio della fase “operatoria” per i casi elettivi.

In letteratura vi è comunque un’altra classificazione che va a raggrupparli in “emergenze”, “urgenze” ed “elettivi” [2] (con la differenza sostanziale rispetto alla classificazione precedente che le “emergenze” vanno trattate al massimo entro 2 ore dal momento in cui si presentano e le “urgenze” vanno invece trattate comunque entro poche ore).

Un’ulteriore classificazione identifica invece i casi chirurgici per categorie ed in questo caso si ha una prima categoria denominata A che va ad includere i casi elettivi che si vengono a verificare con una certa frequenza, una seconda categoria denominata B che va ad includere i casi elettivi ma che si vengono a presentare più raramente ed infine una terza categoria denominata C che va invece ad includere le emergenze [3].

La fase di pianificazione deve dunque tenere conto di questi vari aspetti nella loro possibile incertezza e deve indicare strategicamente i mezzi, gli strumenti, le azioni da adottare e le modalità con cui è necessario intervenire al fine di rendere i processi gestionali più coordinati, razionali, adeguati ed efficienti possibile.

A questa fase si affianca quella di non minor importanza di gestione delle risorse, la quale va a trattare invece l’esigenza di identificare con chiarezza tutti i principali processi gestionali ed organizzativi che vanno ad intervenire nell’intero “*processo clinico*”, definito come detto nel paragrafo precedente come quell’insieme di attività che assorbe

risorse ed utilizza metodiche e tecnologie finalizzate a produrre un risultato in termini di output che abbia un certo valore per il paziente.

Le attività che avvengono dunque all'interno di una sala operatoria sono organizzate per processi ed al fine di gestirli in maniera migliore è conveniente distinguerli in “*primari*”, che sono quelli fondamentali e che sono anche denominati “processi-chiave”, ed in “*secondari*”, che sono invece quelli supporto ai suddetti processi primari.

I processi primari che vanno a caratterizzare la fase di erogazione del servizio sanitario sono, per sommi capi, i seguenti:

1. addestramento e gestione del personale medico e paramedico;
2. pianificazione delle sale operatorie;
3. setup adeguato delle sale operatorie che devono erogare un determinato intervento chirurgico;
4. esecuzione degli interventi chirurgici;
5. gestione delle eventuali emergenze/urgenze;
6. attività di sterilizzazione e pulizia post-operatoria delle sale operatorie;
7. misurazione dei risultati ottenuti ed analisi di miglioramento.

Tale “mappatura” descrittiva dell'intero processo ne permette un monitoraggio che possa mettere in risalto anche le criticità e le opportunità di miglioramento al fine di ridisegnare successivamente, là dove possibile, i processi organizzativi in maniera ottimale.

La fase di controllo prevede per l'appunto l'identificazione degli aspetti gestionali-organizzativi al fine di evidenziarne eventuali peculiarità e criticità per poter trarne successivamente vantaggio.

Generalmente al fine di identificare e di analizzare le caratteristiche e l'entità delle peculiarità riscontrate nelle fasi di pianificazione e di gestione di una sala operatoria, è conveniente raccogliere i dati storici al fine di esaminare successivamente alcuni aspetti.

Ad esempio un aspetto che può essere importante esaminare è la durata effettiva di un generico intervento chirurgico per confrontarla con la relativa programmazione, delineando così il problema e passando quindi ad analizzarne le cause di un eventuale scostamento avvalendosi magari dell'utilizzo di un classico diagramma “causa-effetto”.

Successivamente quindi si potrebbero identificare le soluzioni, anche mediante un approccio di brainstorming volto a ricavare dalle parti interessate una gamma di possibili soluzioni ed interventi migliorativi.

Tale fase, pertanto, è necessaria dunque per far sì che l'output si mantenga conforme alle specifiche ed agli standard prestabiliti, mantenendo pressoché costanti le modalità

operative e gli elementi variabili del processo stesso andando così ad effettuare, là dove possibile chiaramente, una sorta di standardizzazione delle soluzioni considerando le alternative più plausibili e per le quali sono definite procedure specifiche.

1.3.1 Esigenze di efficienza ed efficacia

Le suddette fasi di pianificazione, di gestione e di controllo della sala operatoria hanno dunque l'obiettivo di far sì che il sistema sanitario funzioni in termini di efficienza ed è proprio per questo che è fondamentale la correttezza della loro elaborazione ed esecuzione, nonché l'adeguato coinvolgimento di tutti gli attori che ne entrano in gioco.

Tuttavia l'efficienza non è l'unico aspetto a cui si deve rispondere in questo contesto, poiché ad essa si affianca anche il concetto di efficacia: per prima cosa bisogna distinguere i due aspetti, che troppe volte potrebbero venir confusi.

C'è da dire allora che mentre l'**efficienza** misura l'impiego economico delle risorse nel "processo produttivo" ed è definita come il rapporto prestazioni/risorse, l'**efficacia** misura invece il contributo dei servizi sanitari al miglioramento dello stato di salute del paziente (definito altresì come stato di "outcome") ed è definita invece come il rapporto prestazioni/salute.

Il principio dell'efficacia sanitaria è dunque che ogni "azione" (un consulto medico, la somministrazione di un farmaco, una seduta e chiaramente anche un intervento chirurgico...) ha valore soltanto se contribuisce a migliorare lo stato di salute del paziente: ogni € speso per una prestazione sanitaria deve tradursi in un certo ammontare di "salute guadagnata".

In termini generali si può dire che l'efficienza è misurata dal numero di prestazioni realizzate (ad esempio il numero di interventi chirurgici effettuati per una determinata specialità), mentre l'efficacia è misurata dal miglioramento dello stato di salute dopo una prestazione sanitaria (ad esempio il progresso dal coma alla piena coscienza di un paziente dopo aver subito un intervento chirurgico di una certa consistenza).

Inoltre una misura sintetica è data dal concetto di "*rendimento*", definito come il rapporto tra salute e risorse.

Andando ad utilizzare un po' di formule matematiche si può dire dunque che:

$$1. \text{ Efficienza} = \frac{\text{Output}}{\text{Input}},$$

$$2. \text{ Efficacia} = \frac{\text{Outcome}}{\text{Output}},$$

$$3. \text{ Rendimento} = \text{Efficienza} * \text{Efficacia} = \frac{\text{Outcome}}{\text{Input}}.$$

In questo caso per input si intendono tutte le risorse che sono impiegate per l'esecuzione del processo clinico (lavoro, capitale...) e per output si intendono invece le prestazioni (intese come visite, analisi, ricoveri...), mentre, come suddetto, per outcome si intende lo stato di salute del paziente.

Il concetto di rendimento è molto importante per la valutazione delle prestazioni sanitarie, poiché un sistema sanitario non deve essere né solo efficiente e né solo efficace ma deve fornire un giusto "trade-off" dei due aspetti: per rendersene ulteriormente conto si può fare un esempio.

Se un intervento chirurgico in una sala operatoria è stato rapido ma le condizioni di salute del paziente sono peggiorate, vuol dire che esistono problemi di inefficacia mentre se per esempio un paziente necessita di un intervento delicato ma lo si è potuto operare solo dopo una settimana di ricovero, vuol dire che esistono problemi di inefficienza.

Unendo le esigenze di efficienza e di efficacia alla già citata possibilità di avere nel sistema sanitario l'arrivo di casi urgenti dobbiamo introdurre altri concetti: si può iniziare dicendo innanzitutto che ogni sala operatoria ha un certo monte-ore lavorativo giornaliero, ossia un intervallo di tempo con un inizio ed una fine specificata e che può variare anche durante i giorni della settimana così come può rimanere sempre uguale.

A seguito della pianificazione delle attività dovuta alla presenza dei casi elettivi da trattare ed alla possibilità che essa sia soggetta a modifiche in corso d'opera a causa dell'arrivo di casi urgenti, ad ogni sala operatoria andrà un certo carico di lavoro che è possibile ovviamente quantificare in ore.

Si avrà allora che la differenza tra il carico di lavoro di ogni sala operatoria ed il numero di ore giornaliere di lavoro regolare forniranno delle indicazioni ben precise: se il numero di ore giornaliere di lavoro regolare eccederà il carico di lavoro della sala operatoria stessa, si andrà a parlare allora del concetto di "sotto-utilizzo" della sala operatoria, dato che vi è una porzione di tempo in cui la sala operativa rimane inattiva (proprio questa differenza si va ad identificare come "idle-time", ovvero per l'appunto "tempo di inattività").

Viceversa qualora il carico di lavoro della sala operatoria andrà ad eccedere il numero di ore giornaliere di lavoro regolare della sala operatoria si andrà invece a parlare del cosiddetto concetto di "sovra-utilizzo" ed ad esso si affianca anche quello di "over-time" (od "overtime", ma anche semplicemente "OT"), che rappresenta il lasso di tempo di lavoro straordinario di una sala operatoria, che deve chiaramente essere retribuito

maggiormente al personale medico e paramedico che lo svolge e che comporta quindi dei costi straordinari di gestione.

Chiaramente è comprensibile che questi costi di overtime, se necessari perché il carico di lavoro eccede il numero di ore lavorative regolari, non possono essere evitati poiché un intervento chirurgico non gode della cosiddetta proprietà di “preemption” (una volta iniziati bisogna chiaramente portarli a compimento, non si può di certo interromperli perché è scaduto l’orario di lavoro regolare e continuarli in istanza successiva!).

Quindi si è compreso come anche i cosiddetti concetti di “sotto-utilizzo” e di “sovra-utilizzo” della sala operatoria vadano ad incidere sulla misura della sua efficienza e chi si occupa delle fasi di pianificazione e di gestione di una sala operatoria non può certo trascurarli.

1.3.2 Esigenze di sicurezza: il concetto di rischio clinico

Accanto alle suddette esigenze di efficacia e di efficienza, vi è all’interno di una sala operatoria anche quella non meno importante di **sicurezza**.

Per sicurezza in sala operatoria si va ad intendere un insieme di tecniche e di metodi volti a minimizzare tutti i fattori di rischio per il paziente.

Lo scopo è quello di andare a soddisfare le necessità di interventi chirurgici, di qualsivoglia natura essi siano, all’interno di un ambiente sicuro ed agibile, con l’utilizzo di figure professionali e di tecnologie appropriate ed in un contesto organizzativo flessibile. Per tali ragioni è d’uopo pianificare e configurare adeguatamente un ambiente operativo che rispecchi le suddette caratteristiche (assegnando, per esempio, una sala operatoria equipaggiata per una certa specialità chirurgica ad un team chirurgico munito di figure professionali che hanno maggior esperienza e dimestichezza con gli strumenti e le tecnologie in essa presenti).

In Italia ormai da anni vi sono continue iniziative volte a garantire maggior sicurezza nei sistemi sanitari, come ad esempio manuali di procedure che comprendono al loro interno una serie di raccomandazioni e di linee-guida da applicare in maniera più o meno rigorosa in contesti specifici per il raggiungimento di determinati obiettivi in base ad alcune problematiche di fondo che potrebbero venire a presentarsi. Inoltre in tali manuali è in genere annoverata l’utilità della metodologia “*check-list*” quale strumento-guida per l’esecuzione dei controlli e con la finalità di favorire l’implementazione degli standard di sicurezza raccomandati al loro interno.

Purtroppo però, malgrado le suddette iniziative, la cronaca quotidiana nazionale racconta, come detto anche nello scorso paragrafo di questo capitolo, troppo spesso di incidenti in sala operatoria che allarmano l'opinione pubblica e che compromettono pertanto la fiducia nella sanità, oltre a coinvolgere legalmente i responsabili della gestione delle stesse strutture.

In generale la sicurezza è data dal controllo degli eventi avversi e dall'annullamento degli stessi: proprio per questo è necessario garantire attività di sorveglianza mediante l'apporto di procedure e di metodologie standardizzate che puntino a ridurre le criticità. Le criticità sono per l'appunto gli aspetti che si ritrovano nell'analisi dei fattori di rischio e vanno esaminate in maniera dettagliata al fine di garantire la sicurezza dei pazienti e degli operatori. Analizzando la natura di questi eventi avversi, ci si può riferire ad una statistica che è stata condotta negli ultimi anni [4] e che va a stabilire le cause dei suddetti eventi all'interno delle sale operatorie. Tale statistica va innanzitutto ad informare che proprio nel "blocco operatorio" si vengono a manifestare addirittura il 52% degli eventi avversi complessivi, che non è certo una percentuale di poca rilevanza e che sottolinea quindi ancor di più quanto sia importante la fase di pianificazione e di gestione di una sala operatoria.

La stessa statistica evidenzia inoltre le cause stesse, specificando che quelle maggiormente incidenti sono dovute ad inefficienze organizzative o a problemi di natura assistenziale ed in minor misura ad aspetti quali carenze ed inadeguatezze tecnologiche, somministrazioni di farmaci vari al paziente in maniera inesatta, errori medici o condizioni fisiche particolari del paziente stesso.

Per farsene maggiormente un'idea, ecco qui di seguito riportati i risultati completi della statistica suddetta:

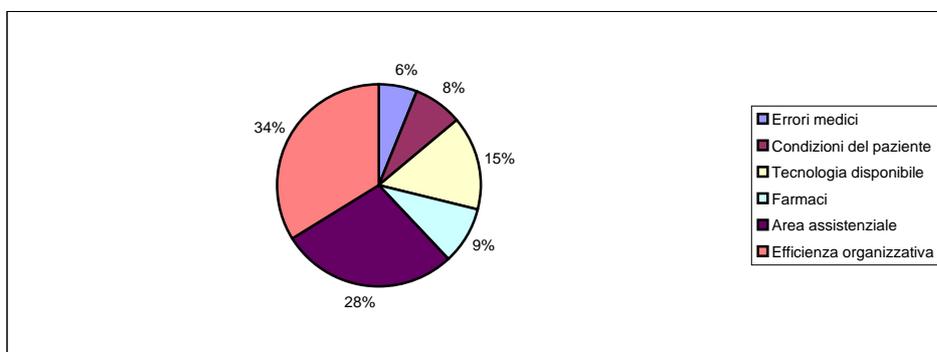


Figura 1.3: cause degli eventi avversi che possono avvenire in una sala operatoria.

Pertanto è facile capire che i fattori di rischio da valutare sono molteplici e che, riferendosi all'assistenza verso il paziente, si può parlare di problemi di identificazione (documentazione del paziente), di mobilitazione, di sterilità e di rischio di ustioni; mentre riferendosi all'efficienza organizzativa, intesa come utilizzo di strumenti e di apparecchiature, vi possono essere invece rischi elettrici, biologici, ionizzanti e microclimatici.

Scendendo di più nel dettaglio si può parlare pertanto di “*rischio clinico*”, una grandezza che si può definire in funzioni di alcune variabili ovvero:

1. il “*rischio chimico*”;
2. il “*rischio meccanico*”;
3. il “*rischio ergonomico*”;
4. il “*rischio biologico*”;
5. il “*rischio elettrico*”;
6. il “*rischio fisico*”;
7. il “*rischio psicologico*”;

Il “*rischio chimico*” va prevenuto mediante l'impiego di impianti di ripulitura e cambiamento dell'aria.

Il “*rischio meccanico*” è inteso come sicurezza nelle movimentazioni di mezzi e macchinari e di posizione del paziente in maniera corretta ed anatomica durante lo svolgimento dell'intervento chirurgico.

Il “*rischio ergonomico*” è legato al fatto che per la sicurezza è necessario che sia il personale medico/paramedico che il paziente siano muniti di tutte le comodità necessarie evitando posizioni scomode).

Il “*rischio biologico*” è invece uno dei fattori preponderanti e si deve prevenire favorendo la sterilità dell'ambiente, degli strumenti e dei materiali nonché rispettando le apposite normative di sicurezza e di qualità.

Il “*rischio elettrico*” comprende possibilità di guasti di natura elettrica che possono contaminare la sicurezza nonché l'incolumità fisica del paziente in particolare e che potrebbero richiedere di studiare ad esempio un apposito piano di isolamento per proteggerlo dalle correnti di dispersione.

Il “*rischio fisico*” considera il fatto che tutti gli attori coinvolti nel processo operatorio, incluso ovviamente il paziente, possono essere minacciati dalla possibilità di pericoli di infortunio talvolta assai gravi e per questo motivo è d'uopo prendere tutti gli accorgimenti

necessari al fine di evitare che macchinari, attrezzature o strumenti vengano utilizzati in maniera scorretta o errata e che si possa minare quindi la sicurezza di qualcuno.

Il “*rischio psicologico*” racchiude invece un insieme di fattori che spesso non vengono ritenuti importanti, ma che possono determinare problemi.

Il personale medico/paramedico durante l'intervento è infatti sottoposto in genere a condizioni di stress e di tensione emotiva, ragion per cui la tecnologia deve essere d'apporto alla sicurezza fornendo macchinari, strumenti ed attrezzature di semplice utilizzo.

Ad esempio può essere d'ausilio la possibilità per il suddetto personale di avere a disposizione durante l'esecuzione dell'intervento chirurgico emettitori di musica che possano ridurre tale stress e tale tensione emotiva, al fine di massimizzare dunque la sicurezza sotto questo punto di vista.

In sintesi dunque di quanto appena menzionato si ha che: *Rischio clinico* = *funzione(Rischio chimico, Rischio meccanico, Rischio ergonomico, Rischio biologico, Rischio elettrico, Rischio fisico, Rischio psicologico)*.

Negli ultimi anni il sistema sanitario ha ricercato delle soluzioni sempre nuove per ridurre il suddetto “*rischio clinico*” del paziente: una di queste è per esempio la “*cartella infermieristica*” che non è copia della nota “*cartella clinica*” ma è piuttosto definibile come un documento sanitario di natura giuridica e di competenza per l'appunto dell'infermiere in autonomia, circa la valutazione dello stato di salute del paziente.

Si tratta perciò di uno strumento informativo capace di registrare, gestire e documentare tutti gli aspetti che possono rivelarsi importanti al fine di garantire al paziente un'assistenza maggiormente qualificata e soprattutto volta a garantirne la sicurezza.

Un altro elemento che può rivelarsi importante sotto questo punto di vista è l'introduzione della metodologia “*check-list*” di autovalutazione per i sistemi sanitari, al fine di raccogliere le informazioni che possono rivelarsi necessarie per prevenire gli eventi avversi studiandone le possibili cause ed identificandone le criticità.

In tale caso si va a porre l'attenzione su aspetti fondamentali quali il monitoraggio continuo del paziente, la pianificazione dell'assistenza, l'accurata gestione del registro operatorio (con la definizione delle responsabilità dei singoli attori che entrano in gioco in tale processo) ed aspetti legati all'igiene del personale medico e paramedico.

I punti di forza dell'utilizzo della suddetta metodologia sono legati al conferimento di un automatismo procedurale ed al miglioramento della cultura della comunicazione fra i vari attori.

Di contro, però, i punti di debolezza sono legati al pericolo di assoggettamento del personale alle sue linee-guida (incapacità di trattare nuovi problemi che possono sorgere, dato che chiaramente non esiste il concetto di assoluta ripetitività degli eventi avversi).

Per migliorare la qualità dell'assistenza bisogna pertanto favorire la continuità delle prestazioni ed organizzare in maniera sistematica le cure a tutto vantaggio dell'efficacia e della sicurezza, dando così garanzia al paziente di ridurre al minimo possibile le probabilità di danno accidentale. Ciò avviene ovviamente in base anche alle conoscenze tecnico-scientifiche attuali ed ai dettami di benchmarking sanitario e dei corsi di addestramento e di formazione per il personale, che nascono proprio con la prerogativa di passare dalla cultura del "*rischio clinico*" alla cultura della "*sicurezza del paziente*" ("*patient safety*").

Per tutte queste considerazioni suddette si viene così a parlare, da un punto di vista filosofico, di "*etica dell'errore*" [5], ovvero di un principio che consiste in alcuni principi fondamentali che devono osservare gli attori che entrano in gioco nel "processo clinico" di una sala operatoria.

Tra questi principi si possono annoverare l'utilizzo sapiente delle competenze per un lavoro diligente con buone attitudini per il teamworking, la riduzione dell'errore in base all'analisi dei processi più critici e l'analisi dettagliata delle cause di un errore, una volta che questo è stato generato.

Inoltre altri principi annoverabili in tale contesto riguardano l'accettazione della sfida per il passaggio dal "*risk management*" alla "*safety culture*" (che non vuol dire annullare completamente la possibilità di errori e di eventi avversi perché è chiaramente un qualcosa di pressoché impossibile, ma quantomeno cercare di far sì che gli errori siano in misura piuttosto contenuta); l'introduzione di una politica "*no blame*" ("nessuna colpa") basata sull'onestà e l'umiltà, ed infine il ruolo del "*feedback*" del paziente e delle persone ad esso più care (al fine di raccogliere una gamma esaustiva di informazioni che possono rivelarsi utili ad evitare errori successivi).

In generale c'è da dire che riconoscere un errore è parte integrante della natura umana, anche per figure con eccellenti standard professionali come possono essere gli attori che entrano in gioco in un processo clinico.

E chiaramente anche in questo settore di riferimento dagli errori si deve solo imparare, cercando di prevenirli successivamente studiandone sia le cause che le possibili soluzioni. Quando si viene a manifestare nel sistema un errore bisognerebbe chiedersi "come e perché" è successo, piuttosto che chiedersi, come succede in genere, "per colpa di chi" è

successo, prestando così in maniera errata maggior enfasi alle responsabilità ad esso legate piuttosto che alle motivazioni che ne sono state realmente causa.

Questo equivale ad introdurre altresì sistemi di verifica al fine di massimizzare la sicurezza, anche se c'è da considerare che è assai difficile per non dire pressoché impossibile raggiungere uno standard in cui il paziente in una sala operatoria non correrà con certezza alcun rischio da errore o da evento avverso.

1.3.3 Il ruolo dell'innovazione ed integrazione

Per il raggiungimento degli standard di efficacia, efficienza e di sicurezza, come illustrato e discusso negli scorsi sottoparagrafi, le sale operatorie moderne possono contare sulla predisposizione di elementi che ne permettono l'esecuzione degli interventi chirurgici in un modo migliore sia per il personale medico (che può operare in un ambiente più sicuro ed efficiente), sia per il paziente (che può contare di essere operato in un ambiente più sicuro ed efficace).

Le sale operatorie moderne, grazie alle tecnologie innovative proposte negli ultimi anni, vantano infatti:

1. dispositivi di integrazione e controllo;
2. utilizzo del “data management”;
3. sistemi di comunicazione avanzati;
4. varie apparecchiature innovative.

Per quanto riguarda i dispositivi di integrazione e controllo c'è da sottolineare proprio quanto in un sistema che si propone di essere sicuro e di qualità siano importanti tali fattori: i locali adibiti all'esecuzione di un intervento chirurgico devono infatti soddisfare in pieno le esigenze del personale medico ed assisterlo in ogni suo bisogno.

Il compito principale di questi dispositivi è proprio quello di andare a supportare il personale medico stesso, assistendolo nel rispetto rigoroso delle singole procedure che compongono un intervento chirurgico al fine di far sì che esso possa essere svolto con una certa qualità.

A tal uopo la tecnologia negli ultimi anni propone l'utilizzo di soluzioni di tipo informatico con telecontrollo o comunque con generici sistemi di video-registrazione per la visione remota delle sale operatorie.

Il loro obiettivo è quello di permettere il controllo dell'operato del personale per verificare non solo eventuali il manifestarsi di disattenzioni (come ad esempio il mancato

rispetto delle norme sull'apposito vestiario che un membro del personale medico/paramedico deve indossare durante l'esecuzione di un intervento chirurgico), ma anche per effettuare specifiche registrazioni sull'intervento specifico al fine di favorire studi futuri di ricerca, sempre nel rispetto della privacy di tutti gli attori che entrano in gioco in questa fase.

Il vantaggio del loro utilizzo sta anche da un punto di vista "psicologico", perché è ovvio pensare che quando una persona sa di essere sempre sotto il controllo vigile di un sistema di registrazione è più predisposta ad agire rispettando adeguatamente tutte le procedure che deve espletare in maniera ferrea e rigorosa, riducendo al minimo le sue possibilità di errore dovute a fattori di distrazione.

Passando a parlare dell'utilizzo dei "data management", c'è innanzitutto da dire che si tratta di un sistema informatico per la raccolta dei dati e delle informazioni relative ad un certo prodotto/servizio: proprio un sistema di questo genere va a costituire la memoria centrale per il controllo dei dati avvalendosi di tutti gli strumenti necessari (stampanti, grafici o tabelle, supporti di registrazione...) per sintetizzare o per ritrasmettere poi i dati e le informazioni da esso memorizzate e che gli sono richieste.

I "data management" più noti in letteratura sono il SIO (acronimo di "*Sistema Informativo Ospedaliero*"), che viene impiegato come una sorta di archivio anagrafico/storico-clinico per il paziente in modo da avere maggior reperibilità delle informazioni, e l'AIDA (acronimo di "*Advanced Image and Data Archivi*"), sistema utilizzato invece per acquisire dati provenienti dalle sale operatorie quali filmati o immagini che sono stati registrati.

Per quanto riguarda invece i sistemi di comunicazione avanzati, la telematica e l'informatica hanno fatto passi enormi in avanti anche sotto questo punto di vista nell'ambito delle sale operatorie e continuano sempre a progredire grazie ad innovazioni tecnologiche sempre nuove.

Le motivazioni d'utilizzo di tali sistemi sono dovute al fatto che la sala operatoria non può essere un luogo isolato, ma deve essere sempre un centro di scambio continuo di dati e di informazioni: la possibilità di inviare all'esterno di un certo campo operatorio registrazioni video/audio permette maggiori possibilità di analisi e di collaborazione tra il personale medico.

Gli elementi principali di questi sistemi sono i telefoni e/o i microfoni (che permettono di comunicare senza il bisogno di allontanarsi dalla sala), lo streaming (ossia la possibilità di trasmettere un flusso di dati video/audio a più destinazioni mediante l'impiego di una rete telematica) e la possibilità di videoconferenze.

In questo campo sono poi sicuramente degne di nota innovazioni quali la “telecardiologia” e la “tele-radiologia”, che prevedono grazie all’utilizzo di sistemi informatici la trasmissione in tempo reale di dati ed immagini inerenti al rispettivo campo di applicazione.

Tali innovazioni permettono vantaggi non irrilevanti perché grazie al loro impiego non è più necessario dover predisporre tutti i macchinari diagnostico-terapeutici in ogni singola sala operatoria.

In sintesi questi sistemi di comunicazione avanzata permettono di trarre grossi benefici sia dal punto di vista dei costi, sia dal punto di vista degli spazi.

Dal punto di vista dei costi, difatti, non è più necessario “sterilizzare” tutti i macchinari, non servono più parecchi materiali specifici nelle sale operatorie ed è inoltre possibile far utilizzare lo stesso macchinario diagnostico-terapeutico e di acquisizione dati a più utenti contemporaneamente ed a più sale operatorie, evitando così di dover predisporre un’apparecchiatura specifica in ogni sala operatoria.

Dal punto di vista degli spazi, invece, non essendo più necessaria la predisposizione di un’apparecchiatura specifica all’interno di ogni sala operatoria, com’è stato appena puntualizzato, il personale medico/paramedico può contare su una migliore possibilità di movimento durante l’esecuzione di un intervento chirurgico e la sala operatoria può anche essere grazie a questo vantaggio pulita in maniera migliore, portando perciò benefici anche dal punto di vista igienico.

Per quanto riguarda infine le apparecchiature innovative, c’è da dire che sotto questo punto di vista l’innovazione tecnologica ha determinato un cambiamento delle metodiche e delle condizioni operative andando a rendere la sala operatoria sempre più adattabile al tipo di specialità chirurgica richiesta da un determinato intervento.

Per prima cosa i tavoli operatori moderni godono di un design in grado di offrire maggior flessibilità, ossia migliori possibilità di configurare il piano operatorio nel modo più convenzionale per l’esecuzione del tipo di intervento chirurgico da svolgere su di esso, grazie anche all’apporto di optional applicabili.

Ad esempio vi è l’utilizzo di “piastre piane” che permettono la stabilità e l’isolamento elettrico del paziente durante l’intervento chirurgico e che consentono inoltre al paziente stesso di essere comodo e di stare nella posizione corretta durante l’esecuzione dell’intervento chirurgico stesso.

Le ultime frontiere tecnologiche permettono anche di scegliere se comandare il movimento del letto operatorio mediante l’utilizzo di un supporto a raggi infrarossi

(telecomando) o uno a cavo e pedaliera, permettendo di sistemare anche il piano-gambe e lo schienale in maniera elettro-motorizzata.

Ma l'innovazione tecnologica circa l'aspetto dei tavoli operatori non si ferma qui: si punta infatti a creare piani trasferibili integrati capaci anche di memorizzare funzioni e posizioni e di visualizzare lo stato del tavolo stesso grazie anche alla presenza di pulsanti retro-illuminati che informano l'operatore in tempo reale sullo stato del sistema (gradi di inclinazione effettuati, posizioni raggiunte...).

Un altro fattore importante è legato invece all'illuminazione e l'innovazione tecnologica sotto questo punto di vista offre oggi delle particolari lampade scialitiche versatili, orientabili, dalla vita utile di lunga durata e che minimizzano la formazione di ombre grazie ad una completa illuminazione sia in superficie che in profondità.

Oltre ad esse, naturalmente, in ogni sala operatoria vi deve essere anche una lampada alogena ausiliaria, che si deve usare in caso di avaria all'impianto luminoso scialitico.

Per quanto riguarda invece la fase anestesiológica, l'innovazione tecnologica mira a migliorare anche sotto questo punto di vista grazie all'introduzione di appositi macchinari moderni denominati "work-stations" e che sono stati studiati per impostare alcuni parametri target in maniera sicura per il paziente.

Tali macchinari sono caratterizzati dalla presenza di miscelatori e raccoglitori interni ed integrati e funzionano inoltre a "circuito chiuso", fattore che offre vantaggi sia per un miglior utilizzo delle risorse sia per i costi dato che i gas alogenati utilizzati sul paziente possono anche essere riutilizzati.

Inoltre essi sono muniti di un unico monitor, cosa che permette al medico anestesista di concentrarsi meglio sul punto visivo e grazie al fatto che essi sono dotati di schermo in modalità "touch-screen" lo stesso medico anestesista può programmare e configurare delle funzioni in maniera migliore, dato che esse erano tradizionalmente configurabili in modalità manuale.

La compattezza e l'avanguardia tecnologica di questi macchinari permettono inoltre una buona integrazione con la sala operatoria, specie nei casi di spazio ridotto.

Da un punto di vista futuro, comunque, non è lontanamente ipotizzabile che l'innovazione tecnologica progredendo sempre più potrà spingere, sfruttando l'ormai integrazione perfetta (simbiosi) tra uomo e macchina, a processi operatori evolutivi ancora più avanzati e che lasciano presagire che all'interno delle sale operatorie potrebbe entrare in gioco la "robotica".

Il paziente, secondo questa configurazione futura, sarebbe sempre assistito durante l'intervento chirurgico (nella fase squisitamente "operatoria" pertanto) dal medico

anestesista, ma la novità potrebbe essere rappresentata dal fatto che ad operarlo sarebbe un robot assistito da un tecnico specializzato, il quale interverrà manualmente solo in caso di emergenza.

Ciò porterebbe vantaggi del tipo minor presenza di personale medico/paramedico, ma di contro servirà personale tecnico specializzato adatto al controllo ed alla gestione di tali impianti ad elevato livello tecnologico.

1.3.4 Qualità nei processi clinici

Per quanto riguarda invece l'importante aspetto della **qualità**, si è già detto nel corso dei paragrafi precedenti che la salute del paziente deve rappresentare il primo obiettivo da raggiungere.

Proprio per il raggiungimento di tale scopo il personale medico/paramedico deve sfruttare al massimo le proprie competenze e le proprie conoscenze, cosicché si possa rispondere nel miglior modo alle esigenze di cura del paziente stesso.

Non a caso il regolamento della “Carta dei servizi pubblici e sanitari” come primo diritto del paziente recita quanto segue [6]: *“Il paziente ha diritto ad essere assistito e curato con premura ed attenzione, nel rispetto della dignità e delle proprie condizioni filosofiche e religiose”*.

Dunque è indispensabile rispondere a dettami di qualità ben specifica al fine di limitare le possibilità che il paziente sia esposto a rischi e pericoli (dei quali si è già ampiamente parlato nei paragrafi precedenti di questo capitolo) ed a tal uopo il tema della ricerca della qualità sanitaria è da anni ormai al centro dell'attenzione.

Tale aspetto appare perciò un elemento essenziale per l'erogazione di un servizio che possa essere il più soddisfacente possibile per il paziente ed i sistemi per il miglioramento della qualità, dunque, rappresentano il fulcro per la realizzazione di un cambio strategico inerente a quest'ambito.

Questo miglioramento è connesso alla misurazione di input (risorse), prestazioni erogate, “processi clinici” ed output (risultati ottenuti) e pertanto è vario anche l'insieme dei soggetti che ne sono portatori di interesse ovvero pazienti, medici, infermieri e figure che si occupano della fase di pianificazione, di gestione e di controllo della sala operatoria.

Chiaramente i pazienti sono i più interessati all'accessibilità ed alla bontà delle prestazioni, mentre medici ed infermieri prestano maggiormente attenzione ad aspetti di qualità tecnica-professionale ed ai risultati clinici ottenuti: l'interazione tra di loro ed il

paziente può così conciliarne alcuni aspetti, dando apporto alla qualità tecnica della cura e rendendo il servizio erogato più appropriato e soddisfacente.

Proprio a tal uopo è necessario sottolineare pertanto che la qualità dell'interazione fra medico e paziente dipende da fattori correlati quale la qualità della comunicazione, le capacità del medico di acquisire la stima e la fiducia del paziente, l'empatia e la sensibilità.

Chi si occupa della fase di pianificazione, di gestione e di controllo della sala operatoria, invece, rivolge comprensibilmente maggior attenzione ad aspetti quali la produttività e l'adeguato utilizzo delle risorse, considerando dunque la qualità dal punto di vista economico.

Dal punto di vista del paziente aspetti che possono essere fondamentali per il suo soddisfacimento e che garantiscano dunque per lui la qualità del servizio sono invece il riconoscimento delle sue preferenze (per esempio dargli la possibilità di scegliere quale medico specialista deve eseguire il suo intervento chirurgico) e la minimizzazione dei tempi d'attesa (aspetto di palese importanza).

A tale aspetto si collega strettamente quello dell'efficacia (definito nel primo sottoparagrafo di questo terzo paragrafo del primo capitolo) ed a mettere in un certo senso a connubio questi due fattori va a rivestire un ruolo di fondamentale importanza la "centralità" del paziente.

Con tale aspetto si intende la condivisione delle decisioni mediche da adottare tra i medesimi medico e paziente, con il fine di migliorarne la "qualità della vita" rispettandone e tutelandone l'autonomia oltre che migliorandone lo stato della salute.

La qualità sanitaria, pertanto, è funzione di alcuni aspetti fondamentali quali la "*qualità strutturale*" (l'adeguatezza degli spazi e dei servizi di accoglienza e tecnici, oltre che la competenza del personale), la "*qualità tecnico-strumentale*" (l'adeguatezza delle metodiche e degli strumenti che vengono utilizzati nell'ambito del "processo produttivo"), la "*qualità relazionale-organizzativa*" (intesa come il grado di interazione tra la struttura sanitaria ed il beneficiario del servizio) e la "*qualità gestionale*" (appropriatezza dell'azione rispetto agli obiettivi da raggiungere).

Dopo aver approfondito il significato del concetto di qualità nell'ambito di una sala operatoria si può passare ora a descriverne le modalità con cui essa viene misurata e verificata.

Innanzitutto bisogna distinguere tre voci principali ovvero "*fattori di qualità*", "*indicatori di qualità*" e "*standard di qualità*".

I “*fattori di qualità*” racchiudono gli aspetti rilevanti per far sì che il paziente percepisca la bontà del servizio (ad esempio la semplicità e la tempestività nella fase di prenotazione di una generica visita di controllo post-operatorio) e possono essere rappresentati sia da aspetti oggettivi che da aspetti soggettivi (rilevabili soltanto raccogliendo le percezioni del paziente).

Gli “*indicatori di qualità*”, invece, sono delle variabili quantitative o comunque dei parametri che vanno a misurare le prestazioni del servizio: ad ogni fattore di qualità (esempio di prima: tempestività nella fase di prenotazione di una visita di controllo post-operatorio) può corrispondere in genere almeno un indicatore di qualità (seguendo l'esempio appena menzionato: tempo di attesa per la prenotazione della visita suddetta, tempo di attesa tra la prenotazione e l'effettuazione della visita...).

Tali indicatori possono essere, a loro volta, di diversi tipi: “*indicatori di struttura*” (legati alle rilevazioni periodiche sullo stato di strutture e procedure), “*indicatori di processo*” (misure o valutazioni sull'effettivo svolgimento delle attività) ed “*indicatori di esito*” (inerenti alla soddisfazione del paziente, raccogliendone le valutazioni con appositi strumenti).

Gli “*standard di qualità*”, invece, sono i valori attesi per determinati indicatori e vanno a rappresentarne perciò gli obiettivi che una volta resi pubblici al paziente vanno a diventare una sorta di “livello promesso” del servizio.

A loro volta gli standard di qualità si dividono in “*standard generali*” (riferiti al complesso della prestazione resa ed espressi generalmente da valori medi statistici o da percentuali di successo attese rispetto ad uno standard specifico) ed in “*standard specifici*” (riferiti invece a ciascuna delle prestazioni erogate al paziente, di modo che esso possa verificarne direttamente il rispetto).

Gli “standard specifici” sono in genere espressi da una soglia minima e da una soglia massima relativa ai valori che esso può assumere, come per esempio l'intervallo di giorni minimo e massimo che può trascorrere dal momento in cui un paziente prenota una visita di controllo post-operatorio al momento in cui la sua richiesta verrà soddisfatta.

1.3.5 Fattori di incertezza

Nel corso dei sottoparagrafi precedenti si sono definite in maniera piuttosto esaustiva le condizioni al contorno nell'ambito di un processo operatorio e da esse si evince in maniera piuttosto evidente che in un sistema sanitario, poiché la medicina in sé è una

scienza probabilistica e non certo deterministica, sia il personale medico e paramedico, sia chi si occupa delle fasi di pianificazione, di gestione e di controllo di una sala operatoria, sia il paziente ovvero in sintesi tutti gli attori che entrano in gioco nell'ambito di questo complicato processo vanno a farlo in condizioni di incertezza.

Per prima cosa è d'uopo specificare che l'incertezza non va identificata con il rischio, nozione della quale si è già discusso e con la quale essa viene a volte confusa.

Questo perché mentre l'incertezza è uno stato di conoscenza limitata in cui è impossibile descrivere esattamente lo stato esistente ed i risultati futuri, il rischio è piuttosto uno stato di incertezza in cui alcuni possibili risultati futuri potrebbero avere con una certa probabilità un effetto indesiderato [7].

Facendo un esempio semplice rimanendo nel contesto di riferimento di questo lavoro di tesi, si può dire che una definizione del tipo "non si sa se domani giungerà al sistema sanitario un caso urgente dovuto alle conseguenze di un sinistro stradale" manifesta una chiara e comune situazione di incertezza.

Si può però applicarne la probabilità di risultati possibili utilizzando ad esempio una valutazione calibrata del tipo "oggi è sabato, quindi stasera è più probabile rispetto alle altre sere che ci siano molte persone che vadano a divertirsi fino alle prime ore dell'alba e quindi, tornando a casa per esempio dopo che qualcuno ha pure bevuto un po' troppo, potrebbero purtroppo verificarsi con maggiore probabilità sinistri stradali".

Quantificando quindi l'incertezza con una definizione basata magari su dati storici, sempre per esempio, del tipo "al 60% domani, ossia domenica, non arriveranno al sistema sanitario casi urgenti dovuti alle conseguenze di un sinistro stradale", si va a stabilire sommariamente in termini di condizioni di valutazione del rischio che "al restante 40% nella giornata di domenica potrebbero presentarsi al sistema sanitario casi urgenti dovuti alle conseguenze di un sinistro stradale".

Dunque le due nozioni di "incertezza" e di "rischio", come ha puntualmente dimostrato l'esempio suddetto, sono correlate ma non sono identiche.

Tornando nello specifico, comunque, si è appena detto che tutti gli attori che entrano in gioco nel processo operatorio devono tenere conto delle condizioni di incertezza e proprio per questo in base al differente ruolo che essi vanno a rivestire, vi sono per l'appunto diverse "classi" di incertezza.

Per quanto riguarda il personale medico si parla pertanto di "*incertezza tecnica*", aggravata dall'insicurezza di tipo "prescrittivo" qualora vi sia mancanza di informazione o di consenso sulle conoscenze professionali più avanzate.

Riguardo al paziente si parla invece di “*incertezza di consumo*”, dato che esso può essere privo delle risorse tecniche per giudicare ciò a cui viene sottoposto.

Infine riguardo invece a chi si occupa delle fasi di pianificazione e di gestione della sala operatoria, si parla invece di “*incertezza decisionale*”, dato che non sono note né l’efficacia del sistema, né il costo di opportunità delle prestazioni e né garanzie sulla qualità.

Una diminuzione del grado di incertezza che caratterizza ciascuno dei suddetti attori si può verificare solo aumentando il grado di conoscenza e di informazione.

La presenza dell’incertezza decisionale spinge perciò ad acquisire una miglior conoscenza dei processi decisionali medesimi, in particolare con l’obiettivo di andare ad identificare le informazioni più rilevanti in base alle specifiche condizioni fisiologiche del paziente.

Scendendo ancora di più nello specifico si può comunque dire che i fattori di incertezza principali a cui si deve fronteggiare nella fase di pianificazione, di gestione e di controllo dei flussi di attività che avvengono nell’ambito di una sala operatoria, riguardano essenzialmente due aspetti:

- incertezza sulla possibilità di arrivo di un caso urgente al sistema sanitario;
- incertezza sulla durata effettiva complessiva di un caso chirurgico.

Focalizzandosi ulteriormente sui due aspetti suddetti, che difatti saranno proprio quelli che saranno considerati nei capitoli successivi di questo lavoro di tesi, si può dire che per quanto riguarda la possibilità di arrivo di un caso urgente al sistema sanitario c’è chiaramente da pianificare il sistema stesso di modo che esso possa essere capace di rispondere tempestivamente e di affrontare l’urgenza stessa in modalità “just in time” (ossia praticamente subito).

In alcuni casi, difatti, il paziente può giungere in condizioni fisiche davvero disperate e non si può attendere assolutamente troppo tempo prima di effettuare su di lui un intervento chirurgico.

La “routine” prevede in questi casi generalmente una chiamata al 118 che la smista immediatamente al pronto soccorso, in cui un addetto del personale paramedico deve raccogliere informazioni circa l’ora dell’accaduto, il sesso e l’età del paziente, la dinamica ed il luogo dell’evento, l’entità della gravità delle sue condizioni fisiche, i primi eventuali provvedimenti adottati e soprattutto i tempi di arrivo al sistema sanitario stimati. In questo caso più che mai i risultati ottenuti sono direttamente proporzionali al livello organizzativo e le capacità professionali rivestono un ruolo di vitale importanza.

In poche parole tutto ciò significa che i processi di pianificazione e di gestione del sistema sanitario devono essere dunque progettati nella maniera più adeguata possibile al fine da poter rispondere prontamente e quanto più possibile efficacemente ed efficientemente a tali urgenze che si possono venire a manifestare all'interno del sistema stesso.

L'altro fattore di incertezza è legato invece alla durata effettiva complessiva di un caso chirurgico: in questo caso bisogna sottolineare innanzitutto che la durata complessiva tiene conto anche dei tempi legati alle fasi a monte ed a valle della fase "operatoria".

Com'è stato detto nel secondo paragrafo di questo capitolo, infatti, il "processo clinico" che avviene nell'ambito di una sala operatoria è composto da tre fasi ovvero quella "pre-operatoria", quella "operatoria" e quella "post-operatoria" ed anche per quanto riguarda la durata complessiva di un intervento chirurgico bisogna tenerne conto.

Per quanto riguarda infatti la fase "pre-operatoria" è richiesto alla sala operatoria un certo margine di tempo per le cosiddette operazioni preliminari e di setup.

Prima di un generico intervento chirurgico infatti bisogna preparare la sala operatoria munendola dapprima di tutto l'occorrente (materiali, strumenti chirurgici che prima dell'uso richiedono di essere sterilizzati secondo determinate procedure, soluzioni, macchinari ed attrezzature specifiche...), così come devono poi prepararsi prima dell'inizio della loro prestazione anche il personale medico e paramedico, rispondendo ovviamente a determinati parametri di igiene e di sterilizzazione che vanno a riguardare anche dettagli che potrebbero sembrare apparentemente di poco conto.

Per esempio, infatti, è bene sapere che rivestono una loro importanza persino le calzature che il personale indossa al momento dell'intervento, perché le soles delle scarpe sono veicolo di numerosi germi.

Infine appartiene a questa fase anche il tempo per la preparazione del paziente, che va posizionato sul letto operatorio in maniera adeguata perché proprio tale aspetto è uno dei presupposti fondamentali per un'agevole esecuzione dell'intervento chirurgico stesso.

Proprio la conoscenza dei tempi e delle peculiarità dei vari interventi permette di posizionare il paziente in modo da esporre adeguatamente la regione del suo corpo che verrà sottoposta alle manovre chirurgiche.

Allo stesso modo per quanto riguarda la fase "post-operatoria" vi è un margine di tempo per le cosiddette operazioni di pulizia post-operatoria, dato che una volta terminato l'intervento chirurgico la sala operatoria va opportunamente pulita e disinfettata così come gli strumenti utilizzati vanno adeguatamente e secondo delle procedure definite sterilizzati.

Ma mentre il fattore di incertezza è piuttosto contenuto se non addirittura trascurabile per quel che riguarda le durate delle operazioni preliminari e di setup (quelle legate alla fase “pre-operatoria”) e per quelle di pulizia dopo l’intervento chirurgico (legate invece alla fase “post-operatoria”), lo stesso non si può dire per quel che riguarda la durata di un intervento chirurgico nella fase esclusivamente operatoria.

Difatti è proprio in questa fase che possono sorgere maggiormente complicazioni, eventi inattesi o rischi: è d’uopo specificare pertanto che vi sono rischi di tipo “generico” (comuni a tutti gli interventi chirurgici e legati alla tipologia di anestesia adottata, all’età del paziente ed alle sue condizioni generali) e rischi di tipo “specifico” (riguardanti per l’appunto il particolare intervento chirurgico in questione e che può comprendere eventi come emorragie, infezioni ad una ferita, ematomi, complicazioni respiratorie, lesioni...).

In generale si può dire che la durata di un intervento chirurgico è variabile e dipende dal tipo di procedura adoperata nonché dalle difficoltà chirurgiche dovute al manifestarsi di qualcuno dei fattori di rischio suddetti.

La durata media dei diversi interventi chirurgici è funzione dunque di due fattori principali ovvero le condizioni cliniche del paziente (per via delle eventuali complicazioni che potrebbero manifestarsi) e l’esperienza e la professionalità del personale che se ne occupa dell’esecuzione.

Tuttavia le varie tecniche chirurgiche ed anestesologiche permettono attualmente di prevedere per la maggior parte degli interventi il valore di questa durata media, tenendo chiaramente conto che quelli di lunga durata sottopongono logicamente il paziente a maggiori rischi.

Vi sono comunque degli interventi particolarmente impegnativi e delicati per cui rimane elevato il fattore di incertezza circa la loro durata media.

Un esempio ne sono i casi di interventi sul cervello, durante i quali bisogna verificare le capacità motorie e cognitive del paziente al fine di accertarsi di non avergli causato lesioni irreversibili.

Tipo di Chirurgia	Intervento	Durata
VASCOLARE E CARDIOCHIRURGIA	safenectomia (asportazioni di varici)	2-2,5 ore
	angioplastica coronarica percutanea	30-60 minuti
	sostituzione valvolare	3-3,5 ore
	chirurgia dell'aorta	2,5-4 ore
	by-pass coronarico	2-3 ore

	chirurgia delle arterie degli arti inferiori	1-2 ore
TORACICA	broncoscopia	15-30 minuti
	riparazione ernia iatale	2 ore
	chirurgia per carcinoma dell'esofago	2-4 ore
	dilatazione endoscopica esofagea	30-40 minuti
	chirurgia per carcinoma polmonare	1,5-2,5 ore
ADDOMINALE	gastrectomia per ulcera o cancro	1-2 ore
	chirurgia della colecisti	1-1,5 ore
	appendicectomia	30-45 minuti
	ernioplastica	30-60 minuti
	emorroidectomia	30-45 minuti
ORTOPEDICA	artroscopia	30-60 minuti
	protesi d'anca o di ginocchio	2-3 ore
	chirurgia della colonna vertebrale	1-3 ore
	chirurgia della mano	30 minuti-2 ore
NEUROCHIRURGIA	asportazione di tumori cerebrali	variabile
	chirurgia vascolare cerebrale	variabile
	chirurgia per traumi cranici	variabile
OFTALMICA	cataratta	20-40 minuti
	distacco della retina	40-60 minuti
	correzione dello strabismo	30 minuti
OTORINO-LARINGOIATRIA	tonsillectomia	20-30 minuti
	adenoidectomia	10 minuti
	laringectomia per neoplasia	2-3 ore
	miringectomia	10-15 minuti
UROLOGICA	resezione prostatica-transuretrale	30-60 minuti
	chirurgia cielo aperto del rene	1-3 ore
GINECOLOGICA	dilatazione e curretage	10-15 minuti

	interruzione di gravidanza	10-15 minuti
	laparoscopia esplorativa	20-30 minuti
	sterilizzazione laparoscopica	20-30 minuti
	isterectomia per via laparotomica	45-90 minuti
	isterectomia per via vaginale	45-90 minuti
	riparazione del piano pelvico	45-90 minuti
OSTETRICA	taglio cesareo	30-60 minuti

Figura 1.4: durate medie di alcuni tra i più comuni interventi chirurgici distinti in base alla specialità chirurgica.

Capitolo 2

Modelli e metodi quantitativi per la pianificazione e gestione delle sale operatorie

2.1 Generalità, livelli e processi decisionali, requisiti e specifiche funzionali

Dopo aver illustrato nel capitolo precedente in maniera esaustiva tutti gli aspetti, i fattori intrinseci e le problematiche legate alla fase di pianificazione, di gestione e di controllo di una sala operatoria, si è a questo punto ben compreso il concetto legato all'importanza di uno svolgimento corretto, adeguato ed ottimale di tale fase.

Tutto questo proprio perché, come detto nello scorso capitolo, una sala operatoria va a rappresentare il punto centrale dell'area critica di un sistema sanitario e non a caso ne è il centro di costo di maggior incidenza.

A conseguenza di tutte le motivazioni suddette diventa d'uopo quindi l'introduzione di soluzioni capaci di progettare al meglio il "processo produttivo" interno ad una sala operatoria.

Questo può avvenire mediante l'utilizzo di strumenti in grado di offrire al decisore che si occupa di tale fase un livello elevato sia di produttività che di efficienza mediante un opportuno trade-off, poiché tali soluzioni sono infatti capaci di ottimizzare l'impiego delle risorse nonché il flusso di attività che avviene all'interno di tale sala operatoria.

Innanzitutto è necessario precisare che tali soluzioni suddette prevedono in genere la definizione a priori di un determinato programma di pianificazione da seguire, che va a stabilire l'ordinamento temporale per l'esecuzione di una serie di interventi chirurgici.

Naturalmente tale programma deve essere modificabile in corso d'opera a causa dell'arrivo dei casi urgenti che, chiaramente, vanno privilegiati in quanto devono godere del diritto di precedenza su quelli elettivi, a meno che non vi sia una sala operatoria dedicata esclusivamente ad esse.

Ciò deve avvenire in un dato orizzonte temporale (generalmente un giorno o una settimana), in maniera da espletare il processo decisionale nell'arco di tempo lungo cui questo si protrae, con l'obiettivo primario dell'ottimizzazione dell'utilizzo delle risorse e del flusso di attività richiesto.

Tutte le caratteristiche suddette a cui questo programma deve rispondere vengono identificate comunemente e racchiuse con l'univoco termine "*scheduling*".

Ma è comunque d'uopo distinguere che, in base alle condizioni che possono essere poste al contorno, esistono tre differenti modalità di scheduling [8]:

1. modalità "open scheduling" (OS);
2. modalità "block scheduling" (BS);
3. modalità "modified block scheduling" (MBS).

Iniziando dalla modalità OS, si può dire che tale strategia consente l'assegnamento dei casi chirurgici ad una qualunque sala operatoria disponibile e, partendo da uno scheduling inizialmente vuoto, si va così a riempire la sala stessa fino a due livelli seguendo l'ordine di arrivo di tali casi (secondo la strategia FIFO, ossia il primo che arriva nel sistema sanitario è il primo ad uscirne con la richiesta soddisfatta).

Il primo livello riguarda la pianificazione dei casi a medio termine, mentre il secondo livello riguarda invece una dettagliata definizione dello scheduling di una sala operatoria a breve termine.

In questa modalità possono essere presentati pertanto casi chirurgici fino al giorno dell'esecuzione dell'intervento chirurgico, dato che l'obiettivo primario è proprio quello di accoglierli e di soddisfarne le richieste di tutti attribuendoli singolarmente alle sale operatorie (prima del giorno previsto per l'esecuzione dell'intervento stesso, chiaramente).

Nella modalità BS, invece, sono assegnati preliminarmente ad uno specifico chirurgo o gruppo di chirurghi (che ne raccoglie l'insieme di tutti i chirurghi appartenenti ad una stessa specialità) un insieme di blocchi temporali (denominati "time-blocks" o più semplicemente TB), che si mantengono generalmente immutati per settimane o addirittura per mesi.

Questa modalità prevede dunque la soluzione di due problemi differenti: il primo riguarda la definizione di un "piano orario capace di definire il numero ed il tipo di sale operatorie rese disponibili in un determinato periodo, il numero di ore in cui esse rimarranno tali, nonché la lista dei chirurghi disponibili durante tale blocco temporale" [9].

Tutto ciò viene identificato con il termine MSS (acronimo di “Master surgical schedule”): quindi, sintetizzando quanto detto finora, il primo problema è dunque quello di costruire un adeguato MSS.

Il secondo problema, invece, è quello di andare a riempire i blocchi temporali suddetti con i casi chirurgici con l’obiettivo che mediamente la loro durata riesca a stare all’interno di questi blocchi stessi, o quanto meno minimizzandone gli overtimes.

Infine, nella modalità MBS, vi è la presenza di blocchi temporali predefiniti ma che in questo caso, per offrire maggior flessibilità al sistema nel processo decisionale, possono essere anche modificati.

In un primo caso solo alcuni tra questi blocchi vengono assegnati, mentre gli altri rimangono vuoti.

In un secondo caso, invece, i blocchi vuoti vengono rilasciati poco prima dell’intervento chirurgico.

Si costruisce anche in questo caso un MSS ma, come detto, alcuni blocchi temporali flessibili non vengono inizialmente assegnati a nessun chirurgo [10].

Tuttavia è comunque d’uopo specificare anche che, in generale, la modalità OS è utilizzata raramente in ambito sanitario, mentre la più comune è senz’altro la modalità BS [11].

Inoltre va puntualizzato che gli MSS riguardano chiaramente la definizione dei soli casi elettivi, ragion per cui la loro efficienza va a diminuire quando si vengono a presentare differenti urgenze.

Proprio per questo motivo scaturiscono due politiche differenti: la prima politica propone di riservare una sala operatoria dedicata alle sole urgenze di modo che esse possano, appena si presentano, essere schedate al più presto (anche subito se tale sala non sta trattando nessun’altra urgenza ed è dunque vuota).

Tale politica minimizza così il tempo di attesa del paziente, ma di contro se si viene a presentare un numero limitato di urgenze in un certo orizzonte temporale più o meno lungo, cosa che può chiaramente succedere benissimo, la sala stessa avrà un tasso di utilizzo basso e ciò ne comporta logicamente un costo annesso di inattività.

La seconda politica, invece, propone in fase di definizione del MSS di riservare parte della capacità della sala operatoria proprio per tali urgenze (questa parte della capacità della sala operatoria viene comunemente chiamata “slack”).

In questo modo appena si verrà a presentare al sistema sanitario un caso urgente, esso potrà essere schedato appena sarà terminato l’intervento chirurgico di un generico

paziente elettivo in tale “slack” e poi si proseguirà con l’esecuzione di quanto pianificato a priori dallo scheduling stesso.

Le due suddette politiche sono state anche testate in una situazione reale in un sistema sanitario olandese di Rotterdam [12] e, misurandone le prestazioni in base ad indicatori quali tempi di attesa, overtimes e tassi di utilizzo delle sale, ne è scaturito che la seconda politica è più conveniente della prima.

A questo punto è d’uopo specificare anche che la letteratura va a distinguere tre differenti livelli decisionali, all’interno dei quali vanno a rientrare tutte le nozioni ed i concetti descritti precedentemente [13]:

1. livello strategico;
2. livello tattico;
3. livello operativo.

Nel livello strategico (noto altresì come “sessione di pianificazione del problema” o SPP), l’obiettivo principale è quello di distribuire il tempo in cui una sala operatoria è utilizzabile ad un chirurgo (o comunque ad un gruppo di chirurghi della stessa specialità), affrontando dunque un problema di “mix della pianificazione”.

Nel livello tattico, invece, una volta assegnati i blocchi temporali di una sala operatoria ai chirurghi si passa alla definizione di un MSS che, come detto in precedenza, va a definire il numero ed il tipo di sale disponibili, l’intervallo in cui ciascuna di esse sarà aperta ed i chirurghi ai quali i blocchi temporali di tale sala sono assegnati.

Nel livello operativo, infine, si va stabilire in genere su base giornaliera la programmazione dei singoli casi elettivi a valle dello sviluppo del MSS, comprendendo aspetti come l’assegnamento di ciascun singolo caso ad una determinata sala operatoria e l’ordine di esecuzione dell’insieme di tali casi elettivi all’interno di una sala operatoria: in sintesi tutto ciò coincide in questa fase con la definizione di uno scheduling opportuno.

C’è naturalmente da sottolineare che i tre livelli decisionali non sono indipendenti l’uno dall’altro, anzi si tratta di un vero e proprio cosiddetto sistema “a cascata” poiché l’output di ciascuna fase diventa l’input per la fase successiva.

Si è inoltre detto che in genere durante l’arco di una giornata una serie di decisioni possono essere ridefinite a causa del verificarsi di un evento non prevedibile a priori.

Tali eventi possono essere:

1. la durata di un caso elettivo che, a causa di complicazioni di qualsiasi natura, si protrae più a lungo del previsto facendo di conseguenza slittare l’inizio dei casi elettivi successivi nell’ambito dello scheduling stesso;

2. la cancellazione di un caso elettivo perché, per esempio, il paziente ha avuto altri problemi che al momento ne rendono impossibile l'esecuzione del suo intervento chirurgico e che rende così necessaria la ridefinizione dello scheduling tenendo conto di tale modifica da apportare;
3. l'arrivo di un caso urgente, che va chiaramente a stravolgere quanto pianificato prima) e proprio per questo motivo uno dei requisiti funzionali principali nella fase di definizione di tale scheduling è la flessibilità, ovvero la capacità di rispondere prontamente al verificarsi di qualcuno degli eventi non prevedibili a priori suddetti.



Figura 2.1: definizione dei livelli decisionali nelle fasi di pianificazione e gestione di una sala operatoria.

2.2 Approcci deterministici, generalità

Dopo aver finora discusso dei caratteri generali legati alle fasi di pianificazione, di gestione e di controllo di una sala operatoria, si può passare ora a vederne più dettagliatamente gli aspetti e gli approcci legati ai modelli ed ai metodi quantitativi presenti in letteratura.

Per prima cosa è d'uopo dire che, poiché una sala operatoria svolge un'importante attività nel settore sanitario, bisogna assicurare che essa svolga prestazioni elevate in termini di efficienza, di qualità e di sicurezza.

Pertanto serve una metodologia di supporto alle decisioni che ne migliori l'organizzazione definendo un'adeguata pianificazione delle attività ed una gestione ottimale dei flussi di lavoro, volta a massimizzarne l'efficienza tenendo chiaramente sotto controllo i costi.

Per tutte le motivazioni suddette, si rende così necessario l'ausilio di uno strumento capace di manifestarsi adeguato nelle attività decisionali, permettendo di operare le decisioni migliori al fine di raggiungere un determinato obiettivo nonostante la presenza di un insieme di vincoli posti da condizioni ben definite e che non sono sotto il controllo del decisore.

In un'ottica relativa ad un sistema di supporto alle decisioni, un modello di ottimizzazione può essere interpretato come una sorta di "cruscotto direzionale" attraverso cui, partendo dalla regolazione di determinati parametri stabilibili dal decisore (ad esempio la lista di interventi chirurgici che bisogna eseguire in una determinata giornata lavorativa), è possibile ottenere delle indicazioni circa le modalità di scelta di gestione della sala operatoria.

I vari modelli di ottimizzazione vengono fuori dopo aver seguito preliminarmente una serie di "step", ovvero:

1. **Identificazione del problema decisionale da affrontare**, in cui si vanno ad identificare i fattori critici per capire l'obiettivo da perseguire;
2. **Definizione dei parametri fondamentali**, in cui si individuano i parametri controllabili dal decisore ed in grado di influenzare il funzionamento del sistema (in questo caso: numero totale di sale operatorie disponibili, monte-ore lavorativo regolare giornaliero...);
3. **Definizione delle variabili decisionali**, in cui individuano invece le scelte effettuabili che in matematica vengono per l'appunto tradotte in variabili (in questo caso: a quale sala operatoria assegnare un determinato caso elettivo o urgente che sia, a quale chirurgo o team chirurgico assegnare un determinato caso...);
4. **Definizione della funzione obiettivo**, ovvero del criterio secondo cui vanno operate le scelte in base ad una certa dimensione di efficienza su cui si va ad incidere (in questo caso: minimizzazione dei costi globali, massimizzazione della produttività, minimizzazione dei tempi d'attesa per un paziente, minimizzazione dei tempi di inattività delle sale operatorie...);
5. **Definizione dei vincoli**, ovvero di tutte quelle condizioni al contorno che possono rendere inammissibile una certa decisione,

limitandone dunque l'insieme complessivo (in questo caso, per esempio, se c'è l'obbligo di schedulare un certo caso urgente entro un numero massimo di ore entro cui esso si viene a presentare, diventa inammissibile una decisione di uno scheduling che vada a stabilire che questo caso urgente venga schedulato oltre questo limite massimo di ore suddetto).

Dopo aver dunque spiegato il ruolo giocato dai sistemi di supporto alle decisioni in quest'ambito e dopo aver illustrato le modalità per lo sviluppo di modelli e metodi quantitativi per il supporto decisionale nelle fasi di pianificazione, di gestione e di controllo delle sale operatorie, bisogna distinguere un'importante classificazione di quanto presente in letteratura per quest'ambito in base a come viene trattata l'incertezza decisionale.

Vi sono infatti presenti in letteratura sia modelli di tipo **deterministico** che modelli di tipo **stocastico**.

Nei modelli deterministici l'incertezza viene praticamente ignorata, tipicamente supponendo per semplicità strutturale che il sistema sanitario in questione non debba trattare casi urgenti o che il sistema stesso “navighi a vista” quando tali casi si vengono a presentare, prendendo dunque delle decisioni “just in time” quando si viene a manifestare un evento di questo tipo).

Nei modelli stocastici invece il grado di incertezza decisionale è trattato opportunamente mediante appositi metodi: di essi, comunque, si parlerà dettagliatamente nel prossimo paragrafo.

Incentrandosi pertanto inizialmente sui modelli di tipo deterministico, si può dire che le loro generalità sono le solite motivazioni legate alla necessità di contenere i costi e tutti i fattori influenti sulla produttività e sull'efficacia del sistema stesso (i tempi di attesa del paziente, i “sotto-utilizzi” ed i tempi di inattività della sala operatoria, i “sovra-utilizzi” ed i tempi di lavoro straordinario, ovvero i cosiddetti “overtimes”, per il personale), in maniera da massimizzare allo stesso tempo anche la soddisfazione del paziente.

Riferendosi ai tre livelli decisionali suddetti, si può dire che la maggior parte di questi modelli si incentrano sul livello operativo poiché il loro obiettivo è quello di costruire uno scheduling, che deve essere seguito all'interno della sala operatoria per pianificare al meglio il flusso delle attività.

Tali modelli deterministici, comunque hanno il grosso limite ovviamente di trascurare l'incertezza e nella fattispecie vanno generalmente, come suddetto, ad ignorare

completamente l'ipotesi che al sistema sanitario giungano casi urgenti, i quali richiedono naturalmente di essere schedulati nell'immediatezza.

Esaminando i modelli deterministici inerenti a quest'ambito applicativo e presenti in letteratura negli ultimi anni in maniera un po' più dettagliata, facendolo in ordine rigorosamente cronologico rispetto al loro anno di pubblicazione, si può partire dal lavoro di Ozkarahan [14], che fu uno dei primi ad occuparsene (il suo lavoro risale infatti al 2000).

Ozkarahan fornì un modello il cui obiettivo è quello di assegnare in ciascuna sala gli interventi chirurgici rispettando tempi e disponibilità considerandone le priorità degli interventi stessi.

Inoltre in questo lavoro viene poi prestata una certa attenzione anche alla fase "post-operatoria", con differenti criteri di ottimalità che possono essere utilizzati sia singolarmente che globalmente (pesati in questo caso da alcuni parametri che tengono conto di quanto essi vadano ad incidere nel complesso).

Proseguendo al 2003 vi è invece il modello Guinet e Chaabane [15], il quale si va a differenziare soprattutto poiché impone la condizione che gli interventi chirurgici con maggior priorità non debbano essere schedulati durante il periodo di overtime.

Tuttavia questo lavoro considera in maniera piuttosto approssimativa che il grado di priorità dell'intervento chirurgico sia direttamente proporzionale alla sua durata effettiva, cosa che chiaramente non è sempre veritiera.

Inoltre, in maniera diversa dal lavoro di Ozkarahan, questo modello incentra la sua attenzione soltanto sulla fase operatoria.

Un lavoro deterministico molto ben strutturato lo si trova invece nel 2006 ed è quello di Jebani, Alouane e Ladet [16], che si basa sul livello operativo ma lo fa considerando anche le fasi a monte ed a valle dell'esecuzione dell'intervento chirurgico sul paziente, mediante l'utilizzo di due modelli.

Il primo è di assegnamento e si occupa per l'appunto di assegnare i casi alle sale operatorie, mentre il secondo è di sequenziamento e si occupa invece di stabilire l'ordine con cui tali casi devono essere schedulati all'interno delle sale cui sono stati assegnati.

Il criterio di ottimalità è la minimizzazione dei costi di ricovero dei pazienti e di quelli di inattività della sala operatoria nel modello di assegnamento, mentre è la minimizzazione dei tempi di overtime nel modello di sequenziamento.

Purtroppo però il grosso limite che ha questo modello, come del resto la maggioranza dei modelli deterministici proposti in letteratura, è quello di non considerare affatto i casi urgenti.

Spostandosi nel 2007 un lavoro degno di nota è certamente quello di Santibanez, Begen ed Atkins [17] che si basa invece per lo più sulla costruzione di una sorta di “macro-scheduling”, in quanto si occupa di assegnare alle sale operatorie le specialità chirurgiche e non semplicemente i casi chirurgici come fanno invece gran parte degli altri modelli deterministici inerenti a quest’ambito.

Si considera inoltre anche la fase “post-operatoria” e lo si fa scegliendo quale criterio di efficienza adoperare tra la massimizzazione della produttività totale e la minimizzazione dei picchi sull’utilizzo delle risorse.

Rimanendo nel 2007 vi è poi da menzionare un modello italiano, proposto da Testi, Tanfani e Torre [18] e basato sulla suddetta suddivisione del processo decisionale in tre livelli.

Tale lavoro propone un approccio deterministico per le prime due fasi (quella strategica, denominata anche SPP come detto nello scorso paragrafo, e quella tattica, in cui come si è detto va definito un MSS). Per la terza fase (operativa, in cui va definito invece lo scheduling delle operazioni, e denominata in questo modello come ECS, acronimo di “*Elective Case Scheduling*”), invece, viene proposto un modello di simulazione.

Il tutto evidenzia comunque la scarsità di lavori in letteratura circa la fase di definizione di un MSS e c’è a tal uopo da dire che questo aspetto è rimasto più o meno uguale tuttora, sebbene a distanza di qualche anno.

Giungendo al 2008 tra i modelli di natura deterministica c’è da citare il lavoro di Arnaout e Kulbashian [19], anch’esso, come molti altri di quest’ambito, incentrato sulla definizione di uno scheduling per la sala operatoria ma con il grosso limite comune a gran parte degli altri modelli deterministici di non considerare i casi urgenti.

Sulla stessa linea, redatto nello stesso anno, vi è anche il modello di Fei, Chu e Meskens [20] che propone la definizione di uno scheduling per la sala operatoria, basandosi dunque anch’esso sul livello decisionale operativo, e con l’obiettivo di minimizzare i costi complessivi, ma con il grosso limite di tralasciare anch’esso i casi urgenti.

Passando al 2009, un lavoro più completo sotto questo punto di vista è senza dubbio quello proposto da Cardoen [21], anch’esso incentrato sul livello decisionale operativo e che propone un modello deterministico adeguato al fine di costruire uno scheduling selezionando tra differenti criteri quello più opportuno al caso del sistema sanitario in questione, grazie alla presenza di una funzione di normalizzazione all’interno della funzione obiettivo del modello stesso.

Proseguendo in questo excursus sui modelli deterministici, c’è poi da annoverare, rimanendo nel 2009, il lavoro di Molina e Framinam [22], il quale prende anch’esso (allo

stesso modo del lavoro di Testi, Tanfani e Torre) come riferimento la suddivisione del processo decisionale in tre livelli, ma che come parecchi altri si incentra poi sul solo livello operativo per la definizione di uno scheduling.

Il suddetto scheduling è definito con l'obiettivo della massimizzazione della qualità che viene calcolata utilizzando i "pesi clinici" dei pazienti, ovvero degli indicatori di importanza legati al tipo di intervento chirurgico che richiedono al sistema sanitario (in base al loro livello di "priorità", dunque).

Redatto anch'esso nel 2009 è poi il lavoro di Pinedo [23], che propone un modello deterministico incentrato però sul solo livello operativo e basato sulle possibili sembianze che assumono, con qualche semplificazione plausibile, il problema di definizione di uno scheduling per una sala operatoria ed il classico problema del "bin-packing".

In tale problema, noto altresì come "problema di caricamento", vi sono un certo numero di oggetti con una loro dimensione (che in questo caso sarebbero gli interventi chirurgici con le loro durate prestabilite) che vanno caricati in un certo numero di contenitori (che in questo caso sarebbero invece le sale operatorie) e lo si fa nell'ottica di minimizzare proprio il numero di contenitori necessari (che in questo caso coinciderebbe con la minimizzazione dei tempi di inattività delle sale operatorie, dato che bisogna "caricarle" di interventi chirurgici quanto più possibile).

Questa panoramica sui modelli deterministici prosegue con il lavoro di Zhang, Murali, Dessouky e Belson [24], anch'esso del 2009 e che rispetto ad altri modelli di questo tipo va a fare un passo in avanti distinguendo i casi chirurgici in ordinari ed emergenti anche se con delle dovute semplificazioni, plausibili per il fatto che chiaramente con un modello deterministico non è possibile trattare adeguatamente il fattore di incertezza in un sistema sanitario.

Per il resto c'è da dire che anche il loro modello, come gran parte di tutti quelli che sono stati citati finora, si propone di fornire uno scheduling degli interventi chirurgici da eseguire all'interno di una sala operatoria in base alla minimizzazione dei costi globali (legati a ritardi o penalità per la domanda insoddisfatta).

Questo excursus lo si può infine concludere giungendo nel 2011 e citando un lavoro di Liu, Chu e Wang [25] che prendendo come riferimento il già menzionato modello di Fei, Chu e Meskens hanno proposto un possibile metodo euristico per la costruzione di uno scheduling per una sala operatoria basato sulla modalità OS. Tale metodo appare essere efficiente, ma il limite complessivo è dovuto al fatto che non sempre è possibile programmare in una sala operatoria uno scheduling seguendo esclusivamente tale modalità ed inoltre il modello di riferimento, come suddetto, non considera i casi urgenti.

Autore/i (anno)	Obiettivi del modello	Peculiarità del modello
Ozkarahan (2000)	Minimizzazione di vari criteri, considerabili sia singolarmente che globalmente.	Modello multisala che considera sia la fase “operatoria” che quella “post-operatoria”
Guinet e Chaabane (2003)	Minimizzazione dei costi totali.	Modello multisala riferito alla sola fase “operatoria” ed impedisce che i casi con maggiore priorità siano trattati durante il periodo di overtime.
Jebani, Alouane e Ladet (2006)	Minimizzazione degli overtime (fase di sequenziamento) e dei costi di ricovero, di undertime e di overtime (fase di assegnamento).	Modello multisala che considera sia la fase “pre-operatoria” che quella “post-operatoria” utilizzando due modelli distinti (che non sono a cascata), ma che non considera i casi urgenti.
Santibanez, Begen ed Atkins (2007)	Scelta di vari criteri da max/min.	Modello multisala che assegna le specialità e non i casi alle sale operatorie e che considera anche la fase “post-operatoria”.
Testi, Tanfani e Torre (2007)	Minimizzazione dei costi e dei tempi di attesa per il paziente.	Lavoro basato sulle tre fasi e che propone un modello deterministico per le prime due fasi (SPP e MSS) e di simulazione per la terza fase (ECS).
Arnaut e Kulbashian (2008)	Riduzione dei costi e miglioramento dell’efficienza.	Ipotizza che in ogni sala sia possibile effettuare qualsiasi tipo di operazione ed ignora i casi urgenti.
Fei, Chu e Meskens	Minimizzazione dei	Modello multisala che

(2008)	costi totali.	considera dato l'insieme degli interventi chirurgici che ogni chirurgo deve trattare, ma che non considera i casi urgenti.
Cardoen (2009)	Riduzione dei costi ed aumento della soddisfazione del paziente.	Utilizza una funzione di normalizzazione al fine di considerare più criteri nella funzione obiettivo.
Molina e Framinam (2009)	Massimizzazione della qualità (calcolata utilizzando i "pesi clinici" dei pazienti).	Lavoro basato sulle tre fasi ma che poi si incentra solo su quella operativa proponendo un modello multisala.
Pinedo (2009)	Minimizzazione dei costi e dei tempi di attesa per il paziente, nonché dei tempi di inattività.	Approccio basato sul modello di "bin packing".
Zhang, Murali, Dessouky e Belson (2009)	Minimizzazione dei costi totali (di ritardo e di penalità per gli interventi chirurgici insoddisfatti).	Modello multisala riferito alla sola fase "operatoria" e che considera anche i casi urgenti, ma senza considerare i costi straordinari.
Liu, Chu, Wang (2011)	Minimizzazione dei costi totali.	Prendendo come riferimento il modello di Fei, Chu e Meskens (vedere pagina precedente), propone un metodo euristico per costruire uno scheduling in modalità OS.

Figura 2.2: principali modelli deterministici proposti per la pianificazione e gestione delle sale operatorie.

Facendo un'analisi complessiva di quanto discusso in questo paragrafo c'è da sottolineare innanzitutto che la letteratura basata sull'approccio deterministico, ovviamente inerente a quest'ambito, ha centrato nel corso degli anni uno dei problemi di fondo legati alla fase di pianificazione e di gestione di una sala operatoria e che, scendendo ulteriormente nel dettaglio, va a collocarsi nel livello operativo del processo decisionale.

Quanto suddetto allude ovviamente alla costruzione di uno scheduling, basato su criteri specifici quali la massimizzazione dell'efficienza in termini di produttività o la minimizzazione di costi globali o straordinari che siano.

Tuttavia l'aspetto deterministico appare inconsistente per quest'ambito applicativo in quanto non permette di trattare in maniera adeguata i fattori di incertezza che sorgono in un sistema del genere.

Gran parte di questi modelli, come si è più volte ripetuto, non considera infatti la possibilità d'arrivo dei casi urgenti e la rimanente parte che prova a farlo, lo fa con semplificazioni che nel caso reale non sono sempre accettabili.

In conclusione dunque si può dire che l'utilizzo dell'approccio deterministico per supportare un processo decisionale così complesso come lo è quello che sorge in quest'ambito, è quasi utopia o che quantomeno lo si può fare con sistemi sanitari di piccole dimensioni (i quali, per esempio, suppongono la possibilità di inviare altrove le eventuali richieste d'urgenze che gli possono pervenire).

Se si vuole invece pensare ad un approccio che sia più consono alle realtà vigenti in un sistema sanitario odierno di medie-grandi dimensioni, l'idea di affrontare i problemi decisionali che sorgono in questa fase di pianificazione e di gestione di una sala operatoria con un approccio prettamente deterministico non è certo quella più indicata.

Proprio per questo motivo, come sarà possibile vedere nel prosieguo di questo capitolo e poi specificamente nei capitoli successivi, questo lavoro di tesi si avvale dell'utilizzo di un approccio decisionale che offre la possibilità di ovviare alle suddette carenze legate all'utilizzo di un approccio deterministico.

2.3 Approcci stocastici

Nel paragrafo precedente si è illustrato ampiamente come gli approcci deterministici non siano totalmente in grado di offrire modelli e metodi quantitativi capaci di trattare al meglio tutti gli aspetti e le peculiarità legate ad un contesto in cui sorgono condizioni di

incertezza come lo è una sala operatoria nella sua fase di pianificazione, di gestione e di controllo.

A questo punto, dunque, in questo paragrafo si può passare finalmente ad incentrarsi maggiormente sugli strumenti che la Ricerca Operativa offre per trattare al meglio le suddette situazioni ed ai quali il lavoro proposto nell'ambito di questa tesi di dottorato fa riferimento, proprio come differenti modelli inerenti anch'essi a quest'ambito applicativo e che sono presenti in letteratura.

La branca della Ricerca Operativa che si interessa di trattare modelli di programmazione matematica in condizioni di incertezza è per l'appunto la *Programmazione Stocastica* (PS).

Prima di proseguire, però, è opportuno riportarne qui di seguito alcuni cenni inerenti alle caratteristiche fondamentali.

Innanzitutto c'è da illustrare la formulazione generale di un modello di PS:

$$\begin{aligned} \mathbf{min} \quad & c^T(\omega)x \\ \text{s.v.} \quad & T(\omega)x = h(\omega) \\ & x \in X \end{aligned}$$

In tale formulazione si ha che ω è un elemento di uno spazio di probabilità e che conseguentemente sia il vettore dei costi c , sia la matrice dei coefficienti tecnologici T , sia il vettore dei termini noti h dipendono dalla sua realizzazione.

Il vettore x rappresenta invece, chiaramente, l'insieme delle variabili decisionali del problema ed X ne rappresenta invece l'insieme di ammissibilità.

Il metodo più conosciuto per adoperare con la PS è il cosiddetto "approccio per scenari": tale approccio assume che vi siano un numero finito di realizzazioni delle variabili aleatorie, ognuna delle quali è per l'appunto denominata "*scenario*".

Una delle proprietà di tale approccio è proprio quella di andare a ricondurre, in termini pratici, un modello stocastico ad un modello deterministico equivalente.

Siccome però in genere un parametro aleatorio è caratterizzato dall'aver un intervallo di valori di realizzazione, c'è comunque da specificare che per trattare al meglio un modello stocastico non è sufficiente risolvere il modello deterministico equivalente limitandosi ad imporre per tali parametri il loro valore medio.

In altre parole non è sufficiente “medializzare i parametri” e risolvere il modello pensando di aver trattato così al meglio il fattore dell’incertezza decisionale.

Non è dunque sufficiente pensare che al fine di trattare nel miglior modo possibile l’incertezza decisionale basta risolvere il seguente modello:

$$\mathbf{min} \quad c^T(\omega)x$$

s.v.

$$\bar{T}x = \bar{h}$$

$$x \in X$$

In questo modello si ha naturalmente che \bar{T} rappresenta il valore atteso della matrice $T(\omega)$, così come \bar{h} rappresenta invece il valore atteso del vettore $h(\omega)$.

In alcuni casi può avvenire persino che la soluzione ottima del modello sovrastante può addirittura non essere ammissibile per il modello stocastico di partenza, cosa che non fa altro che dimostrare ulteriormente la scorrettezza e l’improponibilità di questo approccio.

Alla luce di quanto appena illustrato, serve piuttosto considerare gli scenari collettivamente e proprio per tale ragione assume importanza la creazione di un modello a variabili “esplose” (in cui ogni variabile del modello di partenza viene “esplosa” in tante variabili quante il numero complessivo di scenari del modello stesso).

Dati allora gli aspetti appena menzionati, c’è da dire che si utilizzano in genere due approcci principali nella PS:

1. *modelli a due stadi con ricorso;*
2. *modelli con vincoli probabilistici.*

Iniziando a descrivere i **modelli a due stadi con ricorso**, c’è da dire che il vettore delle variabili decisionali del problema in questione va innanzitutto suddiviso in due parti.

La prima parte, che si può denominare x , include le variabili di primo stadio, ovvero l’insieme di quelle decisioni che possono essere prese prima di conoscere gli eventi aleatori che andranno a realizzarsi (denominate a tal proposito “decisioni anticipative”).

La seconda parte, che si può denominare y , include invece le variabili di secondo stadio, ovvero l’insieme delle decisioni di ricorso che vengono invece prese per “correggere” le decisioni di primo stadio una volta noti gli eventi aleatori (tali decisioni sono a tal proposito invece denominate “decisioni adattative”).

Tale ricorso, ovvero la fase necessaria per “aggiustare” le decisioni di primo stadio una volta note le realizzazioni degli eventi aleatori, ha chiaramente un suo costo annesso nonché una struttura legata alle caratteristiche strutturali della matrice dei suoi coefficienti tecnologici (nota altresì come “matrice di ricorso”).

Nella fattispecie si parlerà pertanto di ricorso “fisso” quando tale matrice sarà composta da elementi costanti e di ricorso “fisso semplice” quando invece tale matrice oltre ad essere composta da elementi costanti sarà anche pari alla ben nota in algebra “matrice I” (la “matrice identità”, ossia quella matrice che ha elementi tutti pari ad 1 sulla diagonale principale e tutti i restanti elementi pari invece a 0).

Una nozione molto importante legata a tale approccio è comunque quella di “non-anticipatività” (o “implementabilità”) che consiste nel restringere l’insieme di decisioni di primo stadio che possono essere prese prima di conoscere la realizzazione degli eventi aleatori.

Vi sono vari metodi per rappresentare tale aspetto e c’è da dire in merito che la scelta del metodo da adottare dipende in genere dalla tipologia risolutiva con la quale si decide di trattare il modello stesso.

La formulazione generale di un modello a due stadi con ricorso è la seguente:

$$\mathbf{min} \quad c^T x + \sum_{l \in N} p_l q_l^T y^l$$

s.v.

$$Ax = b$$

$$T^l x + W^l y^l = h^l, \forall l = 1, \dots, N$$

$$x \geq 0, y^l \geq 0, \forall l = 1, \dots, N$$

In questa formulazione l’insieme N rappresenta l’insieme degli scenari realizzabili, p_l rappresenta la probabilità di realizzazione del generico scenario l e q_l rappresenta invece il valore della funzione di ricorso sotto lo stesso generico scenario l , così come chiaramente y^l rappresenta il vettore delle variabili di secondo stadio sotto lo stesso scenario l .

La matrice A ed il vettore b sono invece rispettivamente la matrice dei coefficienti tecnologici ed il vettore dei termini noti riferiti alle variabili di primo stadio x .

Infine si ha che T^l rappresenta la matrice dei coefficienti tecnologici riferita alle variabili di secondo stadio sotto il generico scenario l , così come h^l rappresenta il vettore dei termini noti riferito alle variabili secondo stadio sotto il generico scenario l ed infine W^l rappresenta l'insieme di valori che assume la matrice di ricorso sotto il medesimo scenario l .



Figura 2.3: schema sintetico delle fasi che compongono un modello a due stadi con ricorso.

Accanto all'approccio a due stadi con ricorso ci sono da introdurre anche due grandezze che misurano il livello di accuratezza dei modelli deterministici equivalenti che si possono introdurre al fine di semplificare l'onerosa fase risolutiva di modelli stocastici siffatti.

La prima grandezza è denominata **EVPI** (acronimo di “*Expected Value of Perfect Information*”, ossia “*Valore atteso della perfetta informazione*”) e misura il profitto che il decisore avrebbe se conoscesse a priori tutte le informazioni future.

Tale EVPI è espresso dalla differenza tra il valore di soluzione ottima del problema di ricorso (naturalmente a variabili “esplose” per ogni scenario, come suddetto) ed il valore medio (pesato dalle probabilità di realizzazione di ogni singolo scenario) delle soluzioni ottime di ogni modello deterministico equivalente considerante singolarmente ogni singolo scenario del problema stesso.

C'è da dire che più piccolo sarà il valore dell'EVPI e meno sarà rilevante il fattore dell'incertezza decisionale nel modello in questione.

La seconda grandezza è invece denominata **VSS** (acronimo di “*Value of Stochastic Solution*”, ossia “*Valore della soluzione stocastica*”) e misura il possibile guadagno derivante dalla risoluzione di un modello stocastico rispetto ad uno deterministico creato imponendo ai parametri aleatori il loro valore atteso.

Tale VSS è per l'appunto espresso dalla differenza tra il valore medio (al solito pesato dalle probabilità di realizzazione di ogni singolo scenario) delle soluzioni ottime di ogni

modello deterministico equivalente considerante le variabili di primo stadio già fissate al valore determinato risolvendo il problema che considera i parametri aleatori pari al loro valor medio, ed il valore di soluzione ottima del problema di ricorso.

C'è comunque da aggiungere, in conclusione, che la politica di ricorso non deve avvenire necessariamente in due stadi ma può avvenire anche in un numero maggiore di stadi (e si parla in questo caso di “*modelli multistadio con ricorso*”).

In questi modelli l'evoluzione temporale delle variabili aleatorie si rappresenta mediante l'utilizzo di un “albero degli scenari”.

Passando quindi a descrivere i **modelli con vincoli probabilistici**, c'è da dire che se negli appena menzionati modelli a due stadi con ricorso la logica è quella di dare un prezzo alla violazione introducendo decisioni compensative nel secondo stadio, in questo caso la logica è invece quella di definire dei problemi in cui i vincoli vengono rispettati con una certa probabilità.

Tale probabilità è determinata dal “livello di affidabilità” del vincolo stesso, identificato in genere con un parametro che assume valori compresi tra 0 ed 1 (se assume valore pari ad 1 è chiaramente comprensibile che questo modello ha le stesse proprietà di un normale modello deterministico).

Tali vincoli probabilistici possono essere sia singoli, ossia se vi è un livello di affidabilità differente per ogni vincolo del modello, o congiunti, se invece il livello di affidabilità è uguale per tutti i vincoli.

La scelta della struttura da adottare dipende chiaramente dalle proprietà del problema in questione da rappresentare mediante un modello stocastico siffatto.

Un'importante osservazione in merito è comunque data dal fatto che è dimostrabile che ogni soluzione che soddisfa un problema a vincoli singoli soddisfa anche lo stesso problema a vincoli congiunti.

Eccone la formulazione generale nel caso di un problema a vincoli congiunti:

$$\min c^T x$$

s.v.

$$P[T(\omega)x \geq h(\omega)] \geq p$$

$$x \in X$$

In questo caso il parametro p rappresenta il livello di affidabilità dell'insieme di vincoli probabilistici.

Nel caso di vincoli separati si ha invece la seguente formulazione generale:

$$\begin{aligned} & \mathbf{min} \quad c^T x \\ & \text{s.v.} \\ & P[T_i(\omega)x \geq h_i(\omega)] \geq p_i \\ & x \in X \end{aligned}$$

In questo caso si ha allora che $T_i(\omega)$ rappresenta l' i -esima riga della matrice $T(\omega)$ e che $h_i(\omega)$ rappresenta l' i -esima componente del vettore $h(\omega)$, mentre p_i rappresenta chiaramente il livello di affidabilità per l' i -esimo vincolo del problema.

Dopo aver definito in maniera piuttosto dettagliata gli aspetti fondamentali inerenti alla PS, si può ora passare, come fatto nello scorso paragrafo per quanto riguarda i modelli deterministici, a presentare gli aspetti fondamentali e le peculiarità dei modelli stocastici presenti in letteratura e naturalmente inerenti all'ambito applicativo della pianificazione, della gestione e del controllo delle sale operatorie seguendo un ordine cronologico rispetto alle loro date di pubblicazione.

Il primo modello stocastico degno di nota è quello di Gupta e Denton [26], i quali nel 2003 proposero un lavoro che ha come obiettivo la minimizzazione dei costi globali con un approccio basato sul suddetto "ricorso a due stadi".

Tale modello viene poi trattato mediante un metodo denominato LSB (che è una variante del tradizionale L-Shaped che viene utilizzato in genere per risolvere modelli stocastici a due stadi con ricorso), risolvendo così un problema che considera come riferimento le fasce orarie nel corso della giornata lavorativa e che inoltre presuppone sempre puntuale l'arrivo dei pazienti presso la sala operatoria dove devono essere schedulati.

Nel 2005 invece fu pubblicato un lavoro di Belien e Demeulemeester [27], che ha come obiettivo la minimizzazione della scarsità dei posti letto totali e che propone un modello stocastico senza ricorso e che viene successivamente trattato mediante una linearizzazione del problema stesso: la peculiarità del loro modello sta comunque nel fatto che ci si incentra quasi esclusivamente nella fase "post-operatoria".

L'anno dopo invece furono Lamiri, Xie, Dolgui e Grimaud [28] a pubblicare un modello stocastico senza ricorso e trattato poi mediante l'utilizzo della simulazione Montecarlo, con l'obiettivo di minimizzare i costi e le ore di lavoro straordinario considerando contestualmente sia casi elettivi che casi urgenti, ma con il limite di non introdurre vincoli di capacità sul lavoro straordinario.

Sempre nel 2006, invece, gli olandesi Hans, Wullink, Van Houdenhoven e Kazemier [29] proposero un modello stocastico senza ricorso trattato poi con una tecnica di linearizzazione e che ha come obiettivo la massimizzazione dell'utilizzo delle risorse nonché la minimizzazione degli overtimes, ma che ha comunque il grosso limite di non considerare i casi urgenti.

Passando al 2007, un modello stocastico degno di nota è quello di Denton, Viapiano e Vogl [30]: si tratta di un modello con ricorso a due stadi trattato mediante l'utilizzo di tre tecniche euristiche differenti e che ha come obiettivo la minimizzazione dei tempi d'attesa del paziente nonché quella dei tempi di inattività delle sale operatorie, ma anche in questo caso c'è il grosso limite del fatto che i casi urgenti non sono considerati.

Due anni dopo, lo stesso Denton unitamente a Miller, Balasubramanian ed Huschka [31] propose un nuovo modello stocastico di ricorso a due stadi con la minimizzazione stavolta dei costi e delle ore di lavoro straordinario e che viene trattato mediante una formulazione "robusta".

Proprio tale approccio ne rappresenta comunque un limite dato che per risolvere il modello vengono richieste anche informazioni del suo duale, che andrebbe dunque risolto preliminarmente.

Nel 2009, inoltre, anche Mancilla e Storer [32] proposero un modello stocastico con ricorso a due stadi trattandolo poi mediante una tecnica euristica basata sulla "decomposizione di Benders".

Tale modello ha come obiettivo la minimizzazione dei costi di lavoro straordinario, nonché quella dei tempi di inattività delle sale operatorie.

Questo lavoro presenta comunque grossi limiti perché oltre ad essere un modello "monosala", va anche ad ignorare la presenza dei casi urgenti.

Sempre nel 2009 fu pubblicato anche un modello stocastico di Lamiri, Grimaud e Xie [33], senza ricorso e che viene trattato poi mediante il metodo della simulazione Montecarlo con l'apporto di varie tecniche metaeuristiche.

Tale modello ha come obiettivo la minimizzazione del valore atteso dei costi globali ma un limite è dovuto al fatto che i casi urgenti vengono sì considerati, ma si suppone poi di "navigare a vista" al fine di trattarli.

Giungendo quindi al 2011, l'ultimo modello stocastico degno di nota in questo excursus è quello di Batun, Denton, Huschka e Schaefer [34]: si tratta di un modello con ricorso a due stadi trattato mediante il tradizionale algoritmo L-Shaped, che ha come obiettivo la minimizzazione dei costi di apertura di una sala operatoria e dei costi legati sia agli overtimes che ai tempi di inattività di una sala operatoria. Il loro modello considera dei vincoli di precedenza tra casi chirurgici (supponendo un parametro che indichi se un certo caso deve essere schedato prima di un altro), ma non va distinguere in maniera definita la differenza tra casi elettivi e casi urgenti.

Andando quindi a fare una sorta di riepilogo generale:

Autore/i (anno)	Obiettivo del modello	Tipologia del modello	Tipologia di risoluzione	Peculiarità del modello
Gupta-Denton (2003)	Minimizzare i costi globali.	Ricorso a 2 stadi.	Metodo LSB, variante del tradizionale L-Shaped.	Considera le fasce orarie supponendo sempre puntuali gli arrivi dei pazienti che richiedono l'intervento chirurgico.
Belien-Demeulemeester (2005)	Minimizzare la scarsità dei posti letto totali.	Senza ricorso.	Linearizzazione del problema.	Si incentra di più sulla fase "post-operatoria".
Lamiri-Xie-Dolgui-Grimaud (2006)	Minimizzare i costi e le ore di lavoro straordinario.	Senza ricorso.	Simulazione Montecarlo.	I vincoli di capacità sul lavoro straordinario sono assenti.
Hans-Wullink-Van Houdenhoven-Kazemier (2006)	Massimizzare l'utilizzo delle risorse e minimizzare gli overtimes.	Senza ricorso.	Linearizzazione del problema	I casi urgenti non sono considerati.
Denton-Viapiano-Vogl (2007)	Minimizzare i tempi di attesa del paziente ed i tempi di inattività.	Ricorso a 2 stadi.	Tre differenti euristiche.	I casi urgenti non sono considerati.
Denton-Miller (2009)	Minimizzare i costi ed il numero di ore di lavoro straordinario.	Ricorso a 2 stadi.	Il problema viene formulato con un approccio "robusto".	L'approccio "robusto" non è granché efficace: per risolvere il modello servono le informazioni

				del suo duale.
Mancilla-Storer (2009)	Minimizzare i costi di lavoro straordinario ed i tempi di inattività.	Ricorso a 2 stadi.	Euristica basata sulla “decomposizione di Benders”.	Modello monosala e che non considera neppure i casi urgenti.
Lamiri-Grimaud-Xie (2009)	Minimizzare il valore atteso dei costi globali.	Senza ricorso.	Simulazione Montecarlo con il supporto di varie tecniche per la fase “meta-euristica”.	I casi urgenti sono considerati, ma si suppone di “navigare a vista” per trattarli.
Batun Denton Huschka Schaefer (2011)	Minimizzare i costi di apertura della sala e quelli legati agli overtimes ed ai tempi di inattività.	Ricorso a 2 stadi.	Metodo tradizionale L-Shaped, basato anche su un algoritmo di Branch & Cut.	Tiene conto dei vincoli di precedenza, ma non distingue i casi in elettivi ed urgenti.

Figura 2.4: principali modelli stocastici proposti per la pianificazione e gestione delle sale operatorie..

Andando quindi ed effettuare un’analisi critica complessiva circa l’utilizzo della PS come strumento in grado di definire modelli e metodi quantitativi circa la fase di pianificazione, di gestione e di controllo di una sala operatoria, si può innanzitutto dire che questo campo fino a pochi anni fa era praticamente inesplorato in quest’ambito.

La letteratura nel corso degli ultimi anni, comunque, si è incentrata maggiormente sul livello operativo dato che la maggior parte di questi modelli (come d’altronde anche quelli deterministici, com’è stato illustrato nello scorso paragrafo) si propone di fornire lo scheduling ottimale degli interventi chirurgici da eseguire nell’ambito dell’intero blocco operatorio.

Alcuni di questi modelli lo fanno comunque con il limite di tralasciare i casi urgenti o di presupporre di dover “navigare a vista” per poterli schedulare a differenza di altri modelli che invece considerano adeguatamente le urgenze stesse.

Per quanto riguarda invece gli obiettivi di tali modelli si può dire che il luogo comune è naturalmente quello di minimizzare i costi (globali o straordinari che siano), i tempi (di inattività o di overtime in genere) e solo in qualche caso quello di massimizzare la produttività o l’utilizzo delle risorse.

Per quanto riguarda infine la tipologia del modello, si può notare invece che il tipo di approccio maggiormente utilizzato è il “ricorso a due stadi”, che viene poi trattato nella fase risolutiva mediante svariati metodi.

Tali metodi prevedono approcci di simulazione, tecniche di rilassamento e/o di linearizzazione del problema di partenza, oppure l'utilizzo dei tradizionali metodi risolutivi con qualche modifica o comunque corredati da qualche particolare tecnica euristica (legata alle caratteristiche intrinseche del problema stesso).

Al fine di cercare di superare dunque i limiti di alcuni di questi modelli, serve un modello che consideri al meglio l'incertezza decisionale che sorge in questo ambito senza la presenza di assunzioni o di semplificazioni che si discostino troppo dal caso reale.

Proprio per questo motivo un modello stocastico inerente alla fase di pianificazione, di gestione e di controllo di una sala operatoria non può trascurare innanzitutto la presenza dei casi urgenti.

Inoltre, al fine di schedulare al meglio i vari casi (elettivi od urgenti che siano), esso deve avere una vasta gamma di possibili azioni “correttive” che mantengano sempre il sistema in grado di avere il giusto trade-off tra efficienza ed efficacia, ovvero minimizzandone costi e tempi globali ma anche garantendo la qualità e la soddisfazione del paziente.

Capitolo 3

Modelli di Programmazione Stocastica a due stadi con ricorso per la gestione operativa delle sale operatorie

3.1 Architettura dei modelli

Riepilogando in maniera sintetica, nel primo capitolo di questa tesi di dottorato si è discusso in maniera approfondita delle caratteristiche intrinseche e dei vari fattori che entrano in gioco nell'ambito della definizione di una fase di pianificazione e di gestione di una sala operatoria: ne sono stati definiti così all'interno di tale capitolo i vari concetti di base, ne sono state approfondite le diverse tematiche legate agli aspetti tecnici ed organizzativi, prestando particolare attenzione alle esigenze ed ai fattori di incertezza che condizionano in maniera significativa le decisioni nell'ambito di tale contesto, che si dimostra essere dunque di basilare importanza nonché di una certa complessità.

Proprio il suddetto fattore dell'incertezza è stato una sorta di “filo conduttore” per quanto si è poi discusso nel secondo capitolo, nel quale invece sono stati definiti i modelli ed i metodi quantitativi che lo stato dell'arte ha offerto nel corso dell'ultimo decennio, illustrando pertanto in termini generali il ruolo che la **Ricerca Operativa** può andare a rivestire in questo contesto decisionale.

Si sono così distinte, proprio in base alle modalità con cui viene trattato il rilevante fattore dell'incertezza, due differenti tassonomie di quanto presentato dallo stato dell'arte stesso: la prima riguarda i “*modelli deterministici*”, i quali, come ha illustrato un'analisi critica posta come appendice a tale tassonomia, si rivelano essere piuttosto inadeguati per trattare al meglio i vari aspetti legati alla fase di pianificazione e di gestione di una sala operatoria in quanto tali modelli infatti trattano in maniera troppo marginale il fattore dell'incertezza e difatti nella maggior parte dei casi, come si era detto, trascurano la

presenza degli interventi chirurgici urgenti o comunque nella loro struttura vanno ad includere approssimazioni e semplificazioni che non appaiono essere sempre plausibili ed accettabili in linea alla situazione reale.

Per tutti i suddetti motivi, allora, nel corso del secondo capitolo si è dapprima giustificato e poi si è presentato il ruolo della **Programmazione Stocastica**, ovvero come detto di quella branca della Ricerca Operativa che va a trattare specificamente i problemi decisionali in cui sorge l'incertezza.

Dopo averne delineato gli aspetti fondamentali ed alcuni semplici richiami, a questo punto nello stesso capitolo si è introdotta la seconda tassonomia di quanto offre lo stato dell'arte, ovvero i "*modelli stocastici*" corredando alla fine il tutto con un'analisi critica che va a discuterne delle peculiarità generali.

Alla luce di quanto detto sinora è tempo quindi di entrare nello specifico e di presentare quindi l'oggetto principale di questo lavoro di tesi di dottorato, ovvero un **modello di Programmazione Intera Mista** (o MIP, "Mixed Integer Programming") **stocastico a due stadi con ricorso**, che ha come obiettivo l'ottimizzazione della fase di gestione operativa di un blocco operatorio e che chiaramente si discosti da quanto lo stato dell'arte ha già presentato nell'ambito della Programmazione Stocastica.

Nel corso del primo paragrafo del secondo capitolo l'intero processo di pianificazione e di gestione era stato scisso in tre livelli decisionali conseguenti, ovvero quello strategico, quello tattico e quello operativo.

Si era poi discusso di quanto avviene nello specifico all'interno di ciascuno dei livelli decisionali suddetti, mentre poi nel proseguo dello stesso capitolo è stato illustrato come lo stato dell'arte (sia per quanto riguarda l'approccio deterministico che quello stocastico) si va ad incentrare soprattutto nel livello operativo, in cui l'obiettivo è quello di andare a fornire un adeguato scheduling delle operazioni da eseguire nell'ambito di una sala operatoria, andandone così a pianificare al suo interno la giornata lavorativa in maniera da contenere quanto più possibile i costi globali ed i tempi di inattività delle sale operatorie nonché quelli di lavoro straordinario.

Nel corso del medesimo primo paragrafo del secondo capitolo si era altresì detto che per la definizione di uno scheduling vi sono delle differenti modalità e che comunque quella che corrisponde maggiormente alla situazione reale è la cosiddetta BS (modalità Block Scheduling, in cui come si era detto si va preliminarmente ad assegnare un insieme di blocchi temporali, che rimangono invariati a volte per settimane o addirittura per mesi, ad uno specifico chirurgo, gruppo chirurgico o ad una specifica specialità chirurgica e quindi si vanno poi a "riempire" tali blocchi all'interno delle differenti sale operatorie

assegnandovi i casi chirurgici che devono essere schedulati nell'arco della giornata lavorativa in maniera da rispettare i suddetti criteri di efficienza.

In linea a tutto ciò, pertanto, il modello stocastico a due stadi con ricorso proposto in questo capitolo è per l'appunto incentrato sul livello decisionale **operativo** e lo fa seguendo una struttura basata sulla suddetta modalità **BS**.

Entrando quindi gradualmente nello specifico, c'è innanzitutto da dire che tale modello è su base **giornaliera** ovvero si propone di fornire quotidianamente lo scheduling delle operazioni che devono essere eseguite all'interno del blocco operatorio, dove per blocco operatorio si intende l'insieme complessivo delle sale operatorie che compongono il sistema sanitario in questione.

Questo ha anticipato dunque il fatto che il modello presentato è di tipo "*multisala*" e che considera come dati iniziali oltre all'insieme di sale operatorie di cui è munito il medesimo sistema sanitario, anche l'insieme di casi elettivi in attesa di assegnamento, l'insieme di casi urgenti che potrebbero venire a presentarsi nell'arco della giornata lavorativa, l'insieme delle specialità chirurgiche che il sistema stesso si propone di trattare (chiaramente ogni intervento chirurgico elettivo od urgente che sia richiede una sola di queste specialità) e l'insieme dei possibili scenari con annessa probabilità di realizzazione, al fine di trattare l'incertezza decisionale legata sia alle durate effettive di un intervento chirurgico (ogni durata è comprensiva del tempo di setup "pre-operatorio" per la sala operatoria, del tempo di esecuzione dell'intervento stesso all'interno della sala operatoria e del tempo di pulizia "post-operatoria" della sala operatoria stessa) sia alla possibilità che un caso urgente possa venire a presentarsi o meno: sotto ogni scenario di riferimento un caso elettivo avrà una durata complessiva differente, mentre per ogni caso urgente oltre ad esservi una durata complessiva differente vi può essere anche la possibilità che esso non si presenti al sistema sanitario (in termini sintetici, se un caso urgente non si viene a presentare nella giornata lavorativa al sistema sanitario sotto un certo scenario di riferimento, vorrà dire che per quello scenario stesso vi sarà una durata complessiva nulla dell'intervento chirurgico stesso).

Di ogni intervento chirurgico elettivo in attesa di assegnamento è poi noto il cosiddetto "*livello di priorità*", ovvero il suo grado di complessità dell'intervento chirurgico il quale, quantificato numericamente, può essere anche inteso come il costo di ricovero del paziente stesso.

Entrando nello specifico c'è da dire che nell'ambito di questo modello sono stati distinti cinque differenti livelli (classi) di priorità e che ad ognuno di essi è assegnato un certo codice identificato da un colore.

Procedendo in ordine crescente di priorità del caso elettivo c'è da dire allora che il livello più basso è identificato dal “*codice bianco*”, il quale è assegnato a pazienti affetti da patologie assolutamente non gravi e di cui in genere le prestazioni, che hanno consuetamente una durata esecutiva piuttosto contenuta e che in termini pratici si aggira in un intervallo tra i 20 ed i 45 minuti, possono essere soggette al pagamento di un ticket sanitario salvo casi di esenzione inerenti alle normative regionali vigenti: un esempio di caso elettivo al quale si può assegnare tale codice può essere riferito alla specialità chirurgica oftalmica, come la correzione di uno strabismo o la rimozione di una cataratta, interventi che con le tecnologie odierne vengono svolti generalmente in un tempo massimo di una quarantina di minuti alla fine di ovviare a tali patologie che non sono di certo gravi a tal punto da compromettere le funzioni vitali del paziente, cosicché un tempo di attesa prolungato non comporterebbe alcuna conseguenza particolare.

Proseguendo vi è poi il “*codice azzurro*”, il quale è assegnato invece a pazienti con patologie non gravi e dalla medio-bassa priorità in quanto il rischio di complicazioni è sì maggiore rispetto agli interventi elettivi con codice bianco, ma è comunque piuttosto contenuto anche in questo caso così come la durata complessiva d'esecuzione che in genere si aggira attorno nell'intervallo dei 45-75 minuti: un esempio di intervento elettivo alla quale si potrebbe assegnare tale caso può riferirsi alla specialità addominale e riguardare un'appendicectomia, al fine di ovviare ad una patologia certamente non grave e che ha una durata esecutiva della prestazione chirurgica piuttosto contenuta, cosicché un tempo di attesa prolungato per il paziente neanche in questo caso andrebbe a comportare conseguenze particolari.

Aumentando il livello di priorità vi è quindi il “*codice verde*”, inerente a patologie non molto gravi ma che potrebbero avere delle complicazioni e che per i quali il tempo di attesa potrebbe essere anche prolungato a secondo delle condizioni fisiche del paziente stesso, mentre la durata complessiva d'esecuzione di un intervento di questo codice si aggira in un intervallo di 60-90 minuti: un esempio di intervento elettivo alla quale si può assegnare tale codice può essere riferita alla specialità chirurgica ortopedica e riguardare l'introduzione di una protesi ad un ginocchio.

Proseguendo vi è poi il “*codice giallo*”, inerente a patologie gravi di medio-alta priorità e per le quali il tempo d'attesa deve necessariamente rimanere contenuto perché il paziente può essere soggetto a complicazioni che ne potrebbero riguardare le funzioni vitali e con una durata complessiva che può variare in un notevole intervallo di tempo: un esempio inerente a tale codice può riferirsi alla specialità chirurgica toracica e riguardare un intervento dovuto ad un carcinoma dell'esofago.

Infine il livello di priorità più alto è rappresentato dal “*codice rosso*” e riguarda tutti quegli interventi chirurgici per i quali il paziente è affetto da una patologia di una certa delicatezza e che potrebbe comprometterne le funzioni vitali: la durata esecutiva di tali interventi elettivi può essere anche significativamente elevata (e superare ad esempio le 4-5 ore) ed i tempi di attesa del paziente devono essere minori quanto più possibile proprio a causa delle complicazioni alle quali il paziente può essere soggetto: un esempio per il quale può essere assegnato tale codice è riferito alla specialità chirurgica vascolare-cardiochirurgica e può riguardare un intervento di chirurgia dell’aorta.

Dopo aver definito le tipologie di casi elettivi che il modello si propone di andare a schedulare, si può passare ora a definire la struttura dei blocchi temporali: si assume che la giornata lavorativa regolare duri 12 ore, durante le quali una sala operatoria generica può schedulare al suo interno un certo numero di casi elettivi od urgenti che siano, e che tale monte-ore sia partizionato in base al numero di specialità chirurgiche che il sistema sanitario offre, creando così blocchi temporali all’interno di ciascuna sala di uguale dimensione (ad esempio in una giornata generica una sala operatoria di un sistema sanitario che offre 3 differenti specialità chirurgiche e che come detto è aperta per un tempo regolare giornaliero di 12 ore, avrà tanti blocchi temporali quanto il numero di specialità chirurgiche offerte dal sistema ovvero tre blocchi dalla durata di 4 ore ciascuno: in sintesi tale sala operatoria dedica in tale giornata lavorativa 4 ore regolari a ciascuna delle specialità ed in questi blocchi lo scheduling deve proporsi di andare ad inserire i casi elettivi ed urgenti richiedenti una certa specialità chirurgica e che devono essere per l’appunto trattati nella sala operatoria stessa).

Ovviamente ogni intervento chirurgico non gode, come si era detto nel primo capitolo, della proprietà di “preemption” e questo vuole chiaramente significare che in una giornata lavorativa una volta iniziato tale intervento non può essere interrotto e poi completato in una giornata lavorativa seguente a quella d’inizio: chiaramente potrebbero presentarsi dei casi in cui, anche a causa di complicazioni di qualsivoglia genere, un intervento può protrarsi per una durata che va oltre quella dei suddetti blocchi temporali previsti e per capirne di più su come comportarsi in tali casi si può fare un esempio. Si supponga allora un sistema sanitario di riferimento che offre 3 differenti specialità chirurgiche, tra cui quella di tipo vascolare-cardiochirurgica e che vi sia da schedulare in una giornata lavorativa un intervento di chirurgia all’aorta (che, come suddetto, avrà sicuramente assegnato un “*codice rosso*”), richiedente per l’appunto tale specialità, e che viene assegnato in una sala operatoria che ha come blocco temporale predefinito e dedicato a tale stessa specialità un periodo di 4 ore lavorative regolari (supponendo, sempre a titolo

esemplificativo, che nel corso della stessa giornata lavorativa all'interno di questa sala non sia stato schedulato nessun altro intervento chirurgico richiedente la stessa specialità chirurgica vascolare-cardiochirurgica). Se per ipotesi a causa di complicazioni che potrebbero insorgere nel corso dello svolgimento del suddetto intervento la durata complessiva va a superare le 4 ore suddette, raggiungendo per esempio una durata complessiva di 4 ore e 30 minuti, di sicuro l'intervento stesso andrà completato anche una volta scaduto tale margine e, pertanto, sarà richiesto un **overtime** di 30 minuti. Tale overtime racchiude dunque la durata del tempo lavorativo straordinario (in quanto va oltre al tempo lavorativo regolare) e deve essere chiaramente retribuito al personale medico e paramedico sostenendo un costo straordinario: il modello considera dunque come dato anche il **costo orario** di tale overtime.

Dopo aver definito la struttura dei dati richieste dal modello, si può passare ora a definire le decisioni che il modello stesso si occupa di stabilire: trattandosi come detto di un modello stocastico a due stadi con ricorso, avremo che nel primo stadio lo scheduling si occuperà di fissare i casi elettivi assegnandoli ad una sala operatoria che deve occuparsene dell'intervento chirurgico o decidendo comunque di non schedularli nella giornata lavorativa corrente. Nel secondo stadio, invece, quando sarà noto il manifestarsi degli eventi su cui vi era incertezza iniziale, lo scheduling stabilirà anche in quale sala operatoria deve essere schedulato ciascun caso urgente che sotto un determinato scenario si è venuto a presentare, nonché poi come altro insieme di variabili anche l'ammontare degli overtimes relativi ad ogni specialità chirurgica in ogni sala operatoria sotto ogni determinato scenario.

Ma il modello stesso include anche altre due possibilità di ricorso oltre all'overtime e che, come sarà illustrato dettagliatamente nel successivo paragrafo di questo capitolo, vengono però considerate singolarmente: la prima possibilità è quella del cosiddetto "**swapping**", ovvero nel secondo stadio sotto un determinato scenario si va a spostare di sala operatoria un caso elettivo che la soluzione di primo stadio aveva assegnato ad un'altra sala operatoria (sostenendo un annesso costo di swapping) mentre la seconda possibilità è quella del "**posticipo**", ovvero nel secondo stadio sotto un determinato scenario si va a posticipare un caso elettivo che la soluzione di primo stadio aveva già assegnato ad una sala operatoria (e questo lo si fa naturalmente sostenendo l'annesso costo di posticipo, dovuto al fatto che il paziente elettivo che doveva essere operato nel corso di questa giornata lavorativa deve essere tenuto ancora in lista di attesa).

Dopo aver presentato le caratteristiche generali del modello di ricorso a due stadi, si può ora passare a discuterne le condizioni al contorno ovvero le assunzioni che vengono fatte,

presentando anche un'analisi critica del modello stesso: si è detto che il modello in questione è di tipo “multisala giornaliero” e che si incentra esclusivamente sul livello “operatorio” considerando i casi urgenti ed i casi elettivi, con questi ultimi distinti in base al codice di priorità annesso.

Una prima condizione al contorno che viene posta è quella di supporre nullo o trascurabile il costo di apertura di una sala operatoria, supponendo così che tutte le sale operatorie del sistema sanitario di riferimento siano tutte disponibili nell'arco della giornata lavorativa e che nessuna di loro sia dedicata ai soli casi urgenti ma che tutte siano in grado di trattare sia interventi elettivi che urgenti.

Una seconda condizione al contorno è invece sempre inerente alle caratteristiche assunte per tali sale operatorie: si assume infatti che esse siano multifunzionali, ovvero in grado di trattare qualunque intervento richiedente una delle specialità chirurgiche che il sistema sanitario di riferimento offre essendone equipaggiate adeguatamente: questo potrebbe essere un limite del modello stesso, in quanto tale assunzione non è sempre data per scontata (vi possono essere infatti in un sistema sanitario alcune sale operatorie non equipaggiate per trattare interventi chirurgici di un certo tipo).

Una terza condizione al contorno è invece riferita ai teams chirurgici (dove per team si intende un insieme composto generalmente, come si era detto anche nel corso del primo capitolo, da 5 figure professionali ovvero un medico senior specialista che assumere il ruolo di “capo-team”, un medico junior specialista, un anestesista e due infermieri), i quali all'interno di questo modello non vengono considerati e tale assunzione può comunque essere giocata in maniere diverse. Si può infatti assumere che per ogni intervento sia deciso a priori il team specifico che deve occuparsi dell'esecuzione della prestazione chirurgica (magari anche per scelta del paziente stesso, al quale per aumentarne la soddisfazione si può dare a priori la possibilità di scegliere il team che lo dovrà operare in base per esempio alla fiducia che nutre nei confronti del “capo-team” o di qualche altro membro stesso che potrebbe per ipotesi averlo assistito nelle fasi preliminari all'operazione, o comunque per scelta a priori di chi si occupa della fase di gestione operativa del blocco operatorio) oppure, in maniera più semplificata, che vi siano dei team multifunzionali in grado dunque di trattare casi chirurgici, elettivi o urgenti che siano, richiedenti una qualsiasi specialità chirurgica tra quelle offerte dal sistema sanitario e supponendo che magari ad ogni sala operatoria del sistema stesso sia stato assegnato a priori uno di questi suddetti team.

Una quarta condizione al contorno è invece legata al costo dell'overtime che è supposto lineare, ossia sempre uguale ora dopo ora, mentre invece vi sono dei sistemi sanitari che

retribuiscono al personale medico e paramedico il periodo di overtime con delle particolari tariffe (per esempio se un'ora di overtime è retribuita ad una generica figura professionale con 50 € in più, due ore consecutive di overtime anziché essere retribuite con 100 € supponendo dunque un costo orario lineare di 50 €/h per l'overtime potrebbero essere ad esempio retribuite con 120 €, con 20 € ulteriori come “premio” per lo straordinario e, sempre per esempio, allo stesso modo tre ore consecutive di overtime potrebbero essere retribuite anziché che con 150 € complessive con 180 € aumentando in questo caso a 30 € il “premio” per lo straordinario suddetto e così via).

Una quinta condizione al contorno, invece, è dovuta al fatto di supportare sempre puntuale l'arrivo dei pazienti e di non considerare per quanto riguarda i casi elettivi il fatto che durante la giornata un caso del genere potrebbe essere cancellato (ad esempio perché il paziente ha subito improvvisamente delle complicazioni che ne rendono impossibile al momento l'esecuzione del suo intervento chirurgico oppure a causa dell'indisponibilità del team chirurgico che se ne doveva occupare e che non può essere rimpiazzato nell'immediatezza): in queste situazioni lo scheduling andrebbe ridefinito da capo tenendo quindi conto che il caso elettivo inerente al paziente che ha avuto delle complicazioni che al momento ne rendono impossibile la schedulazione e va dunque cancellato dalla lista d'attesa giornaliera dei casi elettivi, il che potrebbe quindi rappresentare un altro possibile limite del modello stesso anche se di certo non si tratta di un qualcosa di insormontabile.

Un'ultima condizione al contorno è invece legata ai casi urgenti: si impone infatti che il sistema sanitario debba obbligatoriamente trattarli tutti appena questi si vengono a presentare e non è dunque possibile sostenere nessun costo di penalità per smistarli ad altri sistemi sanitari, come si potrebbe anche assumere di fare.

In conclusione di questo paragrafo si possono definire gli obiettivi del modello: si è già detto che la sua finalità è quella di andare a costruire uno scheduling dei casi all'interno delle sale operatorie, ma il criterio con cui si va a farlo tiene conto della minimizzazione dei costi di non-scheduling dei casi elettivi (che coincide in sintesi con la minimizzazione dei tempi di inattività delle sale operatorie, poiché è chiaro che la minimizzazione dei costi di non-scheduling va ad implicare la massimizzazione del numero di casi elettivi trattati nell'arco della giornata dal sistema sanitario e quindi conseguentemente, poiché per fare ciò bisogna rendere le sale operatorie il più operative possibile, con la minimizzazione dei tempi di inattività) e della minimizzazione dei costi dovuti a decisioni di ricorso quali gli overtimes, le operazioni di swapping o le operazioni di posticipo: il tutto, in maniera sintetica, corrisponde dunque con la *minimizzazione dei costi*

straordinari (infatti le decisioni di non-scheduling di qualche caso elettivo, le decisioni di overtimes, le decisioni di operazioni di swapping o quelle di operazioni di posticipo possono essere tutte viste come un unico insieme di decisioni di tipo straordinario a ciascuna delle quali è chiaramente imputabile un costo annesso: l'obiettivo è minimizzare la somma di tutti questi costi, definendo pertanto uno scheduling capace di ricorrere nella minor misura possibile a queste suddette alternative).

3.2 Sviluppo dei modelli

Dopo aver descritto in maniera sommaria gli aspetti trattati dal modello stocastico oggetto di questo lavoro di tesi di dottorato, si può passare ora a descriverne in maniera dettagliata la formalizzazione.

Preliminarmente, però, è innanzitutto d'uopo specificare che, com'è stato anticipato nello scorso paragrafo, il modello considera singolarmente le varie strategie di ricorso che sono state presentate: per tale ragione il modello è composto da tre varianti le quali considerano rispettivamente:

- la sola possibilità di ricorrere a tempi di overtimes (prima variante del modello, denominata “**MG_OVER**”);
- la possibilità di ricorrere a tempi di overtimes nonché quella di effettuare operazioni di swapping (seconda variante del modello, denominata “**MG_OVER+SWAP**”);
- la possibilità di ricorrere a tempi di overtimes nonché quella di effettuare operazioni di posticipo (terza variante del modello, denominata “**MG_OVER+POST**”).

Quanto suddetto si può giustificare in maniera plausibile con il fatto che essendo questo modello stocastico proposto come una sorta di DSS per chi si occupa della fase di pianificazione e di gestione del blocco operatorio, la scelta di ricorrere singolarmente ad una sola delle tre strategie suddette rende meno “caotico” e più lineare il flusso di attività che avviene all'interno di ogni sala operatoria (anche a livello psicologico per gli attori coinvolti in questo “processo produttivo” sarebbe meno indicato l'idea di poter avere oltre ai tempi di overtimes, anche contemporaneamente sia operazioni di posticipo che operazioni di swapping per i casi elettivi e proprio per questo motivo si rende

giustificabile l'idea di scindere il modello nelle varianti suddette, garantendo allo stesso tempo anche una complessità computazionale minore del modello stocastico stesso, il che non è sicuramente un dettaglio marginale: il decisore può così valutare singolarmente ciascuna di queste suddette strategie e di scegliere quindi conseguentemente ai vari criteri di ottimalità la più conveniente tra esse).

A questo punto si può iniziare a presentare ogni aspetto inerente alla rappresentazione matematica ed alla notazione di ciascuna delle tre varianti suddette:

- **INSIEMI:**

I = insieme dei casi elettivi in attesa di assegnamento, con indice i ;

E = insieme dei casi urgenti, con indice e ;

K = insieme di sale operatorie (OR), con indice k ;

S = insieme degli scenari, con indice s ;

J = insieme di specialità chirurgiche, con indice j ;

I_j = sottoinsieme dei casi elettivi richiedenti la specialità chirurgica j ($\bigcup_{j \in J} I_j = I$, dato

che chiaramente ogni caso elettivo richiede una sola specialità);

E_j = sottoinsieme dei casi urgenti richiedenti la specialità chirurgica j

($\bigcup_{j \in J} E_j = E$, dato che chiaramente ogni caso urgente richiede una sola

specialità).

- **PARAMETRI:**

d_{is} = durata del caso elettivo i sotto lo scenario s [ore];

d_{es} = durata del caso urgente e sotto lo scenario s [ore];

t_{jk} = tempo in cui la sala operatoria k è disponibile per la specialità chirurgica j [ore];

pri_i = priorità del caso elettivo i ;

p_s = probabilità dello scenario s ;

Cov = costo orario dell'overtime [€/ora];

$Cswap_i$ = costo per lo spostamento del caso elettivo i ad un'altra sala operatoria [€]

(parametro presente solo nella seconda variante del modello);

C_{post}_i = costo per il posticipo del caso elettivo i ad un giorno successivo [€] (parametro presente solo nella terza variante del modello).

• **VARIABILI:**

$x_{ik} = 1$ se il caso elettivo i è assegnato alla sala operatoria k , 0 altrimenti.

$y_{eks} = 1$ se il caso urgente e è assegnato alla sala operatoria k sotto lo scenario s , 0 altrimenti;

$o_{jks} =$ quantità di overtime per la specialità chirurgica j nella sala operatoria k sotto lo scenario s [ore];

z_{ikls} (con $l \neq k$) = 1 se il caso elettivo i assegnato alla sala operatoria k nel 1° stadio è poi spostato alla sala operatoria l nella soluzione di 2° stadio sotto lo scenario s , 0 altrimenti (variabile presente solo nella seconda variante del modello);

$x^{-}_{is} = 1$ se il caso elettivo i assegnato ad una generica sala operatoria nel 1° stadio è posticipato ad un giorno successivo nella soluzione di 2° stadio sotto lo scenario s , 0 altrimenti (variabile presente solo nella terza variante del modello).

• **FORMULAZIONE GENERALE DELLA VARIANTE “MG_OVER”:**

$$\min \sum_{i \in I} p r_i \cdot u_i + \sum_{j \in J} \sum_{k \in K} \sum_{s \in S} Cov \cdot p_s \cdot o_{jks} \quad (1)$$

s.v.

$$\sum_{k \in K} x_{ik} + u_i = 1, \forall i \in I \quad (2)$$

$$\sum_{i \in I_j} d_{is} \cdot x_{ik} + \sum_{e \in E_j} d_{es} \cdot y_{eks} \leq t_{jk} + o_{jks}, \forall k \in K, \forall j \in J, \forall s \in S \quad (3)$$

$$\sum_{k \in K} y_{eks} = 1, \forall e \in E, \forall s \in S \quad (4)$$

$$x_{ik}, u_i, y_{eks} \in [0,1], o_{jks} \geq 0, \forall i \in I, \forall e \in E, \forall k \in K, \forall j \in J, \forall s \in S \quad (5)$$

- (1): funzione obiettivo del modello, che si propone di minimizzare i costi straordinari dovuti al non-scheduling dei casi elettivi ed ai costi totali di overtimes;
- (2): insieme di vincoli che assicura che ciascun caso elettivo i nel 1° stadio venga assegnato ad una sala operatoria per essere schedulato o che non venga schedulato nella giornata lavorativa corrente;
- (3): insieme di vincoli che assicura che in ogni sala operatoria k , per ogni specialità chirurgica j e sotto ogni scenario s la durata complessiva dei casi elettivi e dei casi urgenti che si decide di schedulare (tra 1° e 2° stadio) non ecceda il tempo totale operativo della sala (determinato dalla somma tra il blocco temporale giornaliero di ogni specialità chirurgica in ogni sala operatoria più i possibili overtimes dipendenti da ciascuna specialità chirurgica e che potrebbero sorgere nella stessa sala operatoria);
- (4): insieme di vincoli che impone che sotto ogni scenario s quando si viene a presentare al sistema sanitario un caso urgente e , esso deve essere schedulato in giornata in qualche sala operatoria k del sistema stesso;
- (5): vincoli di interezza per le variabili binarie e di non negatività per le variabili continue.

• **FORMULAZIONE GENERALE DELLA VARIANTE “MG_OVER+SWAP”:**

$$\min \sum_{i \in I} p r_i \cdot u_i + \sum_{s \in S} p_s \left[\sum_{i \in I} \left(C_{\text{swap}_i} \cdot \sum_{k \in K} \sum_{l \in K: l \neq k} z_{ikls} \right) + \sum_{k \in K} \sum_{j \in J} C_{\text{cov}} \cdot o_{jks} \right] \quad (1)$$

s.v.

$$\sum_{k \in K} x_{ik} + u_i = 1, \forall i \in I \quad (2)$$

$$\sum_{l \in K: l \neq k} z_{ikls} \leq x_{ik}, \forall i \in I, \forall k \in K, \forall s \in S \quad (3)$$

$$\sum_{i \in I_j} d_{is} \cdot \left[x_{ik} - \sum_{l \in K: l \neq k} (z_{ikls} - z_{ilk s}) \right] + \sum_{e \in E_j} d_{es} \cdot y_{eks} \leq t_{jk} + o_{jks}, \forall k \in K, \forall j \in J, \forall s \in S \quad (4)$$

$$\sum_{k \in K} y_{eks} = 1, \forall e \in E, \forall s \in S \quad (5)$$

$$x_{ik}, u_i, y_{eks}, z_{ikls} \in [0,1], o_{jks} \geq 0, \forall i \in I, \forall e \in E, \forall k, l \in K, \forall j \in J, \forall s \in S \quad (6)$$

(1): funzione obiettivo del modello, che si propone di minimizzare i costi straordinari dovuti al non-scheduling dei casi elettivi ed ai costi totali sia di overtimes che dovuti alle operazioni di “swapping” dei casi elettivi nel 2° stadio;

(2): insieme di vincoli che assicura che ciascun caso elettivo i nel 1° stadio venga assegnato ad una sala operatoria per essere schedulato o che non venga schedulato nella giornata lavorativa corrente;

(3): insieme di vincoli che assicura che lo scheduling di un caso elettivo i può essere spostato nel 2° stadio sotto lo scenario s dalla sala operatoria k alla sala operatoria l solo se tale caso elettivo era stato schedulato nel 1° stadio nella sala operatoria k ;

(4): insieme di vincoli che assicura che in ogni sala operatoria k , per ogni specialità chirurgica j e sotto ogni scenario s la durata complessiva dei casi elettivi e dei casi urgenti che si decide di schedulare (tra 1° e 2° stadio) non ecceda il tempo totale operativo della sala (determinato dalla somma tra il blocco temporale giornaliero di ogni specialità chirurgica in ogni sala operatoria più i possibili overtimes dipendenti da ciascuna specialità chirurgica e che potrebbero sorgere nella stessa sala operatoria);

(5): insieme di vincoli che impone che sotto ogni scenario s quando si viene a presentare al sistema sanitario un caso urgente e , esso deve essere schedulato in giornata in qualche sala operatoria k del sistema stesso;

(6): vincoli di interezza per le variabili binarie e di non negatività per le variabili continue.

**• FORMULAZIONE GENERALE DELLA VARIANTE
“MG_OVER+POST”:**

$$\mathbf{min} \sum_{i \in I} pri_i \cdot u_i + \sum_{s \in S} p_s \left[\sum_{i \in I} Cpost_i \cdot x_{is}^- + \sum_{k \in K} \sum_{j \in J} Cov \cdot o_{jks} \right] \quad (1)$$

s.v.

$$\sum_{k \in K} x_{ik} + u_i = 1, \forall i \in I \quad (2)$$

$$x_{is}^- \leq x_{ik}, \forall i \in I, \forall k \in K, \forall s \in S \quad (3)$$

$$\sum_{i \in I_j} d_{is} \cdot (x_{ik} - x_{is}^-) + \sum_{e \in E_j} d_{es} \cdot y_{eks} \leq t_{jk} + o_{jks}, \forall k \in K, \forall j \in J, \forall s \in S \quad (4)$$

$$\sum_{k \in K} y_{eks} = 1, \forall e \in E, \forall s \in S \quad (5)$$

$$x_{ik}, u_i, y_{eks}, x_{is}^- \in [0,1], o_{jks} \geq 0, \forall i \in I, \forall e \in E, \forall k \in K, \forall j \in J, \forall s \in S \quad (6)$$

(1): funzione obiettivo del modello, che si propone di minimizzare i costi straordinari dovuti al non-scheduling dei casi elettivi ed ai costi totali sia di overtimes che dovuti alle operazioni di posticipo dei casi elettivi nel 2° stadio;

(2): insieme di vincoli che assicura che ciascun caso elettivo i nel 1° stadio venga assegnato ad una sala operatoria per essere schedulato o che non venga schedulato nella giornata lavorativa corrente;

(3): insieme di vincoli che assicura che lo scheduling di un caso elettivo i può essere posticipato nel 2° stadio sotto lo scenario s ad una giornata lavorativa successiva solo se tale caso elettivo era stato schedulato nel 1° stadio in qualche sala operatoria;

(4): insieme di vincoli che assicura che in ogni sala operatoria k , per ogni specialità chirurgica j e sotto ogni scenario s la durata complessiva dei casi elettivi e dei casi urgenti che si decide di schedulare (tra 1° e 2° stadio) non ecceda il tempo totale operativo della sala (determinato dalla somma tra il blocco temporale giornaliero di ogni specialità chirurgica in ogni sala operatoria più i possibili overtimes dipendenti da ciascuna specialità chirurgica e che potrebbero sorgere nella stessa sala operatoria);

(5): insieme di vincoli che impone che sotto ogni scenario s quando si viene a presentare al sistema sanitario un caso urgente e , esso deve essere schedulato in giornata in qualche sala operatoria k del sistema stesso;

(6): vincoli di interezza per le variabili binarie e di non negatività per le variabili continue.

3.3 Analisi della struttura e delle proprietà dei modelli

Dopo aver illustrato in maniera esaustiva la formalizzazione di ciascuna delle tre varianti del modello stocastico proposto (in termini di insiemi, di parametri, di variabili decisionali e di formulazione generale di ogni variante del modello stesso), si può passare dunque ad analizzarne in maniera più dettagliata la struttura e le proprietà intrinseche.

Iniziando da un punto di vista prettamente strutturale, si può partire analizzando la *complessità computazionale* del modello stesso, ovvero le risorse minime necessarie per la risoluzione di un'istanza inerente ad ogni sua variante: c'è allora da puntualizzare che, in generale, un modello stocastico con ricorso a due stadi è di per sé un modello computazionalmente complesso e che assumendo certe dimensioni la fase di risoluzione può anche essere molto onerosa in termini di tempi di calcolo richiesti.

Di sicuro per la variante "MG_OVER", dato che in essa non si vanno a considerare oltre ai tempi di overtimes né le possibilità di effettuare operazioni di swapping per i casi elettivi e né le possibilità di posticipare i casi elettivi nel 2° stadio, la complessità computazionale sarà minore rispetto alle altre due varianti in quanto il numero complessivo di variabili e di vincoli di un modello inerente a tale variante sarà senza dubbio minore e questo potrebbe far ipotizzare che conseguentemente sarà minore il suo tempo di risoluzione, cosicché è d'uopo specificare che in generale il tempo di risoluzione di un modello non è sempre direttamente proporzionale al numero complessivo di variabili decisionali e di vincoli.

Di sicuro la complessità computazionale nella seconda variante "MG_OVER+SWAP", a parità di cardinalità degli insiemi dei dati iniziali considerati per una generica istanza del problema, va ad aumentare rispetto alla prima variante, dato che come si è detto in questa seconda variante si vanno a considerare oltre agli overtimes, come possibilità di ricorso, anche le operazioni di swapping per i casi elettivi nel 2° stadio: in questo caso il numero di variabili del modello va ad aumentare sensibilmente dato che vengono introdotte quelle dell'insieme che nel paragrafo precedente è stato denominato z e che include per l'appunto le variabili relative a tali operazioni di "swapping" e conseguentemente, oltre al numero totale di variabili decisionali, andrà rispetto sempre alla prima variante del modello ad aumentare anche il numero totale di vincoli, poiché in questa variante ne è stato introdotto un insieme in più che è per l'appunto quello che va a puntualizzare che un caso elettivo nel 2° stadio sotto un determinato scenario può essere spostato da una certa sala operatoria k ad un'altra sala operatoria l solo se questo caso elettivo era stato schedulato naturalmente dal 1° stadio nella sala operatoria k , poiché chiaramente in caso contrario l'operazione di spostamento di sala operatoria per quel caso elettivo non sarebbe ammissibile.

Allo stesso modo rispetto alla prima variante la complessità computazionale andrà ad aumentare, sempre chiaramente a parità di cardinalità degli insiemi dei dati iniziali considerati per una generica istanza del problema, nella terza variante "MG_OVER+POST", che, riepilogando, va a considerare oltre agli overtimes come

possibilità di ricorso anche le operazioni di posticipo per i casi elettivi nel 2° stadio. Anche in questo caso il numero di variabili del modello va ad aumentare poiché sono introdotte stavolta quelle dell'insieme che nello scorso paragrafo è stato denominato x - e che include per l'appunto le variabili relative a tali operazioni di posticipo: il numero totale di variabili decisionali di un modello inerente a tale variante sarà però, assumendo sempre parità di cardinalità degli insiemi dei dati iniziali considerati per una generica istanza, minore rispetto alla seconda variante e questo perché chiaramente l'insieme delle variabili decisionali z ha dimensioni maggiori rispetto a quello delle variabili decisionali x .

Rispetto alla prima variante del modello, comunque, andrà ad aumentare sensibilmente anche in questa terza variante del modello (come suddetto anche nella seconda variante) il numero complessivo di vincoli, poiché in questa variante è stato introdotto un insieme in più rispetto alla prima variante che è per l'appunto quello che va a puntualizzare che un caso elettivo nel 2° stadio sotto un determinato scenario può essere posticipato ad una giornata lavorativa successiva solo se questo caso elettivo era naturalmente stato schedato dal 1° stadio a qualche sala operatoria, dato che chiaramente contrariamente l'operazione di posticipo per quel caso elettivo non sarebbe ammissibile se già nel 1° stadio stesso si era deciso di non schedarlo nella giornata lavorativa corrente.

Comunque può essere interessante osservare che sebbene, come suddetto (e sempre assumendo di trovarsi a parità di cardinalità degli insiemi dei dati iniziali considerati per una generica istanza del problema), il numero totale di variabili decisionali sia maggiore nella seconda variante del modello rispetto alla terza variante del modello stesso, in queste due varianti il numero totale di vincoli è invece identico e questo avviene perché nella seconda variante del modello l'insieme dei vincoli aggiuntivi rispetto alla prima variante del modello, così come nella terza variante del modello l'insieme dei vincoli aggiuntivi rispetto sempre alla prima variante del modello stesso hanno una stessa cardinalità $|I| * |K| * |S|$ (dove $|I|$ rappresenta la cardinalità dell'insieme dei casi elettivi in lista d'attesa per l'assegnamento, ovvero il numero totale di casi elettivi che il sistema sanitario dovrebbe schedare in giornata (la forma del verbo al modo condizionale è giustificato dal fatto che chiaramente, come detto, un caso elettivo non deve essere, a differenza di un caso urgente, obbligatoriamente schedato nella giornata lavorativa corrente); $|K|$ rappresenta la cardinalità dell'insieme delle sale operatorie del sistema sanitario, ovvero il numero totale di sale operatorie di cui il sistema stesso è munito; e $|S|$ rappresenta la cardinalità dell'insieme degli scenari di riferimento, ovvero il numero totale di scenari considerati).

Dopo aver discusso circa la complessità computazionale di ciascuna delle tre suddette varianti del modello stocastico presentato, si può passare ora a farne un confronto con lo stato dell'arte.

Come si è mostrato in maniera approfondita nello scorso capitolo, la letteratura offre una vasta gamma di modelli (deterministici o stocastici che siano) incentrati per lo più sulla fase "operatoria" e con l'obiettivo di supportarne il livello decisionale operativo fornendo uno scheduling in genere su base giornaliera che vada a rispettare determinati criteri di ottimalità (per lo più minimizzazione dei costi e dei tempi di inattività e/o di overtime): in linea a quanto suddetto anche questo modello segue tali canoni ma lo fa assumendo alcune condizioni al contorno, che sono state illustrate dettagliatamente nel primo paragrafo di questo capitolo, e differenziandosi dagli altri modelli proposti dalla letteratura cerca innanzitutto di superarne i possibili limiti riscontrati in essi da un'analisi critica.

Uno dei punti di forza del modello stocastico proposto in questo lavoro di tesi di dottorato è sicuramente dovuto al fatto che il sistema sanitario di riferimento va qui a considerare prioritariamente i casi urgenti a differenza di tanti altri modelli presenti in letteratura che, com'è stato ampiamente mostrato sempre nello scorso capitolo, vanno ad ignorarli totalmente o a supporre di trattarli ipotizzando di trovare una decisione al momento in cui essi si vengono a presentare: questo modello, a loro differenza, impone che essi qualora si presentino al sistema debbano essere schedulati nell'immediatezza e senza possibilità dunque di inviarli all'esterno per poter far sì che vengano trattati.

Da un punto di vista legato invece all'incertezza decisionale, inoltre, questo modello va a trattarne specificamente i due fattori preponderanti che all'interno di un blocco operatorio, com'è stato ampiamente discusso nel quinto paragrafo del primo capitolo, la rendono maggiormente incidente ovvero le durate degli interventi chirurgici e le possibilità che si possano presentare al sistema dei casi urgenti (cosa che non tutti i modelli della letteratura presentati nello scorso capitolo e che considerano i casi urgenti stessi vanno a fare).

In più è sicuramente vantaggioso l'insieme di soluzioni di ricorso che il modello stocastico presentato in questo capitolo va a proporre per permettere al decisore di avere uno scheduling ottimale in termini di costi straordinari: nessuno tra i modelli presenti in letteratura di cui si è discusso nello scorso capitolo considera infatti oltre alla possibilità di introdurre i tempi di overtime le opzioni di valutare singolarmente (ed unitamente ai tempi di overtime stessi) l'impatto di introdurre anche delle operazioni di scambio della sala operatoria per un caso elettivo rispetto a quanto era stato deciso nel 1° stadio oppure

di ricorrere anche ad operazioni di posticipo per un caso elettivo che il 1° stadio aveva deciso di schedulare.

Il modello stocastico proposto, dunque, si presenta come uno strumento quantitativo efficiente in grado di coadiuvare il decisore nella fase di gestione operativa del blocco operatorio e lo fa ricercando nei suoi aspetti peculiari di superare in qualche modo i limiti riscontrati in un'analisi critica degli altri modelli presenti in letteratura ed esaminati nello scorso capitolo di questo elaborato di tesi.

I capitoli successivi andranno a rafforzare tale supposizione, dimostrando in termini pratici e di natura squisitamente tecnica come il modello stesso sia capace di mantenere un adeguato trade-off legato anche all'efficienza oltre che all'efficacia.

Capitolo 4

Metodi risolutivi

4.1 Generalità su approcci esatti

Dopo aver presentato nel precedente capitolo le tre varianti del modello MIP stocastico a due stadi con ricorso proposto per la fase di gestione operativa del blocco operatorio e dopo averne illustrato in termini sommarî la complessità computazionale legata a ciascuna delle tre varianti, è d'uopo passare ad introdurre ora i metodi risolutivi ai quali si può ricorrere per ottenere la soluzione di un'istanza di ciascuna di tali varianti del modello.

In generale c'è da dire che nonostante per la maggior parte dei problemi di ottimizzazione il processo risolutivo sia di per sé piuttosto complesso, nel corso degli anni sono stati sviluppati numerosi metodi efficaci in grado di risolvere (anche riconducendo all'ottimalità) una vasta gamma di tali problemi ed inerenti a svariati ambiti in un tempo computazionale accettabile.

In generale, le tecniche di risoluzione disponibili possono essere classificate in due grandi classi:

- **approcci (algoritmi o metodi) esatti**, i quali sono in grado di trovare la soluzione ottima di un problema e di dimostrarne l'ottimalità o comunque di assicurare, in caso questo non sia possibile, che il problema non ammette soluzioni realizzabili (cioè che il problema è inammissibile);
- **approcci (algoritmi o metodi) approssimati (euristici)**, i quali sacrificano la garanzia di trovare la soluzione ottima (perché per ottenerla ci vorrebbe troppo tempo) nell'interesse di ottenere soluzioni "buone" (le quali si discostano di non molto dalla soluzione ottima) in un tempo computazionale ragionevolmente accettabile.

Dopo aver specificato questa sostanziale differenza tra le due classi suddette, in questo paragrafo è d'uopo iniziare a presentare le generalità sugli approcci esatti.

Cercando di fornire un'altra definizione di carattere generale, si può dire allora che dato un problema P , un approccio esatto E è una procedura algoritmica che fornisce una soluzione ottima di una qualsivoglia istanza I di P (oppure dimostra che I non ammette soluzione ottima).

I vantaggi di un approccio esatto rispetto ad uno approssimato (euristico) sono notevoli e possono essere riassunti così:

- se un algoritmo esatto ha successo, esso fornirà la soluzione ottima del problema;
- nel caso in cui l'algoritmo si arresti prima del completamento della procedura di ricerca dell'ottimo (magari perché si viene a definire uno dei suoi criteri d'arresto anche prima di trovare la soluzione ottima), vengono comunque fornite preziose informazioni circa i limiti inferiori/superiori della soluzione ottima;
- un metodo esatto va ad escludere le parti dello spazio di ricerca ("regione ammissibile" di un problema) in cui non ci possono essere soluzioni ottime.

Ma di contro un approccio esatto può avere anche degli svantaggi, altrimenti non avrebbe senso l'introduzione di un approccio approssimato (euristico):

- per molti problemi la dimensione delle istanze risolubili nella pratica è piuttosto limitata ed il tempo computazionale necessario aumenta fortemente al crescere della dimensione dell'istanza stessa;
- la memoria necessaria per un approccio esatto può essere anche molto grande e portare così ad un precoce aborto di un programma;
- è difficile estendere un approccio esatto molto efficace per un dato problema, specie nel caso in cui nella formulazione del problema stesso va a variare qualche dettaglio.

Lo stato attuale degli approcci esatti, in generale, vede per un certo tipo di problema che qualche istanza anche di notevole dimensione può essere risolta molto velocemente, mentre per un altro tipo di problema può accadere l'esatto contrario ossia che per qualche istanza di dimensione ridotta il tempo di risoluzione può essere molto elevato.

Tornando al campo di applicazione preso in considerazione per questo lavoro di tesi e riferendosi dunque alla PS (*Programmazione Stocastica*), facendo in particolar modo ai *modelli a due stadi con ricorso fisso* (ricordando che è proprio essa la tipologia di approccio utilizzata da ciascuna variante del modello presentato nello scorso capitolo) c'è da dire che essi possono essere risolti secondo metodi noti in letteratura.

Per distribuzioni discrete si ha in genere che il numero di variabili e di vincoli va moltiplicato per il numero di scenari (che può essere anche elevato poiché sommariamente si può dire che più esso è per l'appunto elevato e meglio è rappresentata l'incertezza): dunque per tante variabili è improponibile utilizzare gli stessi metodi utilizzati nel campo d'applicazione deterministico (ovvero “*algoritmo del simplesso primale*” o “*algoritmo del simplesso duale*” per la PL ed “*algoritmo di Branch & Bound*” per la PLI).

Per tale ragione la letteratura nel corso degli anni ha studiato vari approcci, che sono raggruppabili in due grandi categorie:

- **metodi di decomposizione** (i quali decompongono il problema iniziale in tanti “sottoproblemi” di più piccola dimensione);
- **metodi a punti interni** (i quali operano invece sulle strutture del duale del problema di partenza, sfruttandone le proprietà).

A loro volta i **metodi di decomposizione** sono classificati in **primali** (nei quali si decompone rispetto allo stadio) e **duali** (nei quali si decompone invece rispetto allo scenario).

Il metodo primale più noto è il cosiddetto **metodo L-Shaped** [35] ed è basato sulla “decomposizione di Benders” (che prevede una linearizzazione esterna del problema basata sulla proprietà che la funzione di ricorso è convessa e lineare a tratti su un dominio poliedrale per distribuzioni finite secondo quanto dice in merito un teorema).

In tale metodo si va a risolvere inizialmente un “*rilassato*” di questo problema detto “*problema master*” e si prosegue aggiungendo iterativamente dei cosiddetti “*tagli di ammissibilità*” per rappresentare il valore atteso della funzione di ricorso, con questi tagli che si ottengono risolvendo un insieme di problemi di PL definiti per ogni scenario, i quali hanno l'importante proprietà di essere sempre limitati ed ammissibili.

Quando sarà invece possibile, per qualche soluzione di primo stadio, procedere con la valutazione della funzione di ricorso saranno introdotti invece i cosiddetti “*tagli di ottimalità*”, fino a pervenire così alla soluzione ottima.

Tale metodo presenta anche la variante a “*tagli multipli*” ed in questo caso i tagli introdotti nel problema master assumono una forma disaggregata, dato che allo stesso “problema master” l'informazione passata ad ogni iterazione è maggiormente dettagliata.

Il metodo duale più noto è invece quello della cosiddetta **copertura progressiva** [36], il quale si basa sulla costruzione di una funzione *lagrangiana* “*aumentata*” (nella quale si

vanno ad introdurre nella funzione obiettivo anche i vincoli di “*non-anticipatività*”, opportunamente pesati da un parametro di penalità: tali vincoli, come già detto nel corso del secondo capitolo, in un modello a due stadi con ricorso vanno a restringere le decisioni di primo stadio che possono essere prese prima di conoscere la realizzazione delle variabili aleatorie).

Questo metodo decompone il problema di partenza in un numero di sottoproblemi pari al numero di scenari considerati (usando la tecnica di “*splitting delle colonne*”), con i vincoli di non-anticipatività che vengono espressi mediante un certo vettore, denominato “*incombente*” (per valori fissati di tale vettore si perviene ad un problema completamente separabile rispetto agli scenari e risolubile in maniera agevole, in quanto di dimensione ridotta).

Le soluzioni dipendenti dagli scenari vanno a mano a mano aggregate tra di loro fino ad ottenere un nuovo vettore incombente, finché i suddetti vincoli di “non-anticipatività” non saranno soddisfatti.

Questo metodo, come detto, è annoverato tra i cosiddetti metodi “*lagrangiani aumentati*” poiché va ad introdurre nella funzione obiettivo anche tali vincoli di “non-anticipatività” pesati da opportuni parametri.

C’è infine da dire che la scelta iniziale del vettore incombente può influenzare notevolmente l’efficienza del metodo stesso.

Passando invece ai **metodi a punti interni**, c’è da dire che essi si rivelano essere molto efficienti per risolvere problemi di grande dimensione e con una certa struttura della matrice dei coefficienti tecnologici di una generica istanza del problema stesso: il più noto in questa tipologia di metodi è quello primale/duale di tipo “*path following (PF)*” [37].

L’insieme dei metodi di questa famiglia deve il loro nome ad una loro importante caratteristica: tali metodi, infatti, a differenza del tradizionale “*metodo del simplesso*” utilizzato nella PL non esplorano soltanto i vertici della regione ammissibile (questo perché il teorema fondamentale della PL informa che proprio in un vertice di tale regione ammissibile è possibile trovare la soluzione ottima del problema), ma anche i punti che si trovano dentro tale regione, evitando comunque di avvicinarsi troppo alla frontiera se non in prossimità della soluzione ottima.

La suddetta condizione si introduce andando a sostituire i tradizionali vincoli di “non-negatività”, che sono di solito presenti in un problema di PL, con una “*barriera logaritmica*” pesata da un certo parametro e della quale si tiene conto nella funzione obiettivo.

Il problema che considera tale barriera ha l'importante proprietà di essere convesso con funzione obiettivo strettamente convessa e delle stesse proprietà gode anche il suo duale. A questo punto bisogna quindi ricavarne le cosiddette “*condizioni di Karush-Kuhn-Tucker (KKT)*” e trovarne la soluzione utilizzando il “*metodo di Newton*”.

Tale metodo parte da un punto e determina le direzioni di ricerca per poi risolvere il “*sistema di Newton*”, che oltre alle KKT considera anche gli scostamenti: ci si sposterà così ad un'altra soluzione usando dei vari passi lungo i quali muoversi seguendo tale suddetta direzione di ricerca appena calcolata, in modo da preservare l'ammissibilità della soluzione stessa e fino a trovare così iterativamente la soluzione ottima del problema di partenza.

Come tutti i metodi a punti interni, dunque, anche il PF ha l'onere computazionale di determinare le direzioni di ricerca risolvendo, come suddetto, il sistema di Newton.

Per risolvere tale sistema si possono usare vari metodi quali le fattorizzazioni sparse o le tecniche di “*splitting delle colonne*”, le quali includono esplicitamente i vincoli di “non-anticipatività” ed hanno inoltre il vantaggio di ridurre la densità della matrice dei coefficienti tecnologici, ma di contro accrescono la dimensione del problema stesso e per questo motivo sarebbe d'uopo pensare all'introduzione di tecniche di “*splitting parziale*”, da adoperare dunque sulle sole variabili di primo stadio del problema.

4.2 Generalità su approcci euristici

Dopo aver introdotto gli approcci esatti ed averne presentato una panoramica inerente ai metodi per i modelli a due stadi con ricorso per la PS, si può passare ora a definire le generalità circa gli approcci euristici.

Come detto nello scorso paragrafo, mentre un approccio esatto fornisce la soluzione ottima di un problema, un approccio euristico fornisce invece una soluzione “buona” in quanto abbastanza vicina all'ottimo stesso.

Un approccio algoritmico esatto, difatti, per molti classi di problemi non è in grado di risolvere alcune istanze del generico problema stesso in un tempo computazionale sufficientemente rapido o comunque accettabile: per tale ragione si giustifica l'introduzione di un approccio algoritmico approssimato (euristico).

Data infatti una generica istanza I di un problema P , un algoritmo approssimato H è una procedura che tenta di fornire “rapidamente” (o comunque in tempi computazionali

“accettabili”) una soluzione ammissibile e quanto più possibile vicina all’ottima (se non addirittura proprio quella ottima) di P .

Contestualmente a tale definizione, c’è da aggiungere quella di un algoritmo euristico proprio poiché, come si già è detto nello scorso paragrafo, non tutti gli algoritmi approssimati sono di tipo euristico (per esempio si può pensare ad un problema di PLI e ad un algoritmo esatto di Branch & Bound bloccato alla prima iterazione: si tratta di un approccio approssimato ma non euristico).

In generale allora si può dire che si definisce “*euristico*” (termine che deriva dal verbo greco “εὐρίσκω” (“*eurisco*”), ovvero “trovare”) un algoritmo che su base empirica utilizza una tecnica particolare al fine di determinare una soluzione “buona” (quanto più possibile vicina alla soluzione ottima) riducendo complessità computazionale, capacità di elaborazione e/o tempi di risoluzione di una generica istanza di un problema.

Si può pertanto dire che, data un’istanza ammissibile I di un problema P , si ha che:

- $z^*(I)$ = valore ottimo, assunto dalla funzione obiettivo in corrispondenza della soluzione ottima di I ;
- $zH(I)$ = valore assunto dalla funzione obiettivo in corrispondenza di una soluzione ammissibile fornito dall’algoritmo euristico H per l’istanza I .

Dividendo tra di loro tali due grandezze si introduce il cosiddetto rapporto di prestazione $\rho_H(I) = zH(I) / z^*(I)$, il quale indica quanto l’algoritmo euristico H è stato efficiente per trovare una soluzione ammissibile per l’istanza I di P (in altre parole quantifica la “bontà” della soluzione determinata, specificando quanto essa è vicina o meno alla soluzione ottima dell’istanza I del problema P stesso).

In generale si ha che $\rho_H(I) = 1$ se H fornisce la soluzione ottima di I , ovvero se l’algoritmo euristico in questione si comporta proprio come un normale algoritmo esatto, mentre si ha $\rho_H(I) = 0$ se H fallisce per qualche causa e non è dunque capace di determinare alcuna soluzione ammissibile per I .

Un algoritmo euristico è in genere composto da due fasi distinte: una fase **costruttiva** (nella quale l’obiettivo è quello di determinare una prima soluzione ammissibile per il problema) ed una fase **migliorativa** (applicata in cascata alla fase costruttiva e che ha l’obiettivo di migliorare la qualità della soluzione ammissibile di partenza determinata dalla suddetta fase, generandone così una “migliore”, cioè con miglior valore della funzione obiettivo).

Famiglie di algoritmi appartenenti alla fase puramente costruttiva di un approccio euristico sono gli algoritmi “casuali” (“*random search*”), “*greedy*”, “*ibridi casuali-greedy*”, “*fix-and-relax*”, “*rolling-horizon*”...

Per quanto riguarda invece la fase migliorativa vi sono invece gli algoritmi di “*ricerca locale*”, di “*Tabu Search (TS)*”, di “*Simulated Annealing (SA)*”, gli algoritmi “*genetici e memetici*”, di “*Greedy Randomized Adaptive Searching Procedure (GRASP)*”...

Una volta presentate dunque le caratteristiche generali sia di un approccio esatto che di un approccio euristico, si può dire che è più conveniente pertanto utilizzare un approccio euristico piuttosto che un approccio esatto nei seguenti casi:

- quando le dimensioni del problema sono eccessive rispetto a quelle risolubili mediante algoritmi esatti;
- quando i problemi, anche se di una certa dimensione, devono essere risolti in tempi computazionali piuttosto brevi;
- quando i dati del problema sono approssimati e pertanto non vale la pena di cercare la soluzione ottima;
- quando si risolvono problemi simili ma non identici a quelli affrontati dagli algoritmi esatti.

Ma d'altronde c'è anche da dire che studiare gli algoritmi esatti è sensato perché proprio da un'analisi su di essi si possono sviluppare algoritmi euristici capaci di affrontare in tempi computazionali contenuti istanze di uno stesso problema ma con maggiori dimensioni, si può acquisire conoscenza sulla natura decisionale del problema ed infine si possono valutare esattamente gli errori delle euristiche e dei rilassamenti (quantomeno per problemi di piccola dimensione).

4.3 Disegno e sviluppo dell'euristica proposta

Una volta presentati gli aspetti generali di un approccio di tipo esatto e di un approccio di tipo euristico, si può dire che a causa della complessità computazione di ognuna delle tre varianti del modello stocastico a due stadi con ricorso presentate nello scorso capitolo, si rende necessaria la proposta di un metodo euristico che sia capace di trattare una generica istanza del problema per ciascuna delle suddette tre varianti in maniera efficace (ossia fornendo “buone” soluzioni) ed in un tempo computazionale accettabile.

Per tali ragioni è stato quindi definito un approccio euristico, in cui all'algoritmo generale per ogni specifica variante del modello è aggiunto qualche differente dettaglio al fine di

tenere conto della diversa opzione che la specifica variante del modello consente di scegliere al decisore per definire lo scheduling giornaliero del blocco operatorio (ossia, riepilogando quanto detto nello scorso capitolo, “solo overtime” per la variante “MG_OVER”, “overtime+swapping” per la variante “MG_OVER+SWAP” od “overtime+posticipi” per la variante “MG_OVER+POST”).

Tale approccio euristico è suddiviso in due fasi:

- una prima fase di natura **costruttiva**, il cui obiettivo è quello di fornire una prima soluzione ammissibile per il modello;
- una seconda fase **migliorativa**, in cui chiaramente si punta a migliorare la soluzione ammissibile di partenza rendendola quanto più possibile vicina alla soluzione ottima del problema stesso.

Si può passare ora a definirne dettagliatamente gli aspetti ed iniziando dalla fase **costruttiva**, c'è da dire che essa, a sua volta, è suddivisa in **due parti**:

- una prima parte di costruzione per le variabili **y** (che, secondo quanto detto nello scorso capitolo, rappresentano le decisioni di scheduling sui casi urgenti);
- una seconda fase di costruzione per tutte le rimanenti variabili di ciascuna delle varianti del modello.

Nella **prima parte** di tale fase costruttiva, per ogni specialità e per ogni scenario, si ordinano dapprima tutti i blocchi temporali rimanenti in ogni sala operatoria in ordine decrescente di tempo restante (inizialmente è chiaro che tutti tali blocchi avranno stessa durata restante, pari al massimo valore).

Allo stesso modo, per ogni specialità e per ogni scenario, vengono ordinati in ordine decrescente di durata i casi urgenti che si vengono a presentare.

A questo punto, in maniera molto semplice, si estraggono con disciplina FIFO (acronimo di “*First Input First Output*”, ovvero il primo elemento che entra in lista è anche il primo ad uscirne) i casi urgenti dalla lista e si vanno a schedulare nelle sale operatorie, seguendo il loro ordine decrescente in lista di blocco temporale restante per lo specifico scenario e per la specifica specialità chirurgica.

E si proseguirà ripetendo così quanto suddetto per tutte le specialità chirurgiche e per tutti gli scenari, fino a schedulare pertanto tutti i casi urgenti (dato che chiaramente il

problema ha come vincolo quello di dover obbligatoriamente schedulare in qualche sala operatoria tutti i casi urgenti che si vengono a presentare sotto ogni scenario).

Terminata questa prima parte della fase costruttiva avendo fissato le variabili y , si passa ora alla **seconda parte**.

Per prima cosa viene preliminarmente scelto in maniera del tutto casuale uno degli scenari di riferimento e si suppone banalmente che, avendo per esempio una piena conoscenza degli eventi futuri, sia proprio quello che si viene a manifestare.

Quindi per ogni specialità chirurgica si ordinano dapprima tutti i blocchi temporali rimanenti in ogni sala operatoria in ordine decrescente di tempo restante sotto lo scenario fissato (naturalmente questo avverrà tenendo in considerazione le decisioni sulle variabili già fissate nella prima parte della fase costruttiva stessa).

Allo stesso modo sempre per ogni specialità, e sempre sotto tale scenario, si vanno ad ordinare in ordine decrescente di durata tutti i casi elettivi che si vengono a presentare al sistema sanitario.

Come nella prima parte, in maniera molto semplice si estraggono così ad uno ad uno con disciplina FIFO tutti i casi elettivi in lista e li si va a schedulare nella sala operatoria con più capienza nel blocco temporale di riferimento, valutando se eventualmente per farlo si può incorrere anche parzialmente in un tempo di overtime.

Se per caso dovesse succedere, l'algoritmo valuterà la convenienza di dover schedulare ugualmente il caso elettivo accettando di avere un tempo di overtime oppure se sarà meno costoso prendere su di esso una decisione di non-scheduling.

Alla base di ciò, si proseguirà così per tutte le specialità fino a completare dunque lo scheduling di tutti i casi elettivi (per ognuno si deciderà così se schedularlo in qualche sala operatoria, come specificheranno le variabili x , o se effettuare per esso una decisione di non-scheduling, come specificheranno invece le variabili u).

A questo punto verranno valutati tutti gli eventuali costi di overtime (variabili o) e se si è nella variante "MG_OVER" del modello, la fase costruttiva termina qui e si valuta il costo totale della soluzione appena determinata.

Se si è invece nella variante "MG_OVER+SWAP" si va prima di chiudere con questa fase a valutare per quei casi elettivi schedulati ricorrendo anche parzialmente ad un tempo di overtime cosa succederebbe se li si andasse a scambiare di sala operatoria, sempre sotto lo scenario di riferimento scelto preliminarmente in questa seconda parte, decidendo di rischedularli in un'altra sala operatoria (che tra le sale operatorie rimanenti è quella con più tempo restante, se c'è, o quella con minor tempo di overtime totale).

Se tale mossa sarà conveniente verranno considerate anche tali decisioni sulle variabili z , altrimenti no e così terminerà anche in questo caso la fase costruttiva con la valutazione del costo totale della soluzione determinata.

Allo stesso modo nella variante “MG_OVER+POST” si va prima di chiudere con questa fase a valutare per i casi elettivi schedulati ricorrendo anche parzialmente ad un tempo di overtime cosa succederebbe se li si andasse a posticipare sotto lo scenario di riferimento scelto casualmente a priori.

Se in questo caso tale mossa sarà conveniente verranno considerate anche tali decisioni sulle variabili x , altrimenti no e così terminerà la fase costruttiva anche per quest’altra variante del modello dopo aver al solito valutato il costo della soluzione appena determinata.

A questo punto, per ognuna delle tre varianti del modello, la fase costruttiva ha quindi determinato una soluzione di partenza che è sicuramente ammissibile perché rispetta tutti i vincoli per ogni variante del modello.

Scatta a questo punto la fase migliorativa e per essa c’è invece da specificare che si tratta di una **procedura iterativa**, in cui il tempo di esecuzione è specificato dal decisore oppure può essere imposto per default.

Durante questo tempo di esecuzione verranno svolte quante più possibili iterazioni migliorative e la soluzione di partenza fornita dalla fase costruttiva verrà conseguentemente migliorata, in quanto resa il più possibile vicina alla soluzione ottima del modello.

Dopo aver definito ciò, si può passare a descrivere dettagliatamente cosa avviene in ciascuna iterazione di tale fase.

C’è da specificare che ogni iterazione è composta da un certo numero di **passi** per ogni variante del modello. Una generica iterazione è composta da:

- due passi per la variante “MG_OVER”;
- tre passi per la variante “MG_OVER+SWAP”
- tre passi per la variante “MG_OVER+POST”.

I primi due passi, comunque, sono identici per tutte le varianti del modello mentre naturalmente ciò che va a cambiare specificamente è la natura del terzo passo, là dove esso è chiaramente presente.

Il **primo passo** di tale fase migliorativa va preliminarmente a considerare tutti i casi elettivi che si è deciso di schedulare in qualche sala operatoria k .

Iterativamente se ne prende in maniera casuale uno alla volta e si va a scegliere (sempre in maniera casuale) un'altra sala operatoria **l** passando quindi a valutare cosa succederebbe, tenendo chiaramente conto dell'impatto sui tempi di overtimes, se lo scheduling del caso elettivo corrente fosse ora spostato dalla sala operatoria **k** alla sala operatoria **l**.

Se questa scelta comporterà una diminuzione dei costi, essa verrà effettuata altrimenti no e così via si proseguirà fino a completare con il numero di iterazioni per questo primo passo (numero che sarà chiaramente imposto dal numero di casi elettivi che la soluzione corrente aveva deciso di schedulare) e si potrà così passare al secondo passo.

Il **secondo passo** di tale fase migliorativa, allo stesso modo, va a considerare sempre tutti i casi elettivi che si è deciso di schedulare in qualche sala operatoria **k**.

Iterativamente se ne prende in maniera casuale uno alla volta e si va ora a valutare invece cosa succederebbe se si decidesse ora di non schedularlo più (passando così dal pattern $x_{ik}=1$ e $u_i=0$ al pattern $x_{ik}=0$ e $u_i=1$, con i che è chiaramente il caso elettivo preso singolarmente in esame), tenendo chiaramente conto dell'impatto che avrebbe tale decisione sui tempi di overtimes e sui costi di non-scheduling.

Se questa scelta comporterà una diminuzione dei costi, essa verrà effettuata altrimenti no e così via si proseguirà anche in questo caso fino a completare con il numero di iterazioni per questo secondo passo (che sarà sempre chiaramente imposto dal numero di casi elettivi che la soluzione corrente aveva deciso di schedulare).

Se si è nella variante "MG_OVER" l'iterazione della procedura migliorativa è terminata qui e si andrà a tenere memoria del valore della soluzione trovata (e di tutte le sue variabili) solo se essa sarà migliore di quella dell'iterazione precedente (se si tratta della prima iterazione naturalmente il confronto verrà fatto con il valore della soluzione trovata dalla fase costruttiva).

Se ci sarà ancora tempo restante scatterà un'altra iterazione migliorativa e così via fino al termine del numero di iterazioni migliorative, quando verrà restituita pertanto la migliore soluzione trovata.

Se si è invece in una delle altre due varianti del modello, scatterà invece il **terzo passo**, il quale (come suddetto) differirà in base alla specifica variante.

Iniziando dalla variante "MG_OVER+SWAP", c'è da dire che in questo terzo passo verranno considerati preliminarmente tutti i casi elettivi che si è già deciso di schedulare in qualche sala operatoria **k** e per i quali la soluzione non prevede già inoltre nessuna operazione di swapping sotto qualche scenario.

Iterativamente se ne considera uno alla volta e scegliendo in maniera casuale sia un'altra sala operatoria **I** che un sottoinsieme di cardinalità pari al 10% del numero totale di scenari di riferimento del modello, si va a valutare cosa succederebbe introducendo l'operazione di swapping del caso elettivo corrente dalla sala **k** alla sala **I** singolarmente sotto ognuno di tali scenari considerati.

Se tale scelta comporterà una diminuzione dei costi, essa verrà effettuata altrimenti no e così via si proseguirà fino a completare con il numero di iterazioni possibili per questo passo.

Una volta terminato tale numero di iterazioni possibili, l'iterazione corrente per la fase migliorativa sarà terminata anche per tale variante del modello e si terrà memoria del valore di soluzione trovato solo se esso sarà migliore rispetto a quello di prima. Ugualmente a prima se ci sarà ancora tempo scatterà un'altra iterazione migliorativa, altrimenti no e verrà restituito alla fine il valore della migliore soluzione trovata.

Per quanto riguarda invece la variante "**MG_OVER+POST**" in questo terzo passo, allo stesso modo di quanto detto per la variante che prevede le operazioni di swapping, verranno considerati preliminarmente tutti i casi elettivi che si è già deciso di schedulare in qualche sala operatoria e per i quali la soluzione non prevede già inoltre nessuna operazione di posticipo sotto qualche scenario.

Iterativamente se ne considera uno alla volta e scegliendo anche in questo caso ed in maniera casuale un sottoinsieme di cardinalità pari al 10% del numero totale di scenari di riferimento del modello, si va a valutare cosa succederebbe introducendo una decisione di posticipo sullo scheduling del caso elettivo corrente singolarmente sotto ognuno di tali scenari considerati.

Se tale scelta comporterà una diminuzione dei costi, essa verrà effettuata altrimenti no e così via si proseguirà fino a completare con il numero di iterazioni possibili per questo passo.

Anche in questa variante del modello, una volta terminato tale numero di iterazioni possibili, l'iterazione corrente per la fase migliorativa sarà terminata e si terrà memoria del valore di soluzione trovato solo se esso sarà migliore rispetto a quello di prima.

In maniera sempre del tutto identica a prima se ci sarà ancora tempo scatterà un'altra iterazione migliorativa, altrimenti no e verrà restituito anche in questo caso alla fine il valore della migliore soluzione trovata.

L'approccio euristico presentato è dunque di per sé lungo ed articolato, ma allo stesso modo è anche dotato di una struttura logica ed abbastanza comprensibile. Per fissare

meglio i concetti inerenti ad esso ed articolarne una sua struttura maggiormente comprensibile si può fare riferimento allo schema riassuntivo sottostante:

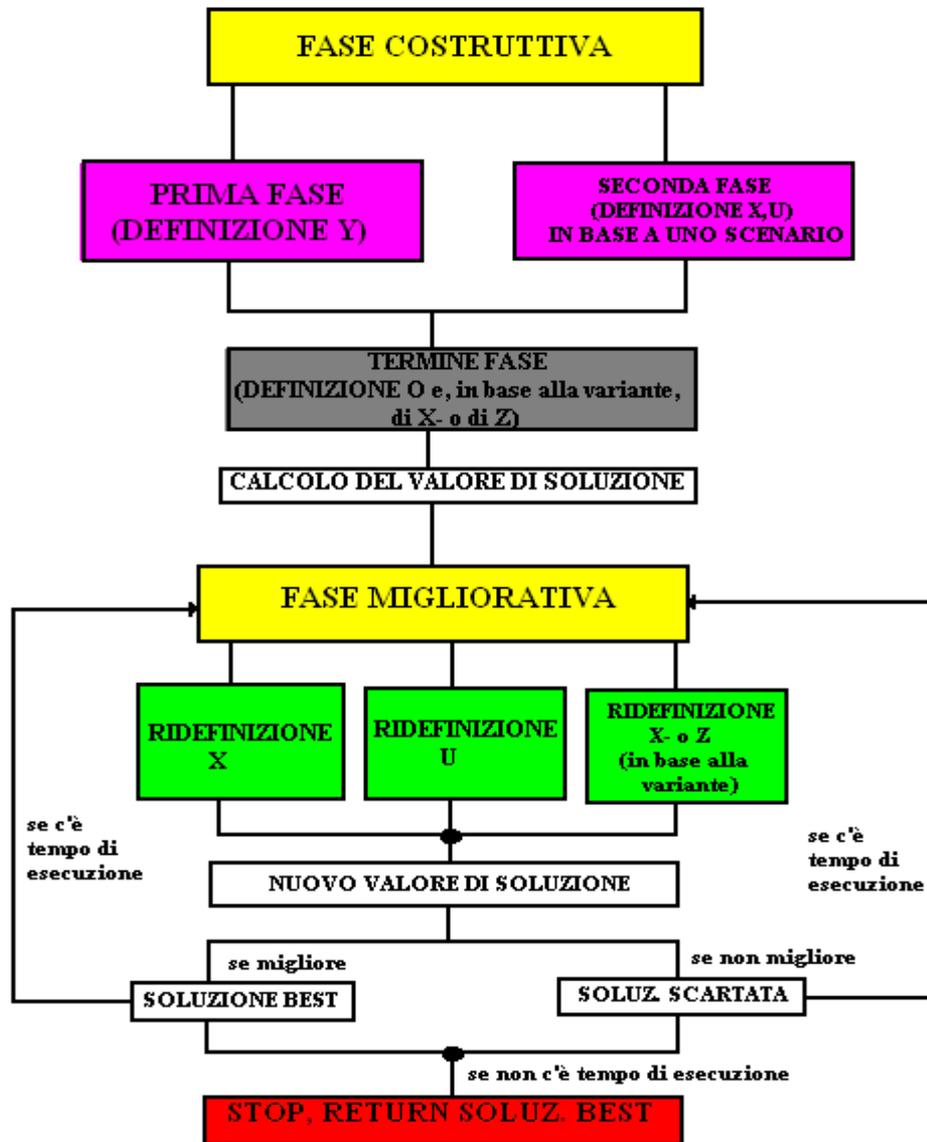


Figura 4.1: tabella riassuntiva delle varie fasi che compongono l'approccio euristico proposto.

L'approccio euristico appena presentato è stato sviluppato in ambiente AIMMS (acronimo di "Advanced Interactive Multidimensional Modeling System"), un sistema software progettato per modellare e risolvere problemi di ottimizzazione di dimensione elevata e che comprende al suo interno un linguaggio di modellazione algebrica. Il CD a corredo del presente tomo fornisce il codice sviluppato con dettagliata descrizione.

Capitolo 5

Esperimenti computazionali

5.1 Strategia sperimentale

Un aspetto da non trascurare è dunque la complessità computazionale di una generica istanza delle varianti del modello, aspetto di cui si è già discusso nell'ultimo paragrafo del terzo capitolo.

In tal senso, è stato introdotto un approccio risolutivo di tipo euristico ed obiettivo di questo capitolo è dimostrare che tale approccio è in grado di fornire “buone” soluzioni (intendendo per “buone” il fatto che esse siano piuttosto vicine all'ottimo) in un tempo risolutivo ragionevolmente accettabile.

Per tutte queste ragioni sono state quindi costruite una serie di istanze inerenti a ciascuna delle tre varianti del modello stocastico proposto nell'ambito del terzo capitolo, nella modalità che sarà illustrata dal paragrafo successivo.

Utilizzando quindi il software *AIMMS* è stata dunque effettuata una fase di testing, al fine di stabilire una base di confronto tra i valori di soluzione ottenuti ed i tempi computazionali del tradizionale risolutore **CPLEX 12.1** (utilizzato dal medesimo software) e tra i valori di soluzione ottenuti ed i tempi computazionali dell'**euristica** proposta nello scorso capitolo, per ogni medesima istanza suddetta.

Gli esiti attesi degli esperimenti di tale fase di testing mirano a dimostrare che l'utilizzo dell'euristica si manifesta come un “buon surrogato” del tradizionale risolutore CPLEX, allorquando quest'ultimo richiede tempi computazionali inaccettabili.

Inoltre si punta ad enfatizzare sul fatto che l'utilizzo dell'euristica, oltre ad abbassarne notevolmente i tempi di risoluzione per una generica istanza di una delle tre varianti del modello, è anche capace di determinare soluzioni non troppo distanti dall'ottimo.

Il tutto al fine di garantire al decisore che tale euristica è in grado di determinare una “buona” soluzione in un tempo computazionale ragionevole.

Quanto suddetto equivale pertanto, utilizzando altri termini, a garantire un giusto “trade-off” d'efficacia e d'efficienza circa l'utilizzo dell'euristica stessa.

Sarebbe infatti inefficace avere un metodo euristico capace di calcolare in un tempo computazionale ristretto una soluzione per un problema che però si discosta notevolmente da quella ottima (ovvero calcolando una soluzione che si può definire “grossolana”).

Ed invece sarebbe inefficiente utilizzare un metodo euristico capace di calcolare una soluzione per un problema che sia molto vicina a quella ottima, ma richiedendo per farlo un tempo computazionale elevato quasi quanto quello di un tradizionale metodo di risoluzione.

Per questo motivo serve quindi un’euristica capace di conciliare entrambi gli aspetti suddetti.

5.2 Costruzione delle istanze per la fase di testing

Procedendo con ordine, è d’uopo partire illustrando le modalità con cui sono state costruite le istanze per ognuna delle tre varianti del modello considerate nella suddetta fase di testing.

Per prima cosa c’è da illustrare la codifica utilizzata per la denominazione di ogni istanza, la quale è infatti composta dalla seguente legenda: “**#CASI_TOTALI . #SALE . #SCENARI - #SPECIALITA**”.

Fornendo un esempio, l’istanza denominata “**30.2.100-3**” si riferisce ad un contesto che nella giornata presa in esame come orizzonte di pianificazione temporale ha **30** casi complessivi (di cui un certo numero elettivi ed un certo numero urgenti).

Tali casi sono da schedulare in **2** sale operatorie disponibili, con **100** possibili scenari che si possono realizzare e con ciascuno dei casi stessi che richiede una tra le **3** specialità chirurgiche che il sistema stesso si propone di trattare.

Gli aspetti inseriti nella suddetta codifica per la denominazione di una generica istanza di una delle varianti del modello, racchiudono comunque tutti i parametri di input che sono stati richiesti all’utente al fine di costruire un’istanza per la fase di testing.

A partire da tali valori, infatti, se ne ricavano altri mediante alcune assunzioni di fondo con il fine di creare un’istanza di un problema per ogni variante del modello.

In tale fase di testing è difatti assunto per semplicità che tra i casi complessivi, l’80% sia riferito ai casi elettivi mentre il restante 20% ai casi urgenti (seguendo dunque l’esempio suddetto, quindi, di questi **30** casi complessivi ve ne sono **24** elettivi ed i restanti **6** urgenti).

Inoltre si è stabilito a priori di avere un tempo regolare lavorativo giornaliero di 12 ore per ogni sala operatoria.

Tale tempo deve essere partizionato in blocchi di uguale durata per ogni specialità chirurgica, poiché si assume che tali sale operatorie siano multifunzionali (quindi avendo assunto nell'esempio suddetto di avere 3 specialità chirurgiche, si stabilisce pertanto di dedicare in ognuna delle 2 sale operatorie del sistema sanitario un blocco di 4 ore giornaliere per ciascuna delle specialità stesse, andando ad inizializzare così la matrice t_{jk}).

Per quanto riguarda gli scenari, si è stabilito che essi siano tutti equiprobabili (seguendo sempre l'esempio suddetto, dunque, si ha che ognuno dei 100 scenari ha quindi una probabilità di realizzazione annessa pari all'1%, inizializzando così il vettore p_s).

Per quanto riguarda invece il costo orario dell'overtime (parametro Cov) si è assunto in tutte le istanze che esso sia pari a 100 €/ora.

Il medesimo software, inoltre, si occupa inoltre di inizializzare i restanti parametri di ingresso, ovvero:

- il livello di priorità di un caso elettivo (ovvero istanziare il vettore pri_i);
- di stabilire quale tra le specialità chirurgiche è richiesta da un caso generico (ovvero istanziare gli insiemi I_j ed E_j);
- la durata di un caso generico sotto ogni scenario assumendo, qualora si tratti di un caso urgente, di avere durata nulla sotto un certo scenario se tale caso urgente non si viene a presentare sotto tale scenario suddetto (ovvero istanziare le matrici d_{is} e d_{es});
- il valore dei costi di “swapping” per i casi elettivi nella variante “MG_OVER+SWAP” (ovvero istanziare il vettore $Cswap_i$);
- il valore dei costi di posticipo per i casi elettivi stabilire nella variante “MG_OVER+POST” (ovvero istanziare il vettore $Cpost_i$).

Nella fattispecie c'è comunque da specificare che per quanto riguarda i casi elettivi, il valore di priorità (i cui valori sono espressi dal vettore pri_i) rispetto al “colore del codice” annesso (che può essere “bianco”, “azzurro”, “verde”, “giallo” o “rosso”, come si è già discusso ampiamente nel corso del primo paragrafo terzo capitolo) è stato specificato a priori per ognuna delle “classi di priorità” suddette.

Allo stesso modo per ogni caso elettivo è stato specificato a priori, in base sempre alla sua “classe di priorità”, la distribuzione di valori che ne assumerà la durata sotto ogni scenario.

L’identica cosa è stata fatta per ogni caso urgente, ma in maniera un po’ differente, dato che i casi urgenti non hanno un “colore di codice” annesso che ne distingue il grado di priorità e considerato che essi non vengono obbligatoriamente a presentarsi sotto tutti gli scenari.

Dopo aver fatto un’analisi di quanto utilizzato in letteratura per trattare l’aspetto suddetto, si è stabilito di ricorrere all’utilizzo di una funzione di distribuzione di tipo “**lognormale**” al fine di istanziare i valori delle durate di un caso sotto ogni scenario (ovvero istanziare le matrici d_{is} e d_{es}).

La suddetta analisi della letteratura dimostra infatti che si ricorre in genere all’utilizzo di una distribuzione uniforme o a quello di una distribuzione lognormale, com’è possibile afferrare facilmente esaminando la tabella sottostante:

<i>Autore/i del modello</i>	<i>Distribuzione utilizzata</i>
Arnaout e Kulbashian	UNIFORM (per classi)
Lamiri, Grimaud e Xie	UNIFORM (0,5 – 3 ore)
Gupta e Denton	UNIFORM
Zhang, Murali, Dessouky e Belson	LOGNORMAL
Jebali, Alouane e Ladet	LOGNORMAL (μ : 3 ore, σ : 1 ora)
Mancilla e Storer	UNIFORM
Guinet e Chaabane	LOGNORMAL

Figura 5.1: tipologie di distribuzioni utilizzate in letteratura per istanziare le durate dei vari casi.

In generale una distribuzione uniforme assume con semplicità che tali durate vadano a variare in un intervallo delimitato da un valore minimo ed un valore massimo che tale grandezza può assumere.

Un limite legato al suo utilizzo è dovuto al fatto che con essa non è dunque possibile rappresentare situazioni in cui a causa di complicazioni un caso assume un valore di durata non comune.

Se per esempio ci si pone in una distribuzione uniforme tra 0,5 e 3 ore non ci sarà la possibilità di rappresentare il fatto che un caso, a causa di una complicazione di qualsivoglia natura, potrà assumere sotto qualche scenario (uno o più che siano) una durata per l’appunto superiore alle 3 ore.

La distribuzione lognormale, invece, è più completa sotto questo punto di vista in quanto è in grado di “fotografare” sotto determinati scenari situazioni di complicazioni che possono portare ad una durata fuori dal comune per un generico caso (elettivo o urgente che esso sia).

Siccome la distribuzione lognormale supera dunque il limite posto dall’utilizzo della distribuzione uniforme, è per questo motivo che la si è scelta in questo contesto.

Tale distribuzione lognormale nel software *AIMMS* richiede in ingresso tre parametri, ossia “**Shape**” (che rappresenta la media dei valori che essa deve andare ad inizializzare), “**LB**” (che rappresenta il “lower-bound” ossia il limite inferiore che ogni valore da essa inizializzato può assumere) e “**Scaling**” (parametro opzionale e che indica, in sintesi, quanto un valore inizializzato può “assomigliare” ad un altro: più tale valore è basso e più tale “somiglianza” può essere notata).

Per illustrare tutti i suddetti aspetti è quindi opportuno introdurre la seguente tabella.

Tale tabella indica per ogni caso elettivo il valore di priorità in base al “colore del codice” annesso e per ogni caso (elettivo o urgente che sia), il valore dei parametri con cui è istanziata la funzione lognormale che si deve occupare, come detto, di stabilirne il valore delle durate sotto ogni scenario:

Priorità (codice) dell'intervento	Pri(i)	Lognormal (shape-LB-scaling) (in ore)
<i>Bassa (BIANCO)</i>	<i>100</i>	<i>0,5-0,375-0,25</i>
<i>Medio-bassa (AZZURRO)</i>	<i>200</i>	<i>1-0,75-0,25</i>
<i>Media (VERDE)</i>	<i>300</i>	<i>1,5-1,125-0,25</i>
<i>Medio-alta (GIALLO)</i>	<i>400</i>	<i>2-1,5-0,25</i>
<i>Alta (ROSSO)</i>	<i>500</i>	<i>2,5-1,875-0,25</i>
<i>Caso urgente</i>	<i>-</i>	<i>3-2,25-0,5</i>

Figura 5.2: corrispondenza tra codice dell'intervento, valore di priorità e parametri di input della funzione “lognormal” per stabilire i valori di durata di un caso sotto i vari scenari.

Al fine di specificare comunque tutti gli aspetti sin qui elencati, è opportuno illustrare per termini sommarî il contenuto nella procedura **Inizializzazione** nella fase di definizione del modello in *AIMMS*.

Innanzitutto c'è da dire che tale procedura stabilisce in maniera casuale a quale specialità chirurgica appartiene un generico caso, elettivo od urgente che sia.

Per ognuno dei casi elettivi si seleziona quindi in maniera casuale un indice numerico al fine di assegnarne un valore di priorità.

A questo punto per ogni caso elettivo, in base al valore di priorità assegnato, si inizializza la matrice delle sue durate passando gli opportuni parametri alla funzione "lognormal", seguendo in questo caso quanto cita la tabella sovrastante in base al valore di priorità dello stesso caso elettivo.

Per ogni caso urgente, invece, un parametro ausiliario permette di stabilire casualmente se esso sotto un determinato scenario si viene a presentare oppure no: se sì, si passano gli opportuni parametri alla funzione "lognormal" (sempre seguendo quanto citato nella tabella sovrastante), altrimenti si impone che il valore di durata del caso urgente sotto tale scenario è chiaramente nullo.

La procedura appena illustrata è quella inerente alla variante "MG_OVER" del modello, ma ciò che va ad aggiungersi nelle altre due varianti è legato soltanto rispettivamente ai costi di swapping (per quanto riguarda la variante "MG_OVER+SWAP") ed ai costi di posticipo per i casi elettivi (per quanto riguarda invece la variante "MG_OVER+POST"). Ci si deve limitare solo ad aggiungere dunque che per la variante "MG_OVER+SWAP" si è imposto per ogni caso elettivo che il suo costo di swapping C_{swap_i} sia pari al 2% del suo valore di priorità e che, invece, per la variante "MG_OVER+POST" si è imposto per ogni caso elettivo che il suo costo di posticipo C_{post_i} sia esattamente pari a tale suo valore di priorità.

5.3 Risultati ottenuti nella fase di testing

Le varie istanze del modello sono state testate al fine di valutare la qualità delle soluzioni ed i tempi computazionali richiesti sia da CPLEX che dall'euristica.

Le seguenti sperimentazioni sono state quindi testate su un notebook *Intel Core 2 con processore da 1,66 GHz e sistema operativo Windows XP Professional* mediante l'utilizzo del software *AIMMS 3.9*.

Come si era detto nel precedente capitolo, l'utente deve stabilire a priori la durata dell'intervallo di tempo entro cui desidera ottenere la soluzione euristica.

In questo intervallo di tempo, la fase migliorativa dell'euristica stessa effettuerà quante più iterazioni possibili al fine di restituire la miglior soluzione trovata una volta che esso sarà scaduto.

Pertanto, al fine di valutare le prestazioni di CPLEX e dell'euristica su uno stesso "orizzonte temporale risolutivo" si è deciso di svolgere una **prima batteria** di sperimentazioni in cui si è concesso a CPLEX il tempo necessario per risolvere all'ottimo ogni istanza del problema per ciascuna delle tre varianti.

Inoltre si è imposto all'euristica un tempo computazionale pari proprio a quello impiegato da CPLEX per determinare la soluzione ottima.

In questo modo si vanno ad apprezzare gli scostamenti ottenuti, al fine di valutare quanto l'euristica si avvicina alla soluzione ottima determinata dal CPLEX.

Ecco quindi i risultati ottenuti per ciascuna delle tre varianti del modello. Iniziando dalla variante "MG_OVER" si è ottenuto:

Test	#variabili	#vincoli	F.O. CPLEX [€]	F.O. Euristica [€]	Scarto % F.O.	CPU time CPLEX [s]	CPU time Euristica [s]
20.2.20-3	253	179	1014,01	1021,87	0,78%	0,062	0,890
20.2.50-3	533	409	1006,05	1010,62	0,45%	1,281	1,830
20.2.75-3	811	623	515,55	552,28	7,12%	1,734	2,880
25.2.50-3	633	457	880,38	883,00	0,30%	1,469	2,670
25.2.75-3	901	666	869,16	882,92	1,58%	1,203	3,000
30.2.20-3	315	206	2146,13	2182,74	1,71%	0,172	2,440
30.2.50-3	699	488	1348,25	1354,99	0,50%	1,047	3,560
30.2.75-3	991	709	1343,45	1356,05	0,94%	2,937	5,220
35.2.50-3	713	493	2408,38	2432,78	1,01%	0,141	4,560
35.2.75-3	1061	742	2192,99	2209,76	0,76%	0,313	7,140
40.3.50-3	1254	708	2141,13	2177,68	1,70%	4,078	6,670
40.3.75-3	1680	1000	1643,94	1662,90	1,71%	70,594	70,860
50.3.50-3	1367	743	3448,34	3517,18	2,00%	379,407	387,81
50.3.75-3	1967	1093	3482,65	3541,96	1,70%	18,812	26,060

Figura 5.3: risultati ottenuti nella variante "MG_OVER" lasciando calcolare a CPLEX la soluzione ottima e valutando l'euristica sullo stesso tempo risolutivo.

Proseguendo quindi con la variante "MG_OVER+SWAP" si è invece ottenuto:

Test	#variabili	#vincoli	F.O. CPLEX [€]	F.O. Euristica [€]	Scarto % F.O.	CPU time CPLEX [s]	CPU time Euristica [s]
20.2.20-3	893	819	1011,68	1014,95	0,32%	36,360	36,420
20.2.50-3	2133	2009	987,73	1006,11	1,86%	4401,891	4402,070
20.2.75-3	3211	3023	509,71	546,29	7,18%	7595,766	7598,460
25.2.50-3	2633	2457	873,61	880,56	0,80%	10022,67	10047,86

25.2.75-3	3901	3666	865,45	879,41	1,61%	7412,079	7421,780
30.2.20-3	1275	1166	2138,03	2176,76	1,81%	8,297	9,550
30.2.50-3	3099	2888	1340,85	1352,91	0,90%	5527,016	5527,31
30.2.75-3	4591	4309	1340,07	1354,91	1,11%	11061,06	11234,87
35.2.50-3	3513	3293	2406,37	2432,66	1,09%	99,875	100,620
35.2.75-3	5261	4942	2192,71	2205,38	0,58%	51,375	58,570
40.3.50-3	10854	5508	2141,09	2152,15	0,52%	8897,28	8905,34
40.3.75-3	16080	8200	1641,29	1667,18	1,58%	5331,062	5381,150
50.3.50-3	13367	6743	3435,61	3513,22	2,26%	6878,44	6886,70
50.3.75-3	19967	10093	3468,69	3529,20	1,74%	5798,015	5798,640

Figura 5.4: risultati ottenuti nella variante “MG_OVER+SWAP” lasciando calcolare a CPLEX la soluzione ottima e valutando l’euristica sullo stesso tempo risolutivo.

E concludendo infine con la variante “MG_OVER+POST” si è ottenuto:

Test	#variabili	#vincoli	F.O. CPLEX [€]	F.O. Euristica [€]	Scarto % F.O.	CPU time CPLEX [s]	CPU time Euristica [s]
20.2.20-3	573	819	1014,01	1016,34	0,23%	0,218	0,970
20.2.50-3	1333	2009	1006,05	1010,47	0,44%	1,328	2,220
20.2.75-3	2011	3023	515,55	552,02	7,07%	3,078	4,200
25.2.50-3	1633	2457	880,38	881,40	0,12%	1,235	2,690
25.2.75-3	2401	3666	869,16	882,00	1,48%	2,109	3,430
30.2.20-3	795	1166	2146,13	2177,70	1,47%	0,469	2,130
30.2.50-3	1899	2888	1348,25	1353,46	0,39%	2,718	3,990
30.2.75-3	2791	4309	1343,45	1354,48	0,82%	4,109	5,960
35.2.50-3	2113	3293	2408,38	2427,79	0,81%	0,313	4,360
35.2.75-3	3161	4942	2192,99	2203,04	0,46%	1,219	9,060
40.3.50-3	2584	5508	2141,13	2155,46	0,67%	6,375	17,200
40.3.75-3	4080	8200	1643,94	1650,49	0,40%	71,407	89,770
50.3.50-3	3367	6743	3448,34	3476,46	0,82%	150,359	150,720
50.3.75-3	4967	10093	3482,65	3508,55	0,74%	40,547	41,750

Figura 5.5: risultati ottenuti nella variante “MG_OVER+POST” lasciando calcolare a CPLEX la soluzione ottima e valutando l’euristica sullo stesso tempo risolutivo.

Al fine di valutare il comportamento delle due metodologie risolutive in un determinato intervallo di tempo, si è poi effettuata una **seconda batteria** di sperimentazioni in cui si è imposto un tempo computazionale di 15 minuti sia per CPLEX che per l’euristica.

Chiaramente per questa seconda batteria di sperimentazioni le dimensioni delle istanze sono maggiori al fine di rendere più complesso e difficoltoso il processo computazionale realizzato da CPLEX.

Iniziando al solito dalla variante “MG_OVER”, ecco i risultati ottenuti:

Test	#variabili	#vincoli	F.O. CPLEX [€]	F.O. Euristicica [€]	Scarto % F.O.	CPU time CPLEX [s]	CPU time Euristicica [s]
30.2.200-3	2447	1812	1324,55 *	1331,74	0,05%	900	905,56
40.3.200-3	2620	4290	1718,81*	1732,08	0,77%	900	924,67
50.3.200-3	4982	2848	3692,71 *	3756,65	1,73%	900	941,88
60.4.200-3	7477	3658	3266,93 *	3356,52	2,74%	900	918,26

Figura 5.6: risultati ottenuti nella variante “MG_OVER” imponendo sia a CPLEX che all’euristica un tempo computazionale di 15 minuti (900 secondi).

Proseguendo poi con la variante “MG_OVER+SWAP”, si è ottenuto invece:

Test	#variabili	#vincoli	F.O. CPLEX [€]	F.O. Euristicica [€]	Scarto % F.O.	CPU time CPLEX [s]	CPU time Euristicica [s]
30.2.200-3	12047	11412	1315,67 *	1331,50	1,20%	900	1145,64
40.3.200-3	42690	21820	1716,23 *	1732,01	0,92%	900	958,38
50.3.200-3	52982	26848	3689,81 *	3752,45	1,70%	900	1006,59
60.4.200-3	122677	42058	3250,20 *	3401,04	4,64%	900	1285,72

Figura 5.7: risultati ottenuti nella variante “MG_OVER+SWAP” imponendo sia a CPLEX che all’euristica un tempo computazionale di 15 minuti (900 secondi).

E concludendo infine con la variante “MG_OVER+POST” si è ottenuto:

Test	#variabili	#vincoli	F.O. CPLEX [€]	F.O. Euristicica [€]	Scarto % F.O.	CPU time CPLEX [s]	CPU time Euristicica [s]
30.2.200-3	7247	11412	1324,55 *	1327,88	0,25%	900	912,05
40.3.200-3	10690	21820	1718,81 *	1717,87	-0,05%	900	904,41
50.3.200-3	12982	26848	3692,12 *	3731,69	1,88%	900	914,31
60.4.200-3	17077	42058	3267,41 *	3314,43	1,44%	900	907,65

Figura 5.8: risultati ottenuti nella variante “MG_OVER+POST” imponendo sia a CPLEX che all’euristica un tempo computazionale di 15 minuti (900 secondi).

Occorre osservare che i risultati denotati dall'asterisco non rappresentano necessariamente la soluzione ottima dell'istanza, bensì il valore della migliore soluzione trovata dal CPLEX dopo un tempo computazionale di 15 minuti.

5.4 Analisi dei risultati ottenuti

In questo paragrafo si cercherà di esaminare e di valutare quindi i risultati numerici ottenuti.

In relazione alla “prima batteria” di sperimentazioni la tabella in figura 5.3, ossia quella riferita ai risultati per la variante “MG_OVER”, mostra nella colonna denominata “Scarto % F.O.” come tale valore si mantenga piuttosto contenuto.

Esso è infatti oscillante tra lo 0,30% riferito alla sperimentazione 25.2.50-3 ed il 2% per la sperimentazione 50.3.50-3, eccezion fatta per la sperimentazione 20.2.75-3 in cui tale valore assume valore massimo pari al 7,12%.

I tempi risolutivi impiegati dai due metodi per risolvere ogni istanza del problema assumono un valore compreso tra pochissimi decimi di secondo per alcune sperimentazioni ai 6 minuti e 27 secondi impiegati dall'euristica per la sperimentazione 50.3.50-3.

La tabella in figura 5.4, ossia quella riferita invece alla variante “MG_OVER+SWAP”, mostra anch'essa nella colonna denominata “Scarto % F.O.” come tale valore si mantenga anche in questo caso piuttosto contenuto.

In particolar modo esso è oscillante tra lo 0,32% riferito alla sperimentazione 20.2.20-3 al 2,26% riferito alla sperimentazione 50.3.50-3, eccezion fatta ancora una volta per la sperimentazione 20.2.75-3 in cui tale valore assume valore massimo pari al 7,18%.

I tempi risolutivi impiegati dai due metodi per risolvere ogni istanza del problema assumono valori inaccettabili per l'utente se non per 4 delle 14 istanze in cui esso varia dagli 8 ai 100 secondi.

Per le rimanenti 10 istanze, invece, esso varia dai 4402 secondi (ovvero 1 ora, 13 minuti e 22 secondi) della sperimentazione 20.2.50-3 ad addirittura 11.235 secondi (3 ore, 7 minuti e 15 secondi) per la sperimentazione 30.2.75-3.

Infine la tabella in figura 5.5, ossia quella riferita invece alla variante “MG_OVER+POST” per la prima batteria di sperimentazioni, mostra al solito anch'essa nella colonna denominata “Scarto % F.O.” come tale valore si mantenga anche in questo caso, al pari dei due casi precedenti, piuttosto contenuto.

In particolar modo esso è oscillante tra lo 0,12% riferito alla sperimentazione 25.2.50-3 all'1,48% riferito invece alla sperimentazione 25.2.75-3, eccezion fatta ancora una volta e non a caso (al pari delle altre due varianti) per la sperimentazione 20.2.75-3 in cui tale valore assume valore massimo pari al 7,07%.

Complessivamente si può dunque dire, per la sperimentazione 20.2.75-3, che in tutte e tre le varianti del modello tale scarto tra la soluzione ottima determinata dal CPLEX e la soluzione determinata nello stesso intervallo di tempo impiegato del CPLEX medesimo dall'euristica si mantiene nell'ordine del 7,10% circa a causa probabilmente del valore della soluzione di partenza determinato dalla fase costruttiva dell'euristica.

Probabilmente per tale sperimentazione, a differenza di tutte le altre, è stato scelto uno scenario di riferimento che ha dato luogo ad una soluzione di partenza particolarmente distante dalla soluzione ottima e la fase migliorativa è stata in grado di ridurre parzialmente tale scarto, abbassandolo così al 7,10% circa in tutte e tre le varianti (che anche se è più alta rispetto a tutte le altre sperimentazioni, non è comunque una percentuale particolarmente distante dall'ottimo).

Questo indica che chiaramente per la fase euristica conta e non in maniera del tutto indifferente il modo con cui sono scelti casualmente determinati parametri.

In generale se si è fortunati in tale scelta (specie nella scelta dello scenario di riferimento per la fase costruttiva dell'euristica stessa) si riescono ad avere percentuali di scarto basse tra la soluzione ottima calcolata dal CPLEX e la migliore soluzione trovata dall'euristica nello stesso intervallo di tempo in cui CPLEX ha determinato l'ottimo.

Per quanto riguarda i tempi risolutivi in questo caso, invece, essi variano da pochissimi decimi di secondo per alcune sperimentazioni ai 2 minuti e 32 secondi per la sperimentazione 50.3.50-3.

Di questa prima batteria di sperimentazioni si può dunque dire che la fase di testing è riuscita complessivamente a dimostrare che l'euristica proposta riesce sia in un tempo computazionale di pochissimi decimi di secondo che di oltre 3 ore a determinare per tutte e tre le varianti valori di soluzione che si avvicinano molto all'ottimo determinato dal CPLEX.

Per ciascuna sperimentazione è inoltre singolarmente apprezzabile il fatto che la percentuale di scarto tra la soluzione ottima del CPLEX e la migliore soluzione trovata dall'euristica nello stesso tempo computazionale del CPLEX varia davvero di poco tra una variante del modello e l'altra.

Si passa ora ad analizzare i risultati ottenuti per la “seconda batteria di sperimentazioni”, quella per cui si è imposto sia a CPLEX che all’euristica un tempo computazionale di 15 minuti.

La tabella in figura 5.6, ossia quella riferita ai risultati per la variante “MG_OVER”, mostra nella colonna denominata “Scarto % F.O.” come tale valore si mantenga sempre piuttosto contenuto.

Esso è infatti in questo caso oscillante tra lo 0,05% riferito alla sperimentazione 30.2.200-3 ed il 2,74% per la sperimentazione 60.4.200-3.

Anche la tabella in figura 5.7, ossia quella riferita ai risultati per la variante “MG_OVER+SWAP” per questa seconda batteria di sperimentazioni, mostra nella stessa colonna ancora una volta come tale valore sia sempre piuttosto contenuto.

In questo caso varia infatti tra lo 0,92% della sperimentazione 40.3.200-3 ed il 4,64% della sperimentazione 60.4.200-3.

Infine lo stesso discorso vale anche per la tabella in figura 5.8, ossia quella riferita ai risultati per la variante “MG_OVER+POST” per questa seconda batteria di sperimentazioni.

In questo caso c’è da apprezzare che tale grandezza varia dal -0,05% della variante 40.3.200-3 all’1,88% per la variante 50.3.200-3.

C’è quindi da segnalare che per la variante 40.3.200-3 l’euristica è stata leggermente migliore del CPLEX, avendo determinato una soluzione migliore.

Un’analisi complessiva dei risultati inerenti a tale seconda batteria di sperimentazioni indica che dunque il CPLEX riesce quasi in tutti i casi a determinare una soluzione migliore di quella dell’euristica.

Occorre precisare che la soluzione determinata dal CPLEX in 15 minuti non è con certezza quella ottima, ma è in ogni caso molto vicina all’ottimo poiché per tutte le sperimentazioni il gap indicato tra il valore di migliore soluzione trovata ed il bound determinato dalla soluzione ottima del problema rilassato rimuovendo i vincoli di interezza è davvero contenuto (tali soluzioni si possono definire con una certa semplicità icastica “pseudo-ottime”).

Questo equivale a dire che in alcuni casi di questa seconda batteria di sperimentazioni il valore della migliore soluzione trovata dopo 15 minuti dal CPLEX potrebbe in alcuni casi essere anche il valore ottimo del problema (con certezza si può dire che non lo è senz’altro per quell’istanza in cui nella variante “MG_OVER+POST” l’euristica è stata in grado di trovare un valore di soluzione migliore del CPLEX dopo 15 minuti. Questo

perché non è ovviamente possibile che l'euristica determini una soluzione migliore della soluzione ottima di CPLEX!).

Alla luce di tale considerazione c'è quindi da dire che anche in un tempo computazionale ristretto di 15 minuti e sicuramente congeniale ed accettabile per l'utente che non i tempi richiesti per le sperimentazioni inerenti alla variante "MG_OVER+SWAP" (variabili nella maggior parte dei casi dall'una alle 3 ore ed oltre, come si è già illustrato), l'euristica è dunque capace di determinare buone soluzioni che si avvicinano all'ottimo o comunque, come si è specificato, a soluzioni davvero vicine all'ottimo determinate dal CPLEX.

Si può comunque aggiungere il fatto che aumentando ancor di più la dimensione di tali sperimentazioni, sarà magari possibile mandare il CPLEX in difficoltà (ossia mettendolo dopo 15 minuti non in condizione di determinare una soluzione vicinissima all'ottimo) ed apprezzare il fatto che l'euristica (sempre in 15 minuti) è capace di determinare invece una soluzione migliore rispetto al CPLEX.

In conclusione si può comunque dire, alla luce della fase di testing svolta, che l'euristica è in grado di determinare sia per tempi computazionali relativamente brevi che relativamente lunghi delle soluzioni sempre piuttosto vicine all'ottimo, con una percentuale di scarto rispetto all'ottimo (o al "pseudo-ottimo") che non va mai ad assumere valori inaccettabili per l'utente.

5.5 Approccio stocastico vs approccio deterministico: un "toy example"

Dopo aver riportato i risultati numerici che mettono a confronto le prestazioni del risolutore CPLEX con quelle dell'euristica proposta su sperimentazioni di una certa dimensione, occorre ora osservare in maniera opportuna le differenze di fondo tra un approccio stocastico ed uno di tipo deterministico.

Si è già discusso ampiamente nel corso del secondo capitolo circa l'utilità della PS come strumento in grado di operare al meglio in presenza di incertezza, ma in questo caso ci si avvale di un esempio numerico al fine di scendere maggiormente nel dettaglio.

L'esempio numerico qui riportato è di dimensioni estremamente ridotte e proprio per questo motivo lo si può definire con semplicità icastica un "toy example" (ovvero un vero e proprio "esempio-giocattolo", utilizzato solo per fini dimostrativi).

Tale toy example è una sperimentazione “10.2.3-3”, composta per l’appunto da un contesto di 10 casi totali (di cui 8 elettivi), con 2 sale operatorie disponibili e con 3 specialità chirurgiche differenti.

Si suppone che ogni sala operatoria sia aperta per un tempo regolare giornaliero di 12 ore, le quali vengono suddivise in tanti blocchi di ugual durata pari al numero delle specialità chirurgiche: si hanno perciò in totale 3 blocchi temporali da 4 ore ciascuno.

Vi sono inoltre 3 scenari supposti equiprobabili, mentre il costo orario dell’overtime è posto al solito pari a 100 €/h.

Inoltre per quanto riguarda la variante “MG_OVER+SWAP” si suppone per ogni caso elettivo un costo di swapping pari al 2% del suo valore di priorità, mentre per quanto riguarda invece la variante “MG_OVER+POST” si suppone per ogni caso elettivo un costo di posticipo esattamente pari al suo valore di priorità.

Le durate dei casi sono state stabilite utilizzando una funzione di distribuzione lognormale, mentre i valori di priorità dei casi elettivi sono stati assegnati casualmente dal software:

s	1	2	3
i			
01	0.935505209	0.861188241	0.958034395
02	1.577412266	1.558427639	1.527328549
03	2.311794047	1.883761911	1.947870284
04	2.437665600	1.990518041	2.226782464
05	0.883928367	0.845762897	1.229504249
06	0.862632481	1.181053403	0.930763419
07	1.943626649	2.025773352	2.653218311
08	1.179202625	2.073078987	1.554944077
09	3.651561038	2.273579086	
10	2.260378633	2.696326352	2.323319812

Figura 5.9: durate dei casi sotto ogni scenario per il “toy example” 10.2.3-3.

ie	
01	200
02	400
03	500
04	500
05	200
06	200
07	500
08	300

Figura 5.10: priorità dei casi elettivi per il “toy example” 10.2.3-3.

A questo punto l'esempio è stato risolto con CPLEX per ciascuna delle tre varianti. I risultati ottenuti sono i seguenti, visualizzabili mediante una rappresentazione grafica di semplice utilizzo:

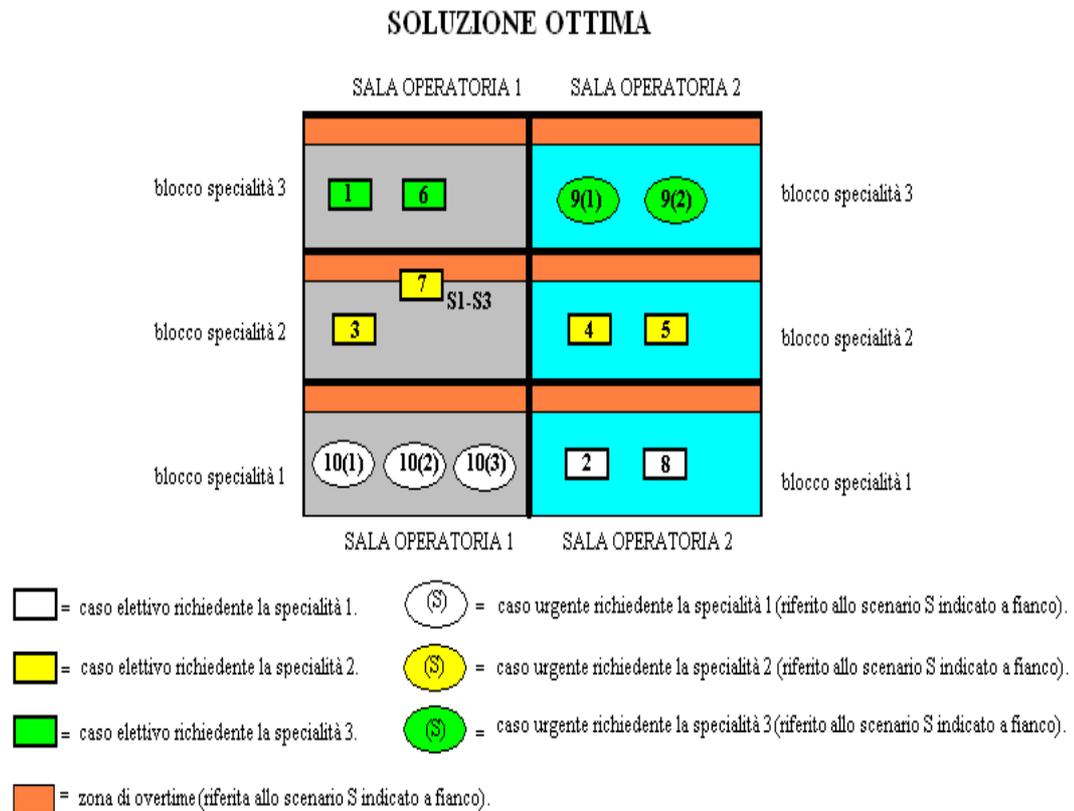


Figura 5.11: risultati per la versione stocastica del toy example (variante MG_OVER).

Tale istanza del modello è composta da 53 variabili e da 32 vincoli ed ha dato un valore ottimo della funzione obiettivo pari a **28,55 €**

La soluzione rappresentata in figura sovrastante mostra subito come tutti i casi elettivi sono stati correttamente schedulati in qualcuna delle due sale operatorie, così come i casi urgenti sotto tutti gli scenari in cui si vengono a presentare.

C'è da segnalare che il solo caso elettivo 7, richiedente la specialità chirurgica 2, finisce per una piccola porzione della sua durata complessiva in overtime sia sotto il primo che sotto il terzo scenario.

A questo punto si può passare pertanto a presentare i risultati ottenuti risolvendo la variante "MG_OVER+SWAP" del modello stesso:

SOLUZIONE OTTIMA

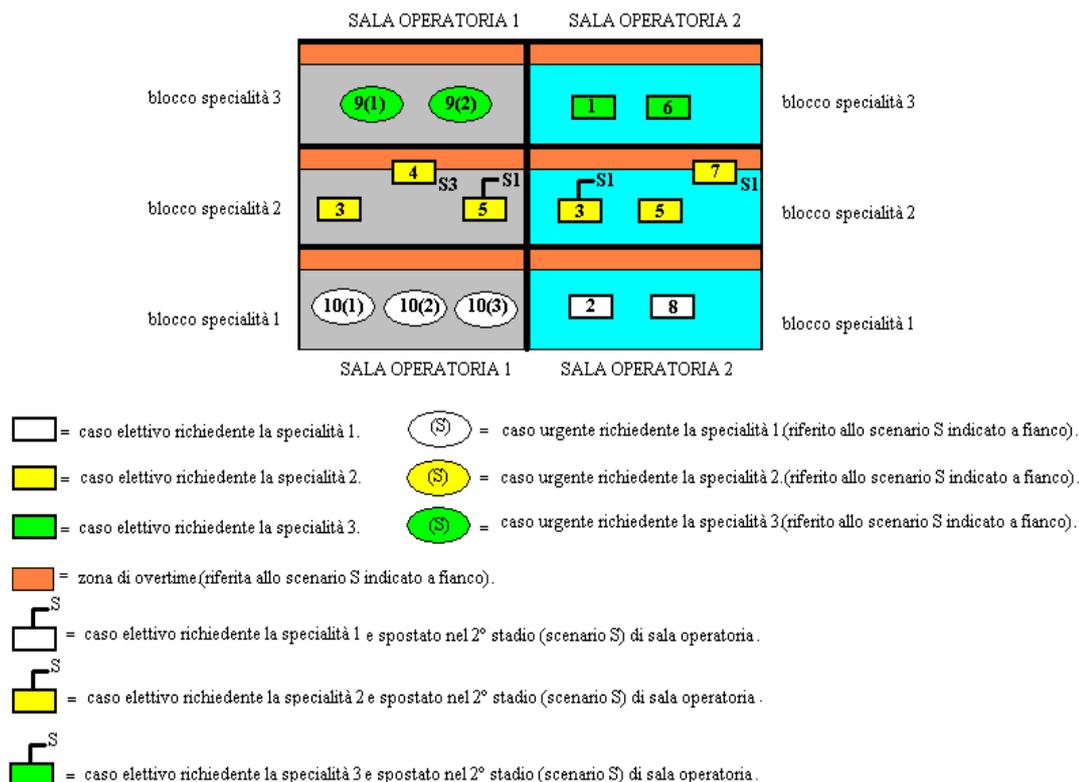


Figura 5.12: risultati per la versione stocastica del toy example (variante MG_OVER+SWAP).

Tale istanza del modello è composta da 101 variabili e da 80 vincoli ed ha dato un valore ottimo della funzione obiettivo pari a **19,00 €**

Anche stavolta la soluzione rappresentata in figura sovrastante mostra subito come tutti i casi elettivi sono stati correttamente schedulati in qualcuna delle due sale operatorie, così come i casi urgenti sotto tutti gli scenari in cui si vengono a presentare.

C'è da segnalare che il solo caso elettivo 7, richiedente la specialità chirurgica 2, finisce per una piccola porzione della sua durata complessiva in overtime soltanto sotto il primo scenario.

Inoltre si hanno due operazioni di swapping: una prevede che sotto il primo scenario il caso elettivo 3, richiedente la specialità chirurgica 2 e schedulato dal 1° stadio nella sala operatoria 1, venga spostato nella sala operatoria 2.

L'altra invece prevede che sotto il primo scenario il caso elettivo 5, anch'esso richiedente la specialità chirurgica 2 e schedulato dal 1° stadio nella sala operatoria 2, venga spostato nella sala operatoria 1.

I risultati ottenuti invece per la variante “MG_OVER+POST” del modello sono i seguenti:

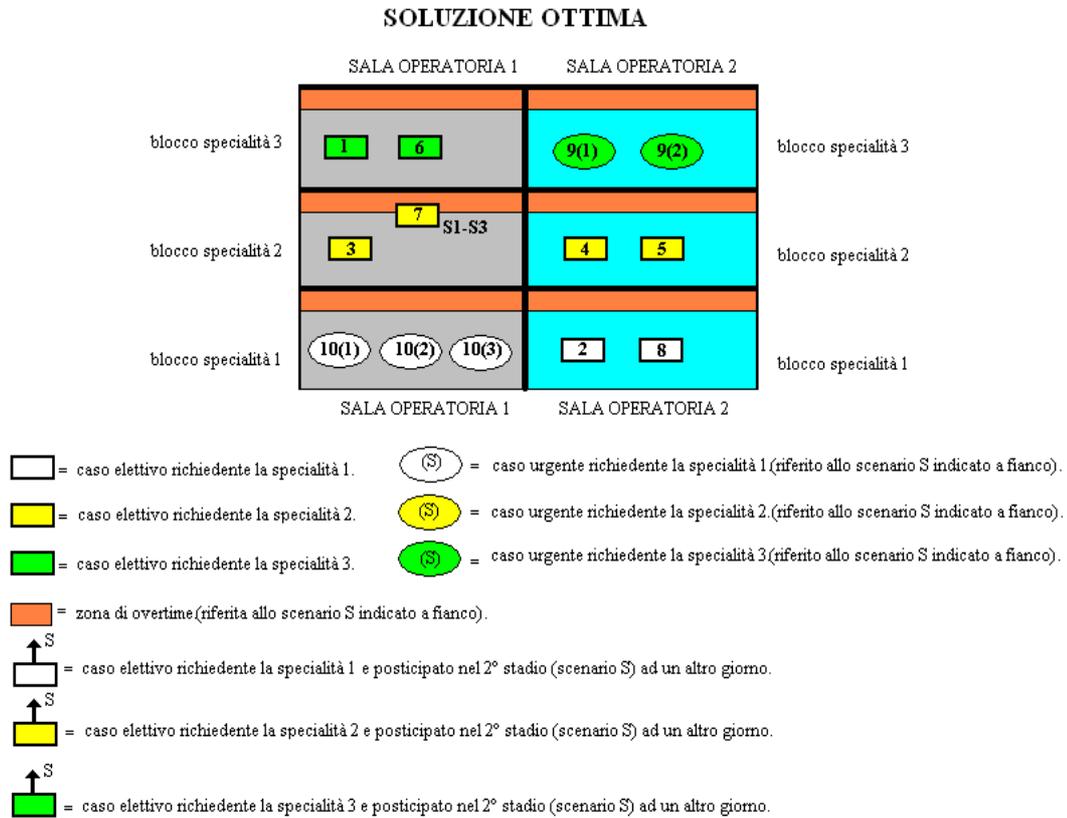


Figura 5.13: risultati per la versione stocastica del toy example (variante MG_OVER+POST).

Tale istanza del modello è composta da 77 variabili e da 80 vincoli ed ha dato un valore ottimo della funzione obiettivo pari a **28,55 €**

La soluzione rappresentata in figura sovrastante mostra subito come essa sia esattamente pari a quella della variante “MG_OVER”, dato che non c’è in questo caso necessità di introdurre nel 2° stadio nessuna operazione di posticipo per i casi elettivi.

A questo punto, dopo aver presentato i risultati ottenuti risolvendo il toy example per ognuna delle tre varianti del modello, si è passato ad istanziare una versione deterministica equivalente dell’esempio stesso.

Tale versione deterministica elimina chiaramente il fattore dell’incertezza e lo fa ponendo pari al loro valor medio i parametri iniziali caratterizzati dall’incertezza.

Il toy example di riferimento è stato pertanto ridotto ad un solo scenario (che si realizzerà dunque con certezza, eliminando così la stocasticità nel modello) e presenta come durata per ciascuno dei casi il suo valor medio (tutti gli altri dati iniziali sono chiaramente rimasti immutati):

s	i	1
01		0.918
02		1.554
03		2.048
04		2.219
05		0.987
06		0.992
07		2.208
08		1.602
09		2.769
10		2.426

Figura 5.14: durate dei casi per il “toy example” in versione deterministica.

Ed anche in questo caso, tale toy example è stato risolto con CPLEX per ciascuna delle tre varianti.

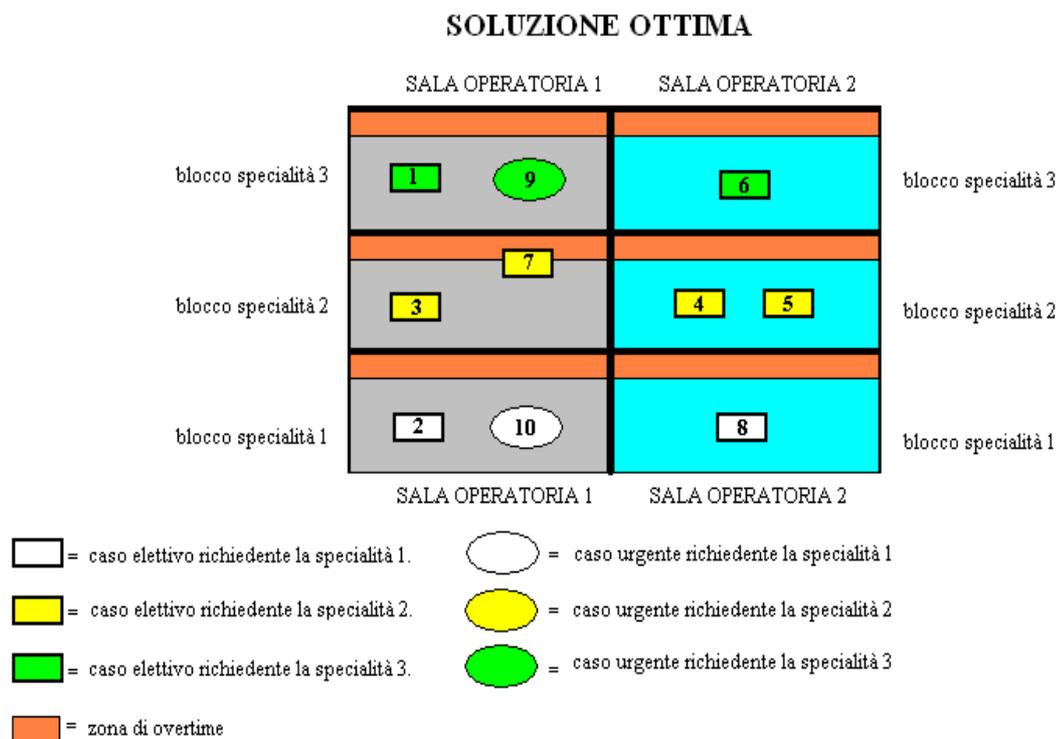


Figura 5.15: risultati per la versione deterministica del toy example (variante MG_OVER).

Tale istanza del modello è composta da 35 variabili e da 17 vincoli ed ha dato un valore ottimo della funzione obiettivo pari a **25,60 €**

La figura dimostra come tutti i casi, elettivi od urgenti che siano, vengono schedulati in qualcuna delle due sale operatorie e che il solo caso elettivo 7 ricade per una piccola porzione della sua durata in overtime.

Per la variante “MG_OVER+SWAP” si è ottenuto invece, in questo caso:

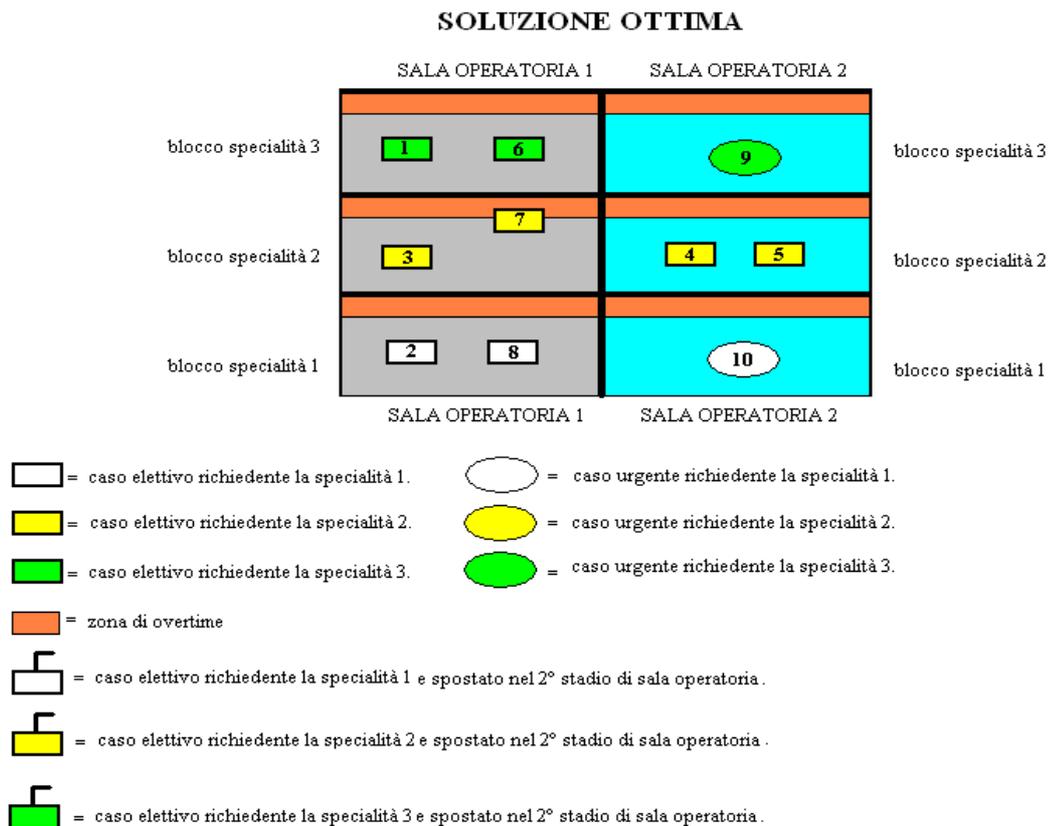


Figura 5.16: risultati per la versione deterministica del toy example (variante MG_OVER+SWAP).

Tale istanza del modello è composta da 51 variabili e da 33 vincoli ed ha dato un valore ottimo della funzione pari a **25,60 €**

La figura dimostra come tutti i casi, elettivi od urgenti che siano, vengono schedulati anche stavolta in qualcuna delle due sale operatorie e che il solo caso elettivo 7 ricade per una piccola porzione della sua durata in overtime.

Non a caso il valore ottimo della funzione obiettivo è uguale a quello ottenuto per la variante “MG_OVER”: l’unica cosa che varia tra le due soluzioni è che alcuni casi vengono schedulati in una sala operatoria piuttosto che in un’altra.

Ad esempio si possono prendere i casi 8 e 10, entrambi richiedenti la specialità chirurgica 3: nella soluzione della variante “MG_OVER” il caso 8 viene schedulato nella sala operatoria 2 ed il caso 10 viene schedulato nella sala operatoria 1, mentre in questa soluzione della variante “MG_OVER+SWAP” accade l’esatto contrario.

Infine per la variante “MG_OVER+POST” si è ottenuto invece, in questo caso:

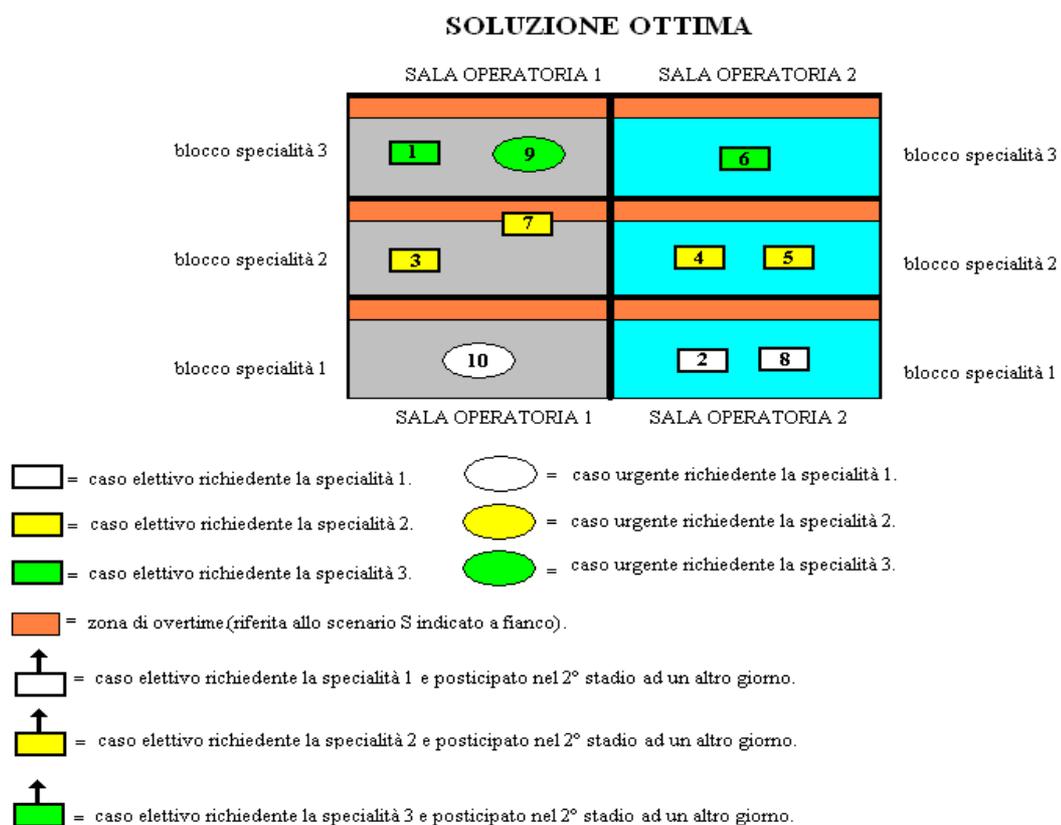


Figura 5.17: risultati per la versione deterministica del toy example (variante MG_OVER+POST).

Tale istanza del modello è composta da 43 variabili e da 33 vincoli ed ha dato un valore ottimo della funzione pari a **25,60 €**

Al pari delle soluzioni ottenute per le altre due varianti del modello, tutti i casi, elettivi od urgenti che siano, vengono schedulati anche stavolta in qualcuna delle due sale operatorie e che il solo caso elettivo 7 ricade per una piccola porzione della sua durata in overtime.

Allo stesso modo il valore ottimo della funzione obiettivo è uguale, mentre ciò che va a variare rispetto alle soluzioni delle altre due varianti è la sala operatoria in cui si decide di schedare qualche caso.

Ad esempio si può prendere il caso elettivo 2, richiedente la specialità chirurgica 3, il quale mentre nelle soluzioni delle altre due varianti veniva schedato nella sala operatoria 1, viene in questo caso schedato nella sala operatoria 2.

Analizzando complessivamente i risultati ottenuti, c'è da sottolineare innanzitutto che i valori di soluzione ottenuti per le varianti "MG_OVER" ed "MG_OVER+POST", risolvendo sia il modello in versione stocastica che in versione deterministica equivalente (ottenuta ponendo pari al loro valor medio i parametri incerti), vanno a mostrare come in questo caso siano leggermente minori per la versione deterministica.

In entrambi i casi, infatti, la funzione obiettivo è minore di **2,95 €** rispetto alla soluzione ottenuta per la versione stocastica.

Questo vuol dire che per questo toy example presentato l'approccio deterministico equivalente si fa leggermente preferire per tali due varianti, ma si può dire che in termini generali si tratta di una casualità dovuta alle dimensioni estremamente ridotte di questo esempio riportato nonché alla struttura dei dati iniziali dell'esempio stesso.

Al crescere delle dimensioni e della complessità computazionale del problema da risolvere, infatti, è veramente assai difficile pensare che ciò possa accadere, così come a volte è persino difficile ottenere per la versione deterministica equivalente (istanziata ponendo pari al loro valor medio i parametri incerti) delle soluzioni ottime che siano ammissibili per la versione stocastica.

Difatti lo dimostrano già i risultati ottenuti per la variante "MG_OVER+SWAP", in cui rispetto alle altre due varianti il numero totale di variabili e di vincoli del problema è maggiore e non a caso l'approccio stocastico si fa stavolta preferire.

La funzione obiettivo all'ottimo assume infatti in questo caso un valore minore di **6,60 €** rispetto a quella ottenuta risolvendo all'ottimo il problema deterministico equivalente (istanziato al solito ponendo pari al loro valor medio i parametri incerti).

Conclusioni e sviluppi futuri

L'obiettivo di questo lavoro di tesi è stato quello di introdurre alcune delle problematiche di natura gestionale connesse ad una fase di pianificazione, di gestione e di controllo di una sala operatoria.

Di immediata comprensione, alla luce delle varie problematiche illustrate nel corso del primo capitolo, è dunque il fatto che l'introduzione di uno strumento quantitativo di supporto alle decisioni ed in grado dunque di coadiuvare la fase progettuale, può assumere in questo contesto un ruolo di fondamentale importanza.

E tutto ciò assume ancor più significato se si considera che le statistiche informano che in un sistema sanitario le sale operatorie rappresentano proprio il centro di costo di maggior incidenza.

Inoltre, a rafforzare ulteriormente tale tesi, c'è da dire che la mancanza di una gestione operativa efficace ed efficiente è sempre più, purtroppo, nel corso degli ultimi anni oggetto di polemica e causa della sfiducia dell'opinione pubblica nei confronti della sanità, specialmente in Italia.

A tal uopo sono stati introdotti, nel corso del secondo capitolo, alcuni modelli di ottimizzazione presenti in letteratura e formulati con l'obiettivo di operare su svariati parametri al fine di rendere il blocco operatorio il più efficiente possibile.

Tali modelli presenti in letteratura sono comunque caratterizzati da diversi limiti ed assunzioni a volte troppo semplicistiche, ragion per cui comunque la tipologia d'approccio più appropriata per un contesto decisionale siffatto, e comunque fortemente condizionato dalla presenza dell'incertezza decisionale, si è manifesta essere la Programmazione Stocastica.

Nel corso di questo percorso triennale di dottorato di ricerca si è così studiato come proporre tale strumento in maniera più adeguata e più consona possibile alle esigenze ed alle problematiche riscontrate per questo sistema.

Prendendo sempre come riferimento l'approccio stocastico, si è partiti così da un modello "monosala" con orizzonte di pianificazione settimanale.

Tale modello "monosala" è stato poi esteso ad un modello multisala, ma si è mantenuto l'orizzonte di pianificazione settimanale.

A questo punto il modello è stato variato, divenendo con orizzonte pianificazione giornaliero e con altre varie assunzioni e condizioni al contorno chiaramente differenti.

Tale modello “multisala” con orizzonte di pianificazione giornaliero è stato quindi ulteriormente modificato ed infine articolato in tre varianti differenti, componendo così la versione definitiva presentata in questo tomo.

Il modello di ottimizzazione che è oggetto di questo lavoro di tesi, presentato nel terzo capitolo, è pertanto di natura stocastica (modello di MIP con ricorso a due stadi) ed è in grado nelle sue suddette tre varianti di fornire una risposta andando a definire i parametri di calibratura.

Esso si differenzia per vari aspetti dagli altri modelli inerenti alla gestione operativa del blocco operatorio, e di cui si è parlato nel secondo capitolo: nessuno tra essi infatti propone una vasta gamma di opzioni possibili, al fine di definire uno scheduling giornaliero che minimizzi i costi straordinari come il modello presentato.

Al suo interno infatti si considerano singolarmente, in ognuna delle sue varianti, tre differenti strategie che il decisore può scegliere di applicare: ricorrere ai soli tempi di overtime (variante “MG_OVER”), ricorrere oltre che ai tempi di overtime alla possibilità di spostare di sala operatoria nel 2° stadio un caso elettivo (variante “MG_OVER+SWAP”) e ricorrere oltre che ai tempi di overtime alla possibilità di posticipare un caso elettivo nel 2° stadio (variante “MG_OVER+POST”).

Il grado di incertezza decisionale è altresì espresso in maniera soddisfacente, cosa che non è sempre possibile riscontrare negli altri modelli presenti in letteratura.

Infatti in taluni modelli suddetti la presenza dei casi urgenti, che rappresentano di fatto un aspetto fondamentale di un sistema siffatto nonché quello caratterizzato da maggior incertezza decisionale, per semplicità strutturale non viene considerata o si fanno assunzioni che nelle situazioni reali non sono particolarmente accettabili.

Tale modello è comunque spesso caratterizzato da un’elevata complessità computazionale che ne potrebbe richiedere un tempo risolutivo estremamente elevato per alcune istanze di grandi dimensioni di ciascuna delle tre varianti suddette.

Per tale ragione è stata introdotta e presentata nel corso del quarto capitolo una metodologia euristica in grado di offrire così maggior flessibilità per quanto riguarda gli approcci risolutivi.

Tale euristica, come hanno dimostrato numericamente i risultati della fase di testing presentati nel corso del quinto capitolo, è in grado di calcolare “buone soluzioni” anche in un tempo computazionale ragionevolmente accettabile per l’utente.

Infatti poiché il modello si propone di definire uno scheduling a livello giornaliero non avrebbe granché utilità offrire un modello efficace ma inefficiente in quanto richiede troppo tempo per fornire la soluzione ottima.

Proprio per questo motivo ci si può accontentare di soluzioni un po' meno efficaci (in quanto non ottime, ma che comunque è dimostrato che non si discostano di molto da esse) ma ottenute quantomeno in un tempo computazionale ragionevole.

Sviluppi futuri potrebbero prevederne l'estensione del modello alla fase "post-operatoria", nella quale si andrebbe anche a considerare la gestione operativa dei posti letto nelle cosiddette "Recovery Room" (RR).

Infatti a seguito di un generico intervento chirurgico particolarmente delicato un paziente può richiedere la necessità dopo aver subito un tale intervento di soggiornare per un certo periodo di tempo in una RR dove riceverà tutte le cure del caso per riabilitarsi e recuperare dai postumi dell'intervento medesimo.

Spesso tali RR sono caratterizzate infatti dalla scarsità dei posti-letto, nonché da un elevato costo di mantenimento e di gestione: per tali ragioni assumerebbe dunque un certo ruolo di interesse estendere il modello presentato in questo lavoro di tesi a considerare contestualmente anche l'aspetto suddetto fornendone anche per essa un'adeguata gestione operativa.

Bibliografia

- **Citazioni:**

[1] = cit. HFMA, "Achieving operating room efficiency through process integration". Technical report, Health Care Financial Management Association Report, 2005.

[2] = cit. M.S. Karpel, "Using patient classifications systems to identify ambulatory care costs", Health Care Financial Management, 1994; cit. D.N.Pham and A.Klinkert, "Surgical case scheduling as a generalized job shop scheduling problem", European Journal of Operational Research, vol. 185, 2008, pagg. 1011-1025.

[3] = cit. M.Van Houdenhoven, J. M. van Oostrum, G. Wullink, E. W. Hans, J. L. Hurink, J. Bakker, and G. Kazemier, "Fewer intensive care unit refusals and a higher capacity utilization by using a cyclic surgical case schedule", Journal of critical care, num. 2, vol. 22, 2008, pagg. 222-226.

[4] = cit. R.Tersigni, "Sicurezza in sala operatoria", Roma, 2008.

[5] = cit. P.L.Deriu, "La cultura della sicurezza nella sala operatoria" (abstract book), Torino, 2008.

[6] = cit. regolamento conforme allo schema generale di riferimento della "Carta dei servizi pubblici e sanitari" di cui al D.P.C.M. del 19 maggio 1995.

[7] = cit. D.W.Hubbard, "How to Measure Anything: Finding the Value of Intangibles in Business", 2nd edition, John Wiley & Sons, 2007.

[8] = cit. P.Patterson, "What makes a well-oiled scheduling system", Journal of OR Manager, num. 9, vol. 12, 1996, pagg. 19-23.

[9] = cit. J.T.Blake, F.Dexter and J.Donald, "Operating room managers use of integer programming for assigning block time to surgical groups: a case of study", Anesthesia Analgesia, vol. 94, 2002, pagg. 143-148.

[10] = cit. F.Dexter, “A strategy to decide whether to move the last case of the day in an operating room to other empty operating room to decrease overtime labor costs”, International Anesthesia Research Society, vol. 91, 2000, pagg. 925-928.

[11] = cit. R.A.Gabel, B.S.Lee and J.C.Kulli, “Operating room management”, Butterworth-Heinemann, 1999.

[13] = cit. M. Kennedy. “Bin-Packing, Knapsack and Chance-Constrained Approaches to Operating Room Scheduling”, PhD thesis, Rensselaer Polytechnic Institute School Of Engineering Department of Decision Sciences Engineering Systems, 1992; cit. R.E.Wachtel and F.Dexter, “Tactical increases in operating room block time for capacity planning should not be based on utilization”, Anesthesia Analgesia, num. 1, vol. 106, 2008, pagg. 215-226.

- **Articoli scientifici:**

[12] = M.Van der Lans, E.W.Hans, J.L.Hurink, G.Wullink, M.Van Houdenhoven, G.Kazemier, “Anticipating urgent surgery in operating room departments”, Working Paper, University of Twente, Twente (Olanda), 2007.

[14]: I.Ozkarahan, “Allocation of surgeries to operating rooms by goal programming”, Journal of medical system, num. 6, vol. 24, 2000, pagg. 339-378.

[15]: A.Guinet e S.Chaabane, “Operating theatre planning”, Int. J. Production Economics, num. 1, vol. 85, 2003, pagg. 69–81.

[16]: A.Jebali, A.B.H.Alouane e P.Ladet, “Operating rooms scheduling”, Int. J. Production Economics, vol. 99, 2006, pagg. 52–62.

[17]: P.Santibanez, M.Begen e D.Atkins, “Surgical block scheduling in a system of hospitals: an application to resource and wait list management in a British Columbia health authority”, Health Care Manage Sci, vol. 10, 2007, pagg. 269–282.

- [18]: A.Testi, E.Tanfani e G.Torre, "A three-phase approach for operating theatre schedules", *Health Care Manage Sci*, vol. 10, 2007, pagg. 163–172.
- [19]: J.P.M.Arnaout e S.Kulbashian, "Maximizing the utilization of operating room with stochastic times using simulation", *Proceedings of the 2008 Winter Simulation Conference*, S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler eds.
- [20]: H.Fei, C.Chu e N.Meskens, "Solving a tactical operating room planning problem by a column-generation-based heuristic procedure with four criteria", *Ann Operational Research*, vol. 166, 2009, pagg. 91-108.
- [21]: B.Cardoen, "Operating room planning and scheduling: solving a surgical case sequencing problem", PhD thesis, Katholieke Universiteit Leuven, Leuven (Belgium), 2009.
- [22]: J.M.Molina e J.Framinam, "Testing planning policies for solving elective case scheduling phase: a real application", University of Sevilla, School of Engineering, Siviglia (Spagna), 2009.
- [23]: M.L.Pinedo, "Planning and Scheduling in Manufacturing and Services - Planning and Scheduling in health care", *Springer Series in Operations Research*, vol. 1, 2009.
- [24]: B.Zhang, P.Murali, M.Dessouky e D.Belson, "A mixed integer programming approach for allocating operating room capacity", University of Southern California, Epstein Department of Industrial and Systems Engineering, Los Angeles (USA), 2009.
- [25]: Y.Liu, C.Chu, K.Wang, "A new heuristic algorithm for the operating room scheduling problem", *Computers & Industrial Engineering*, vol. 61, 2011, pagg. 865-871.
- [26]: B.Denton, D.Gupta, "A sequential bounding approach for optimal appointment scheduling", *IIE Transaction*, vol. 35, 2003, pagg. 1003-1016.
- [27]: J.Belien, E.Demeulemeester, "Building cyclic master surgery schedules with leveled resulting bed occupancy", *European Journal of Operational Research*, vol. 176, 2005, pagg. 1185–1204.

- [28]: M.Lamiri, X.Xie, A.Dolgui e F.Grimaud, "A stochastic model for operating room planning with elective and emergency demand for surgery", *European Journal of Operational Research*, vol. 185, 2006, pagg. 1026–1037.
- [29]: E.Hans, G.Wullink, M.Van Houdenhoven, G.Kazemier, "Robust surgery loading", *European Journal of Operational Research*, vol. 185, 2006, pagg. 1038–1050.
- [30]: B.Denton, J.Viapiano, A.Vogl, "Optimization of surgery sequencing and scheduling decisions under uncertainty", *Health Care Manage Sci*, vol. 10, 2007, 1013-24.
- [31]: B.Denton, A.Miller, H.J.Balasubramanian, T.R.Huschka, "Optimal Allocation of Surgery Blocks to Operating Rooms Under Uncertainty", *Operations Research*, vol. 58, 2010, pagg. 802-816.
- [32]: C.Mancilla e R.H.Storer, "Stochastic Sequencing and Scheduling of an Operating Room", PhD thesis, Lehigh University, Department of Industrial and Systems Engineering, Bethlehem (USA), 2009.
- [33]: M.Lamiri, F.Grimaud e X.Xie, "Optimization methods for a stochastic surgery planning problem", *Int. J.Production Economics*, vol. 120, 2009, pagg. 400–410.
- [34]: S.Batun, B.Denton, T.Huschka, A.J.Schaefer, "Operating Room Pooling and Parallel Surgery Processing Under Uncertainty", *Inform Journal on Computing*, num. 2, vol. 23, 2011, pagg. 220-237.
- [35]: R.Van Slyke e R.B.Wets, "L-Shaped linear program with application to optimal control and stochastic linear programming", *SIAM Journal and Applied Mathematics*, vol. 17, 1969, pagg. 638-663.
- [36]: R.B.Wets e R.T.Rockafellar, "Scenario and policy aggregation in optimization under uncertainty", IISA Working Paper WP-87-119, Austria, 1987.

[37]: R.Birge and D.F.Holmes, "Efficient solution of two-stage stochastic linear programs using interior point methods", Computational Optimization and Applications, vol. 1, 1992, pagg. 245-276.

F.Guerriero, R. Guido, "The Operational Research in the Operating Theatre Management: a survey", Health Care Management Science, num. 1, vol. 14, pagg. 89-114.

B.Cardoen, E.Demeulemeester and J.Belien, "Operating room planning and scheduling: a literature review", European Journal of Operational Research, num. 3, vol. 201, pagg. 921-932.

- **Libri di testo:**

R.Birge, F.Loveaux, "Introduction to stochastic programming", Springer Verlag, New York , 1997.

- **Siti web:**

<http://it.wikipedia.org/>

<http://periodici.caspur.it/>

Ringraziamenti

Una dedica per questo lavoro di tesi a **Giorgio, mio padre**: un uomo esemplare, umile, onesto e laborioso che ha sempre creduto in me e che mi ha sempre sostenuto in silenzio, unendo i suoi sacrifici ai miei... A lui va naturalmente anche un ringraziamento particolare, speciale ed affettuoso...

Un ringraziamento caloroso e speciale a **mia madre**, ai **miei fratelli** ed a tutte le altre persone che mi sono state vicine e che si sono rese disponibili, anche con una semplice frase o con un sorriso, in questi anni, sia moralmente che professionalmente...

Un ringraziamento pieno di riconoscimento e di stima nei confronti dei miei **relatori** che mi hanno seguito con cura, con cortesia e con professionalità durante questi tre anni di percorso...

Un pensiero affettuoso con un dolce ed imperituro ricordo ai miei **nonni paterni**, Domenico e Vincenza...

“Il successo nella vita non è qualcosa che devi rincorrere con le tue forze fisiche, ma è qualcosa che devi attirare a te con le tue capacità, la tua umiltà e con la buona volontà di sfidare ogni giorno i tuoi sogni, fino a renderli realtà... E finchè lo fai credendoci, non darti mai per vinto per nessun motivo...” (**Domenico**).