

Dissertação

Mestrado em Gestão de Sistemas de Informação Médica

***Análises clínicas compostas:
um estudo crítico via simulação***

Ana Granado

Leiria, setembro de 2014



Dissertação

Mestrado em Gestão de Sistemas de Informação Médica

***Análises clínicas compostas:
um estudo crítico via simulação***

Ana Granado

Dissertação de Mestrado realizada sob a orientação do Doutor Rui Santos, Professor da Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria.

Leiria, setembro de 2014

Agradecimentos

Expresso o meu agradecimento ao meu orientar Professor Doutor Rui Santos. Este trabalho foi conseguido com o seu grande apoio, dedicação, disponibilidade e tolerância. Destaco o seu empenho, a capacidade de comunicação e de transmissão de conhecimento, o que tornou a realização deste projeto mais motivador.

Aos meus pais, Domingos e Angelina, obrigada por tudo.

Resumo

Análises clínicas compostas podem ser aplicadas com o objetivo de economizar recursos quando se pretende classificar uma população (identificar todos os indivíduos infectados) na qual a taxa de prevalência é diminuta, apesar de reduzir a fiabilidade da classificação. Neste sentido, o presente trabalho tem como objetivo comparar o desempenho de várias metodologias de classificação (ensaios individuais, metodologia de Dorfman, algoritmos hierárquicos e testes baseados em *arrays* com e sem *master pool*), nomeadamente o custo relativo (número esperado de testes para a classificação de cada indivíduo) e a probabilidade de existência de erros de classificação (medida pela especificidade e pela sensibilidade de cada metodologia). Assim, as usuais técnicas de simulação (realizadas recorrendo ao *software* estatístico R) foram aplicadas a populações com distintos cenários, usando diferentes taxas de prevalência, várias dimensões do grupo, bem como diversos níveis de sensibilidade e de especificidade. Ao longo deste trabalho foi assumido que o agrupamento dos sangues (criação do sangue composto) não afeta a probabilidade de má classificação (ausência do efeito de diluição), como é comprovado por muitas análises qualitativas (presença ou ausência da infeção). As simulações realizadas mostram que os testes compostos só podem ser recomendados em casos com baixas taxas de prevalência e baixas probabilidades de erros de classificação, sendo possível identificar a metodologia mais adequada para cada caso em função da sua taxa de prevalência, sensibilidade e especificidade. Além disso, sempre que a taxa de prevalência, a sensibilidade e a especificidade são conhecidos (ou, pelo menos, estimativas razoáveis estão disponíveis), simulações podem ser realizadas para identificar a metodologia mais adequada, e, deste modo, encontrar um ponto de equilíbrio entre o custo e a fiabilidade da classificação.

Palavras-chave: testes compostos, taxa de prevalência, erros de classificação, custo relativo, simulação.

Abstract

Compound clinical trials can be applied in order to save resources for classification purposes (identification of all infected individuals in a population) whenever low prevalence rates exist, although it decreases accuracy. Our goal is to compare the performance of several classification methodologies (individual trials, hierarchical algorithms, and *array*-based group testing with or without master pool), mainly their relative cost (expected number of tests for the classification of each individual) and their probability of misclassification (measured by the specificity and sensitivity of each methodology). Thus, simulation techniques (performed using Statistical Software R) were employed under different potential population scenarios, using multiple prevalence rates, several group dimensions, diverse sensitivities and specificities. Through this framework it was assumed that pooling does not affect misclassification probability (absence of the dilution effect), as it is corroborated by many qualitative analyses (presence or absence of the infection). The performed simulations show that compound tests can only be advised in cases with low prevalence rates and low probabilities of misclassification. Furthermore, it is impossible to easily identify the most suitable methodology to each case, due its dependence on the prevalence rate, on the sensitivity, and on the specificity. Nevertheless, compound tests allow saving resources (by reducing the number of required tests) and increase the probability of misclassification (mainly decreasing the sensitivity). Moreover, whenever the prevalence rate, the sensitivity, and the specificity are known (or at least reasonable estimates are available), simulations can be performed in order to identify the most appropriate methodology, and thereby to balance the cost with the accuracy.

Keywords: compound tests, prevalence rate, misclassification, relative cost, simulation.

Índice de Figuras

Figura 1: Realização de um teste composto	8
Figura 2: Evolução da proporção de ocorrência de face nacional em 20 sequências de 20000 lançamentos de uma moeda ao ar.....	19
Figura 3: Metodologia baseada em <i>arrays</i> com um indivíduo infetado	41
Figura 4: Metodologia baseada em <i>arrays</i> com dois indivíduos infetado	41

Índice de Quadros

Quadro 1: Diferentes tipos de classificação com base na relação entre o resultado dos testes e o estado dos indivíduos.....	13
Quadro 2: Custos relativos obtidos na simulação da metodologia de Dorfman sem erros de classificação.....	26
Quadro 3: Valores ótimos utilizando a metodologia de Dorfman	28
Quadro 4: Eficiência e fiabilidade obtidas na simulação da metodologia de Dorfman considerando erros de classificação de 5 por cento.....	30
Quadro 5: Valores ótimos, com erros de classificação de 5%, utilizando a metodologia de Dorfman	32
Quadro 6: Custos relativos obtidos na simulação da metodologia hierárquica sem erros de classificação.....	35
Quadro 7: Valores ótimos utilizando a metodologia hierárquica	36
Quadro 8: Eficiência e fiabilidade obtidas na simulação da metodologia hierárquica considerando erros de classificação de 5 por cento.....	38
Quadro 9: Valores ótimos, com erros de classificação de 5%, utilizando a metodologia hierárquica	39
Quadro 10: Custos relativos obtidos na simulação da metodologia baseada em <i>arrays</i> sem master pool nem erros de classificação	43
Quadro 11: Valores ótimos utilizando a metodologia baseada em <i>arrays</i> sem master pool.....	44
Quadro 12: Eficiência e fiabilidade obtidas na simulação da metodologia baseada em <i>arrays</i> sem master pool considerando erros de classificação de 5 por cento.....	46
Quadro 13: Valores ótimos, com erros de classificação de 5%, utilizando a metodologia baseada em <i>arrays</i>	47
Quadro 14: Custos relativos obtidos na simulação da metodologia baseada em <i>arrays</i> com master pool sem erros de classificação.....	50
Quadro 15: Valores ótimos utilizando a metodologia baseada em <i>arrays</i> com master pool	52
Quadro 16: Eficiência e fiabilidade obtidas na simulação da metodologia baseada em <i>arrays</i> com <i>master pool</i> considerando erros de classificação de 5 por cento.....	53
Quadro 17: Valores ótimos, com erros de classificação de 5%, utilizando a metodologia baseada em <i>arrays</i> com <i>master pool</i>	55
Quadro 18: Simulação sem erros de classificação: Especificidade e Sensibilidade a 100%	58
Quadro 19: Simulação sem erros de classificação: Especificidade e Sensibilidade a 99%.....	60
Quadro 20: Simulação sem erros de classificação: Especificidade e Sensibilidade a 95%.....	61

Lista de Siglas

Siglas	Designação
VIH	Vírus da imunodeficiência humana
N	Dimensão da população
n	Dimensão do grupo
p	Taxa de prevalência
E	Especificidade
S	Sensibilidade
EC	Especificidade composta
SC	Sensibilidade composta
EM	Especificidade da metodologia de classificação
SM	Sensibilidade da metodologia de classificação
CR	Custo relativo
$M1$	Metodologia de Dorfman
$M2$	Metodologia hierárquica (divisão dos grupos positivos em dois grupos e, se positivos, testes individuais)
$M3$	Metodologia baseada em <i>arrays</i> (recorrendo a matrizes quadradas) sem <i>master pool</i>
$M4$	Metodologia baseada em <i>arrays</i> (recorrendo a matrizes quadradas) com <i>master pool</i>

Índice

AGRADECIMENTOS	I
RESUMO	III
ABSTRACT	V
ÍNDICE DE FIGURAS	VII
ÍNDICE DE QUADROS	IX
LISTA DE SIGLAS	XI
ÍNDICE.....	XIII
1 – INTRODUÇÃO	1
2 – NOÇÕES FUNDAMENTAIS DE TESTES CLÍNICOS	7
2.1 – TESTES INDIVIDUAIS <i>VERSUS</i> TESTES COMPOSTOS.....	8
2.2 – TESTES QUALITATIVOS <i>VERSUS</i> TESTES QUANTITATIVOS	10
2.3 – ERROS DE CLASSIFICAÇÃO E MEDIDAS DE FIABILIDADE DOS TESTES	12
2.4 – CUSTO RELATIVO <i>VERSUS</i> EFICIÊNCIA.....	17
2.5 – A UTILIZAÇÃO DA SIMULAÇÃO NA ANÁLISE DA PERFORMANCE DE TESTES CLÍNICOS	18
3 – METODOLOGIAS DE CLASSIFICAÇÃO	23
3.1 – <i>M1</i> – METODOLOGIA DE DORFMAN.....	24
3.2 – <i>M2</i> – METODOLOGIAS HIERÁRQUICAS	33
3.3 – <i>M3</i> – METODOLOGIAS BASEADAS EM ARRAYS SEM <i>MASTER POOL</i>	40
3.4 – <i>M4</i> – METODOLOGIAS BASEADAS EM ARRAYS COM <i>MASTER POOL</i>	48
4 – COMPARAÇÃO E DISCUSSÃO DE RESULTADOS.....	57
4.1 – SIMULAÇÃO SEM ERROS DE CLASSIFICAÇÃO	58
4.2 – SIMULAÇÃO COM ERROS DE CLASSIFICAÇÃO A 99 POR CENTO	59
4.3 – SIMULAÇÃO COM ERROS DE CLASSIFICAÇÃO A 95 POR CENTO	60
5 – CONCLUSÃO	63
BIBLIOGRAFIA	67
ANEXOS	71

1 – Introdução

Dorfman (1943), durante a segunda guerra mundial, recorre à utilização de análises clínicas compostas (nas quais o sangue de distintos indivíduos é misturado e é efetuada uma análise a essa mistura) com o objetivo de obter uma metodologia que permita efetuar a identificação de todos os soldados infetados com sífilis de forma mais eficiente do que a usual realização de análises individuais. Nesta metodologia a população começa por ser dividida em grupos de n indivíduos, sendo realizado um teste conjunto a cada grupo. Se o resultado do teste conjunto for negativo implicará que ninguém está infetado nesse grupo. Se o resultado do teste conjunto for positivo significará que pelo menos um dos membros está infetado e, neste caso, é necessário realizar posteriormente testes individuais a cada membro do grupo de forma a identificar quais os indivíduos dentro do grupo que estão infetados.

O principal objetivo desta metodologia é encontrar a dimensão mais eficiente para cada grupo, em função da taxa de prevalência da infeção, no sentido de minimizar o número esperado de testes a serem realizados até à identificação de todos os indivíduos infetados. De facto, Dorfman conclui que, para taxas de prevalência baixas, a utilização de análises compostas permite economizar bastantes recursos.

Desde então diversos algoritmos foram propostos com o objetivo de minimizar o número de testes necessário para a correta classificação de todos os indivíduos de uma população (Lancaster e Keller-McNulty (1998), Hughes-Oliver (2006) e Sousa (2012) fazem uma boa descrição da evolução de algumas destas metodologias) nomeadamente generalizações da metodologia de Dorfman (denominados por algoritmos hierárquicos, nos quais os grupos que são classificados como infetados são iterativamente divididos em grupos disjuntos mais pequenos, até serem realizados testes individuais para identificar os indivíduos infetados, cf. Sterret (1957), Sobel e Groll (1959), Finucan (1964), entre outros) ou

estratégias mais complexas como nos casos em que o delineamento da realização dos testes compostos é baseado em *arrays* (onde os indivíduos são colocados numa matriz bidimensional na qual cada posição é ocupada por um indivíduo, sendo posteriormente realizados testes compostos a todas as linhas e todas as colunas que formam essa matriz, pelo que cada indivíduo pertence a dois grupos, cf. Phatarfod e Sudbury (1994), Woodbury *et al.* (1995), Kim *et al.* (2007), entre outros) ou mesmo algoritmos baseados em *arrays* multidimensionais (uma extensão para *arrays* com dimensão superior a 2, cf. Berger (2000) ou Roederer (2003), entre outros).

Por outro lado, no trabalho inicial de Dorfman, tal como ocorre em muitas outras aplicações mais recentes, é considerado que não existem erros de classificação (a análise clínica identifica sempre corretamente a infeção) o que não corresponde à realidade na maioria das situações. Deste modo, os erros de má classificação, medidos pela sensibilidade e especificidade do teste, foram incluídos no modelo. Salientemos que na maioria das aplicações é considerado que a probabilidade de ocorrência de uma classificação errónea é igual nos testes individuais e nos testes compostos (isto é, considera-se válida a hipótese de ausência do efeito de diluição, cf. Tu *et al.* (1994, 1995) e Liu (2011)), ou que o número de infetados num grupo não afeta esta probabilidade (Hwang, 1976). Todavia, com recursos a modelos mais complexos podemos incluir o efeito de diluição, conforme Wein e Zenios (1996) bem como Zenios e Wein (1998) utilizam para descrever a utilização de algoritmos hierárquicos com o objetivo identificar indivíduos infetados (classificação) bem como de estimar a taxa de prevalência do VIH (estimação).

No presente trabalho, as análises descritas são qualitativas, através das quais unicamente se pretende identificar a presença *versus* ausência de uma determinada característica no sangue, pelo que iremos considerar a hipóteses de ausência do efeito de diluição. No entanto, saliente-se que a principal desvantagem da análise composta é a sua maior probabilidade de erros de classificação (diminuição da sensibilidade e/ou da especificidade), principalmente quando se verifica o efeito de diluição (veja-se, por exemplo, Santos *et al.* 2013).

Por este motivo o recurso a testes compostos só deve ser realizado se a possibilidade de ocorrência de erros de classificação for monitorizado, permitindo equilibrar as suas vantagens, em termos de diminuição do número de testes necessários, com a fiabilidade, aumento da probabilidade de ocorrência de erros de classificação. Deste modo, a metodologia mais apropriada bem como a dimensão ideal do grupo dependem da taxa de prevalência da infeção, das probabilidades associadas à correta classificação, bem como do objetivo do decisor (quanto está disposto a perder em termos de fiabilidade de forma a ganhar eficiência, isto é, diminuir o seu custo).

De referir ainda que, apesar de neste trabalho unicamente ser investigado o objetivo de classificar uma população (problema de classificação), as metodologias apresentadas são igualmente aplicada para obter “boas” estimativas da taxa de prevalência (problema de estimação), nomeadamente através da metodologia proposta por Sobel e Elashoff (1975), Lower (1983), Garner (1989), Chen e Swallow (1990), Hung e Swallow (1999) e posteriores generalizações. Neste sentido, o package *binGroup* para o *software* R (Bilder *et al.*, 2010) já tem incorporado os principais estimadores (pontuais e intervalares) baseados em testes compostos na presença de erros de classificação. Todavia, no presente trabalho, a análise apresentada é restrita ao problema de classificação.

Por fim, sublinhe-se que as aplicações destas metodologias são atualmente bastante mais vastas, uma vez que podem ser aplicadas a qualquer substância (em particular qualquer fluido) cujas amostras provenientes de diferentes indivíduos possam ser misturadas para a realização de uma análise composta. Desta forma, estas metodologias são muito aplicadas, de forma semelhante, no controlo de qualidade (por exemplo, de produtos alimentares ou na indústria farmacêutica), conforme Johnson *et al.* (1991) descreve. Boswell *et al.* (1996) apresenta múltiplos exemplos de aplicação, além do controlo de qualidade, destas metodologias e suas generalizações.

Sob o contexto previamente descrito, o presente trabalho tem como principal objetivo comparar, via simulação efetuada através de programação em R, a performance de diferentes algoritmos de classificação (baseados em testes compostos) de forma a, sabendo o valor da taxa de prevalência, da sensibilidade e da especificidade, identificar qual das metodologias é mais adequada ou, pelo menos, fornecer os valores referentes ao custo

relativo (medida de eficiência) e à fiabilidade de utilização de cada metodologia de forma a permitir uma decisão informada, com consciência dos ganhos obtidos (em termos de número de testes realizados) e perdas admitidas (em termos de sensibilidade e especificidade dos testes) em comparação com a utilização de testes individuais.

Para atingirmos este objetivo, no segundo capítulo (Noções fundamentais de testes clínicos) iremos apresentar as noções fundamentais de testes clínicos, desde a definição de testes compostos, testes qualitativos, medidas que permitem avaliar a ocorrência de má classificação (sensibilidade, especificidade, valor preditivo positivo e valor preditivo negativo), custo relativo (eficiência comparada com os testes individuais) e simulação.

No terceiro capítulo (Metodologias de classificação) iremos apresentar diversos algoritmos de classificação baseados em testes compostos, desde o original (metodologia de Dorfman), as suas generalizações (metodologias hierárquicas) bem como metodologias baseadas em *arrays* (bidimensionais) com e sem a aplicação do *master pool* (teste prévio realizado a todos os indivíduos que compõem a matriz que, caso seja negativo, ficam logo todos os elementos dessa matriz classificados). Estas diferentes metodologias em estudo serão alvo de recriação ao nível de programação em R de forma a simularmos a aplicação destes métodos uma vez que, repetido um grande número de vezes, podemos obter valores aproximados do custo relativo e da fiabilidade de cada algoritmo. Deste modo, em cada metodologia será analisado, via simulação, a dimensão de grupo mais adequada para cada taxa de prevalência, na ausência e na presença de erros de classificação (e, neste caso, considerando a sensibilidade e a especificidade do teste).

No quarto capítulo utilizaremos os ótimos de cada metodologia de forma a podermos mais adequadamente comparar a performance das metodologias apresentadas (sabendo o melhor que podemos obter com cada metodologia podemos comparar as diferentes metodologias). Algumas das nossas conclusões foram já apresentadas no 2nd *IPLeiria Internacional Health Congress* que decorreu em maio de 2014, cf. Granado e Santos (2014), apesar de não ter sido incluído nessa comunicação a metodologia 2 – modelos hierárquicos, por ainda estar em fase de programação.

Refira-se igualmente que esta comparação não pretende, de modo algum, ser exaustiva. Pelo contrário, tem como único objetivo ilustrar a comparação que pode ser realizada com recurso aos programas de simulação para cada metodologia (que são apresentados em Anexo e que podem ser utilizados para analisar quaisquer outras situações que apresentem taxas de prevalência, sensibilidade ou especificidade distintas dos casos apresentados ao longo do presente trabalho).

Deste modo, para uma dada infeção, conhecendo a sua taxa de prevalência bem como a sensibilidade e especificidade do teste (ou valores próximos, isto é, estimativas razoáveis destes valores) podemos consultar as tabelas correspondentes (ou simular os valores correspondentes, nos casos não contemplados nas tabelas apresentadas ao longo do presente trabalho) de forma a obtermos a descrição da eficiência (custo relativo, que mede o número de testes realizados) e da fiabilidade (sensibilidade e especificidade) de cada metodologia, que são informações fundamentais para a adequada tomada de decisão sobre a metodologia que deve ser aplicada em cada situação.

2 – Noções fundamentais de testes clínicos

A identificação de uma infecção ou doença num indivíduo pode ser efetuada recorrendo a análises clínicas ao sangue (poderá ser outro fluido, como a urina, mas ao longo deste trabalho, com o objetivo de simplificar a sua apresentação, iremos restringir-nos às análises ao sangue). Neste sentido, após a recolha do sangue a um indivíduo, o sangue é sujeito a análises para determinar a presença ou ausência de determinada característica que permite classificar o indivíduo em infetado ou saudável (não infetado), ao que denominaremos por classificação do indivíduo. As referidas características que permitem através dos resultados dos testes clínicos inferir sobre a existência da infecção são interpretados com base em conhecimentos na área de saúde, as quais serão consideradas conhecidas ao longo do presente trabalho.

Assim, consideremos que o nosso objetivo é classificar um conjunto de N indivíduos que constituem a população em estudo (N representará, ao longo do trabalho, a **dimensão da população** em análise). Deste modo, o nosso objetivo é investigar diferentes metodologias de organização da realização de análises clínicas de forma a identificar todos os indivíduos infetados nesta população (isto é, **metodologias de classificação** que visam a classificação de todos os N indivíduos em infetados ou saudáveis). Um parâmetro basilar nesta análise corresponde à **taxa de prevalência** da infecção, isto é, à proporção de indivíduos infetados na população que representaremos por p . Por conseguinte, a taxa de prevalência pode ser determinada por:

$$p = \frac{\text{Número de indivíduos com característica de infetado}}{N},$$

que, multiplicando por 100, indica a percentagem de indivíduos infetados na população.

2.1 – Testes individuais *versus* testes compostos

Para a classificação de toda a população é usualmente realizado um teste clínico a cada um dos indivíduos que compõem a população, isto é, através de **testes individuais** nos quais cada indivíduo é analisado isoladamente. Assim sendo, para atingir o objetivo de classificar todos os indivíduos de uma população, o número de testes necessário corresponderá ao número total de indivíduos da população em estudo, pelo que teremos de realizar N testes. Contudo, o número de testes necessários para atingir este objetivo pode ser diminuído (em algumas situações) caso em cada teste se investigue mais do que um indivíduo, através do recurso aos **testes compostos** (testes que são aplicado ao sangue composto que resulta da mistura do sangue de diferentes indivíduos).

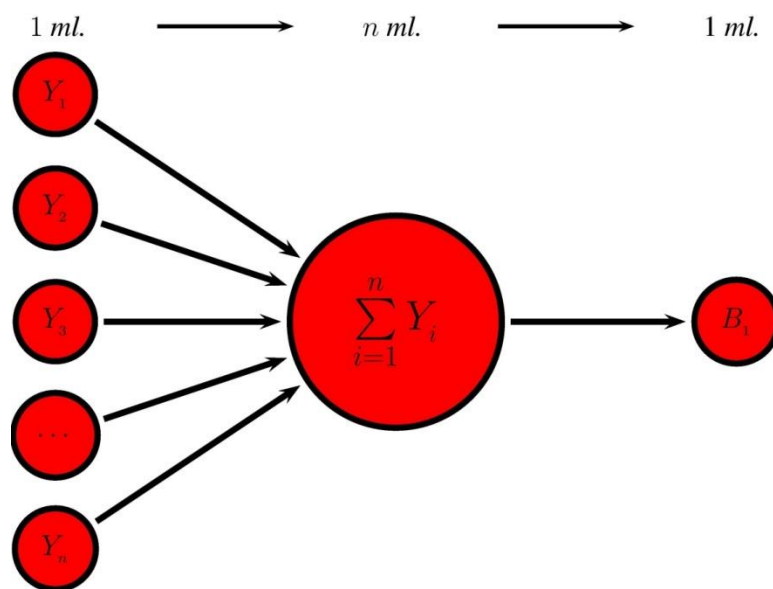


Figura 1: Realização de um teste composto

Consideremos que pretendemos analisar um grupo de n indivíduos (n representará, ao longo do trabalho, a **dimensão de cada grupo** utilizado para uma análise conjunta). Os testes compostos são realizados recolhendo a mesma quantidade de sangue a cada um dos n indivíduos, sendo este sangue misturado (de forma a ser obtida um sangue composto homogêneo) o qual será analisado. Por exemplo, conforme ilustra a Figura 1, pode ser recolhido um mililitro de sangue a cada um dos n elementos de um grupo, que após se misturar obtém-se n mililitros do sangue composto (sangue homogêneo constituído pela

mistura de n sangues distintos), do qual é retirado a quantidade adequada para o teste que se pretende aplicar (por exemplo, 1 mililitro). Se nesta análise composta o resultado do teste clínico ao sangue for negativo (não foi identificada presença de infecção) então **nenhum indivíduo do grupo está infectado** (que denominaremos por **grupo saudável**) e através de um único teste foi possível classificar n indivíduos. Caso contrário, se o resultado do teste clínico ao sangue composto for positivo implicará que há **pelo menos um elemento do grupo infectado** (grupo que designaremos por **grupo infectado**, bastando, para tal, que pelo menos um dos seus elementos esteja infectado), apesar de o teste não identificar quais os indivíduos infectados nem quantos indivíduos estão infectados dentro do grupo e, portanto, outros testes têm de ser aplicados a estes indivíduos de forma a atingirmos o nosso objetivo de classificar todos os indivíduos da população (diferentes metodologias de utilização de testes compostos para a classificação de todos os indivíduos de uma população serão descritas no capítulo referente às metodologias de classificação – capítulo 3). Assim, com este tipo de testes pretendemos, tal como na aplicação de testes individuais, classificar todos os indivíduos de uma população. A diferença reside na quantidade de testes que se tem de realizar (bem como na fiabilidade dos resultados, como iremos analisar no subcapítulo 2.3 referente à existência de erros de classificação), podendo ser bastante inferior quando aplicamos testes compostos.

Refira-se que, apesar de não ser objetivo deste trabalho, naturalmente que nem em todo o tipo de análises é possível realizar testes compostos, pois a mistura de sangues (principalmente se forem de tipos sanguíneos distintos) pode desencadear reações químicas que impeçam uma análise fiável dos resultados obtidos através da análise ao sangue composto. Todavia, estas situações podem ser facilmente investigadas e identificadas (por profissionais da saúde) e, nestes casos, não poderão ser aplicadas as metodologias baseadas em testes compostos que neste trabalho propomos. Por conseguinte, doravante iremos considerar unicamente infecções cuja mistura de sangues provenientes de diferentes indivíduos não desencadeia reações químicas que prejudiquem a fiabilidade dos testes compostos.

2.2 – Testes qualitativos *versus* testes quantitativos

As análises clínicas permitem, através de uma análise ao sangue, verificar a existência de problemas ao nível da saúde dos indivíduos. A realização dos exames irá permitir distinguir os indivíduos infetados (resultados positivos nos testes clínicos) dos não infetados (resultados negativos nos testes clínicos). No âmbito das análises clínicas podemos ter dois tipos de testes, os qualitativos e os quantitativos. Os **testes qualitativos** identificam a presença ou ausência de determinada substância e/ou característica no sangue analisado, resultando em testes positivos e negativos respetivamente. Neste tipo de testes não há a contagem ou medição de nenhuma substância no sangue, unicamente a identificação da presença de determinada substância e/ou característica. Desta forma, em relação à aplicação de testes compostos, sublinhemos que a realização de um teste qualitativo composto continuará a consistir unicamente na identificação da presença *versus* ausência dessa substância/característica, sendo, deste modo, o teste composto análogo ao teste individual (e, portanto, todas as características do teste composto deverão ser análogas às características do teste individual).

Caso seja necessária a contagem ou medição de determinada substância para a classificação então o teste será considerado **quantitativo**. Neste caso, é definido um patamar que delinea os valores que distinguem os indivíduos infetados dos não infetados e, em cada indivíduo, caso na sua análise a contagem seja superior a esse patamar então o indivíduo é classificado como infetado, caso contrário será considerado saudável (a desigualdade oposta também é por vezes utilizada, mas é menos frequente). Deste modo, os testes quantitativos individuais, além de permitirem identificar a presença (ou ausência) da infeção ou doença, podem ainda possibilitar a determinação do nível da infeção. Notemos que neste tipo de testes o procedimento para a aplicação do teste composto é bastante mais complexo. Recuperando o exemplo previamente ilustrado através da Figura 1, começamos por recolher um mililitro de sangue de cada um dos n indivíduos e consideremos que, na análise em questão, é medida a quantidade de determinada substância. Representaremos por Y_i a quantidade dessa substância observada no indivíduo i . Consideremos ainda que um indivíduo é considerado infetado se esta quantidade for superior a um determinado patamar que representamos por t . Desta forma, conhecendo os valores das variáveis Y_i (valores individuais obtidos através de testes individuais) podemos

classificar os n indivíduos: se $Y_i > t$ o indivíduo é classificado como infetado, caso contrário será classificado como saudável. Se misturarmos o sangue dos n indivíduos então vamos obter n mililitros de sangue e a quantidade da substância neste sangue composto será igual à soma das quantidades que cada indivíduo tinha (soma dos valores assumidos pelas variáveis Y_i dos n indivíduos que pertencem ao grupo). Por fim, vamos recolher um mililitro do sangue composto (depois de o sangue ser bem misturado e tornar-se homogêneo) que será a quantidade na qual será realizada a análise (este processo visa ter a mesma quantidade de sangue em ambas as análises, na individual e na composta, pois cada teste deve ser realizado com uma determinada quantidade de sangue fixa). Deste modo, no teste composto iremos unicamente medir a quantidade de substância neste mililitro em análise que corresponderá aproximadamente à média das variáveis Y_i , pelo que a conclusão se o grupo está infetado ou saudável nem sempre é possível (consulte-se, por exemplo, Santos *et al.* (2014) que identificam em que casos o teste quantitativo composto pode ser realizado com fiabilidade). Em pormenor, podemos concluir que unicamente é observada a média do grupo (única informação proveniente da análise ao sangue composto), mas pretendemos saber o máximo do grupo (isto é, se algum dos indivíduos que pertencem ao grupo ultrapassa o patamar t). De forma intuitiva, o problema de realização dos testes compostos quantitativos está ligado à existência de diluição e consequente rarefação da substância quando misturamos o sangue proveniente de um indivíduo infetado com o sangue de diversos indivíduos saudáveis (mesmo que um indivíduo tenha uma grande quantidade desta substância no sangue, a junção do seu sangue com o sangue de outros indivíduos com uma quantidade diminuta da substância pode provocar uma leitura errónea da análise composta). Sublinhemos também, que mesmo nos casos nos quais a realização dos testes quantitativos compostos é adequada, a definição do patamar que separa o sangue composto saudável do sangue composto infetado não é trivial (pois será necessariamente distinto do valor t definido para os testes individuais e dependerá da dimensão n do grupo utilizado, isto é, do número de indivíduos que pertencem ao grupo).

Assim, nos testes quantitativos compostos a situação é extremamente mais complexa, uma vez que os patamares para identificar se um sangue composto de n indivíduos está infetado teriam de ser adaptados devido à diluição e rarefação da substância em análise, conforme claramente destacam Hwang (1976), Hung e Swallow (1999). Santos *et al.* (2013) consideram a contagem do número de determinado tipo de bactérias num mililitro de

sangue. Neste caso, se juntar o sangue proveniente de um indivíduo infetado (portanto com um número elevado de bactérias) com o de diversos indivíduos saudáveis (com um número reduzido de bactérias) então a análise a um mililitro do sangue composto irá mostrar um número de bactérias próximo da média dos n indivíduos, pelo que o patamar para os testes individuais não pode ser tido em consideração nos testes compostos. Nestes casos, a adequação da aplicação de testes composto deve ter em consideração as usuais medidas de correta classificação, nomeadamente a especificidade (probabilidade de um indivíduo saudável ser classificado como tal) e a sensibilidade (probabilidade de um indivíduo infetado ser identificado pelo processo de classificação) e dependerá das distribuições de probabilidade que descrevem o número de bactérias num indivíduo saudável e num indivíduo infetado. Consulte-se, por exemplo, aplicações na identificação de infetados com VIH em Tu *et al.* (1994, 1995), Wein e Zenios (1996) e Zenios e Wein (1998).

Pelos motivos referidos, e de forma a evitar esta complexidade nas simulações que irão ser realizadas, neste projeto iremos restringir a nossa análise à utilização de testes qualitativos (ou testes quantitativos nos quais se considera que a mistura não influencia a capacidade do teste em identificar a infeção) uma vez que iremos supor que as características relacionadas com a fiabilidade dos testes são iguais nos testes individuais e nos testes compostos. De referir igualmente que há ainda casos (infeções) para os quais é aconselhado um número máximo de elementos no grupo (utilização de um valor de n inferior a determinado patamar), por se considerar que ao ser utilizado um número superior o teste composto poderá perder fiabilidade.

2.3 – Erros de classificação e medidas de fiabilidade dos testes

Os testes clínicos individuais permitem classificar um indivíduo em infetado (resultado do teste clínico positivo) ou não infetado (resultado do teste clínico negativo) mas podem existir erros de classificação uma vez que nem todos os testes executados resultam numa correta identificação e classificação de infetado e/ ou não infetado. Por este motivo, com o objetivo de medir a fiabilidade das análises clínicas, isto é, a capacidade de um teste

classificar corretamente um indivíduo, existem diversas medidas, tais como a sensibilidade, a especificidade, o valor preditivo positivo e o valor preditivo negativo. Para compreendermos estes conceitos, destaquemos que existem quatro possibilidades distintas, conforme ilustra o Quadro 1, que relacionam a classificação obtida (teste positivo *versus* teste negativo) com as características do indivíduo (infetado *versus* saudável ou não infetado):

		Estado do indivíduo	
		Infetado	Não Infetado
Resultado do teste	Positivo	Verdadeiro Positivo	Falso Positivo
	Negativo	Falso Negativo	Verdadeiro Negativo

Quadro 1: Diferentes tipos de classificação com base na relação entre o resultado dos testes e o estado dos indivíduos

Deste modo, as quatro possíveis situações resultantes de uma análise clínica são:

- **Verdadeiro positivo** – o resultado do teste clínico foi positivo (o indivíduo foi classificado como infetado) e está realmente infetado.
- **Falso positivo** – o resultado do teste clínico foi positivo (o indivíduo foi classificado como infetado) mas não está infetado (está saudável) pelo que houve um erro de classificação.
- **Falso negativo** – o resultado do teste clínico foi negativo (o indivíduo foi classificado como saudável) mas está infetado, pelo que este indivíduo foi incorretamente classificado.
- **Verdadeiro negativo** – o resultado do teste clínico foi negativo (o indivíduo foi classificado como saudável) e está saudável.

Com base nestas quatro possibilidades de resultado, podemos definir as quatro medidas usualmente utilizadas para avaliar a fiabilidade dos testes.

A **especificidade** (que representaremos por E) corresponde à probabilidade de um indivíduo saudável ser efetivamente classificado como saudável através da realização de um teste individual, isto é,

$$E = \frac{\text{Número de indivíduos saudáveis com teste negativo}}{\text{Número de indivíduos saudáveis}},$$

correspondendo à proporção (ou percentagem, se multiplicado por 100) de indivíduos saudáveis que são corretamente classificados, pelo que mede a capacidade do teste em identificar os indivíduos saudáveis. Para determinar esta medida devemos dividir o número de verdadeiros negativos pelo número de indivíduos saudáveis (e, caso se pretenda apresentar em forma de percentagem, como faremos em todas as tabelas, bastará multiplicar por 100). Com esta medida podemos igualmente avaliar a possibilidade de ocorrência de falsos positivos (pois a percentagem complementar, isto é, o que falta para os 100 por cento, corresponde à percentagem de indivíduos saudáveis que são classificados como infetados).

Por outro lado, a **sensibilidade** (representada por S) corresponde à probabilidade de um indivíduo infetado ser efetivamente classificado como infetado através da aplicação de um teste individual, isto é,

$$S = \frac{\text{Número de indivíduos infetados com teste positivo}}{\text{Número de indivíduos infetados}},$$

o que corresponde à proporção (ou percentagem, se multiplicado por 100) de indivíduos infetados que são corretamente classificados, pelo que mede a capacidade do teste em identificar os indivíduos infetados. Esta medida é definida pela divisão do número de verdadeiros positivos pelo número de indivíduos infetados (no caso de se pretender em percentagem bastará multiplicar este valor por 100). De forma análoga à especificidade, esta medida permite avaliar a ocorrência de falsos negativo uma vez que a percentagem complementar (o que falta para os 100 por cento) corresponde à percentagem de indivíduos infetados que são classificados como saudáveis.

De referir que, apesar de estas duas medidas serem igualmente utilizadas na caracterização da fiabilidade das análises clínicas, tal não implica que tenham a mesma importância, pois em muitos casos o erro de classificação relativo aos falsos negativos poderá ter consequências mais graves que uma situação de falso positivo, podendo mesmo corresponder a uma situação que deve ser evitada a todo o custo. Tal situação ocorre nos casos de possíveis contágios, nos quais a não identificação de um indivíduo infetado

poderá implicar a existência de mais indivíduos infetados no futuro. Outro caso ocorre se a infeção tiver um impacto significativo na vida do indivíduo, nomeadamente se a cura de tal problema só for possível se a infeção for diagnosticada com antecedência. Notemos igualmente que quando um indivíduo é classificado como infetado (teste positivo) outras medidas de diagnóstico podem ser utilizadas para confirmar a existência da infeção. Por este motivo, apesar de a fiabilidade dos testes ser analisada através destas duas medidas, sensibilidade e especificidade, tal não significa que estas tenham a mesma importância, pois nos casos referidos a sensibilidade terá inequivocamente uma importância e um impacto superior relativamente à especificidade.

Apesar de não utilizarmos nas nossas análises (nos resultados que apresentaremos das simulações), há outras medidas (que naturalmente estão associadas às previamente apresentadas) que permitem avaliar a fiabilidade das análises, tais como o **valor preditivo positivo**, que corresponde à probabilidade de um indivíduo classificado como infetado estar efetivamente infetado (logo, mede a fiabilidade para quem recebe um resultado positivo) e o **valor preditivo negativo**, que corresponde à probabilidade de um indivíduo classificado como saudável estar efetivamente saudável (medindo a fiabilidade de quem recebe um resultado negativo). Contudo, como pretendemos sobretudo comparar a fiabilidade das metodologias aplicadas, restringimos a nossa análises à sensibilidade e especificidade.

Com base nos conceitos de sensibilidade e especificidade, podemos igualmente definir a especificidade composta e a sensibilidade composta, isto é, as mesmas medidas mas relativamente à realização de testes compostos. Deste modo, **especificidade composta** (que representaremos por *EC*) corresponde à probabilidade de um grupo saudável (nenhum elemento infetado) ser efetivamente classificado como saudável, isto é,

$$EC = \frac{\text{Número de grupos saudáveis com teste negativo}}{\text{Número de grupos saudáveis}},$$

que corresponde à proporção (ou percentagem) de grupos saudáveis que são corretamente classificados e, de forma análoga, a **sensibilidade composta** (representada por *SC*) corresponde à probabilidade de um grupo infetado ser efetivamente classificado como infetado (proporção de grupo infetados que são corretamente classificados), isto é,

$$SC = \frac{\text{Número de grupos infetados com teste positivo}}{\text{Número de grupos infetados}}.$$

Como esta dissertação irá ter como base unicamente a realização de testes qualitativos (nos quais o teste composto é igual ao teste individual, conforme referido no subcapítulo anterior, sendo utilizado um sangue composto em vez de um sangue proveniente de um único indivíduo), considerando igualmente a ausência do efeito de diluição nestes casos (que a característica desapareça com a junção dos sangues), a sensibilidade e a especificidade dos testes compostos serão considerados iguais às correspondentes medidas dos testes individuais.

Conforme previamente referido, os testes compostos não permitem identificar quais os indivíduos realmente infetados dentro do grupo uma vez que unicamente classifica o grupo e um grupo infetado é um conjunto de indivíduos nos quais há pelo menos um indivíduo infetado. Deste modo, no próximo capítulo serão apresentadas metodologias que permitem identificar quais os indivíduos infetados num grupo infetado. Assim, há ainda que definir a sensibilidade e a especificidade de cada metodologia que corresponderá à capacidade da metodologia classificar corretamente cada indivíduo. Por conseguinte, a **especificidade de uma metodologia** (representada por *EM*) corresponde à probabilidade de um indivíduo saudável ser efetivamente classificado como saudável na aplicação dessa metodologia (isto é, corresponde à proporção ou percentagem de indivíduos saudáveis que são corretamente classificados quando é utilizada a metodologia na classificação da população). De forma semelhante, a **sensibilidade de uma metodologia** (representada por *SM*) corresponde à probabilidade de um indivíduo infetado ser classificado como infetado através da aplicação dessa metodologia (proporção ou percentagem de indivíduos infetados que são corretamente classificados através da aplicação da metodologia em análise).

Naturalmente, a sensibilidade e a especificidade de cada metodologia serão fatores fundamentais para ter em conta na escolha da metodologia a utilizar em cada caso/infeção, pelo que serão determinadas nas simulações das metodologias. Neste sentido, além de uma análise aos custos associados à aplicação de cada metodologia (nomeadamente da análise do número de testes realizados para cada caso) há igualmente uma investigação à fiabilidade de cada uma (pois o recurso a testes compostos poderá comprometer a fiabilidade das análises).

2.4 – Custo relativo *versus* eficiência

O principal objetivo da criação de metodologias de classificação baseadas em testes compostos foi minimizar o número de testes que é necessário realizar para classificar uma população de dimensão N . Deste modo, e com o objetivo de medir esta redução e de se obter uma medida que não depende da dimensão da população é habitualmente utilizado o custo relativo (identificado por CR) que compara o número de testes de cada metodologia com o número de testes quando aplicamos testes individuais. Notemos, antes de definirmos esta medida, que o número de testes a serem realizados na aplicação de testes individuais é igual ao número de indivíduos, conforme previamente referido. Contudo, quando aplicamos testes compostos, este número não será determinístico, pois pode ocorrer que os indivíduos infetados estejam todos no mesmo grupo, ou podem estar espalhados por grupos distintos. Deste modo, não teremos um número determinístico de testes, mas unicamente o número esperado de testes (o que em média ocorre para esta situação) e que dependerá da taxa de prevalência bem como do número n de indivíduos em cada grupo. Assim sendo, o **custo relativo** de uma metodologia é definida pelo número esperado de testes na aplicação dessa metodologia a dividir pela dimensão da população (que corresponde ao número de testes realizados se aplicarmos testes individuais) a multiplicar por 100 (para que seja apresentada em forma de percentagem, conforme apresentaremos os nossos resultados nas simulações):

$$CR = \frac{\text{Número de testes realizados}}{N} \times 100.$$

Sublinhemos, então, que se este número for superior a 100 significará que o número esperado de testes é superior a N pelo que é preferível utilizar testes individuais (pois, neste caso só teremos de realizar N testes). Por exemplo, se for 120, significará que (em média) teremos de realizar 120 testes para classificar 100 indivíduos. Nos casos em que o custo relativo é inferior a 100 significará que (em média) necessitamos de realizar menos testes com a aplicação da metodologia. Por exemplo, se for igual a 40, significará que (em média) teremos de realizar 40 testes para classificar 100 indivíduos. Neste caso, e tendo em consideração que teríamos de realizar 100 testes no caso de aplicação de testes individuais, poupamos 60 testes (houve uma redução, em média, de 60 por cento dos testes necessários para a classificação da população). Assim, uma vez que pretendemos minimizar o número

de testes, tal implica minimizar o custo relativo uma vez que quanto menor for o valor deste coeficiente maior será a **eficiência** da metodologia em análise (menos testes são necessários, em média, para classificar o mesmo número de indivíduos).

Em suma, o custo relativo traduz, através de percentagem, os ganhos conseguidos em termos de número de testes realizados para cada metodologia de classificação comparativamente com a quantidade de testes individuais que é necessário realizar na classificação dos indivíduos.

Neste sentido, esta medida é central neste trabalho, uma vez que o objetivo do projeto é minimizar o número de testes necessário para a classificação de todos os indivíduos da população em estudo através do recurso à realização de testes clínicos compostos. Sublinhemos ainda que, segundo Liu (2011), o custo associado à realização dos testes é praticamente proporcional ao número de testes, pelo que o custo associado à mistura poderá (na maioria das aplicações) ser negligenciado (pois já existem máquinas que automaticamente concretizam a mistura homogênea dos sangues provenientes de diferentes indivíduos pelo que, após o investimento inicial nessa máquina pelo laboratório, não haverá mais custos significativos com este processo). Por conseguinte, a minimização do número de testes implica necessariamente uma diminuição do custo associado à classificação da população.

2.5 – A utilização da simulação na análise da performance de testes clínicos

A **simulação** é um processo artificial que imita o comportamento de um fenómeno aleatório (não determinístico), permitindo obter uma descrição aproximada das características de um processo (neste caso, os resultados de aplicação de distintas metodologias de classificação) uma vez que a sua realização é executada um elevado número de vezes (o que é acessível com o recurso a um computador). Por exemplo, através da simulação do lançamento de uma moeda ao ar um grande número de vezes é possível identificar que quanto maior for o número de lançamentos da moeda ao ar realizados, as

frequências relativas (valor aproximado da probabilidade) vão ficando cada vez mais próximas dos 50 por cento (uma vez que a probabilidade considerada foi de 0.5), conforme ilustra a Figura 2.

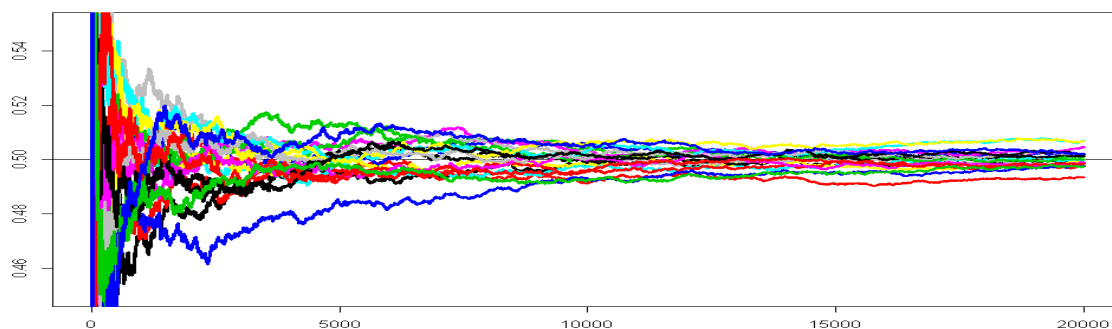


Figura 2: Evolução da proporção de ocorrência de face nacional em 20 sequências de 20000 lançamentos de uma moeda ao ar

Para realizar esta experiência elementar recorreremos a um comando fundamental em todas as simulações (Ross, 2006) que permite gerar números aleatórios (em rigor são número pseudoaleatórios uma vez que são gerados por algoritmos conhecidos e, como tal, conhecido o número inicial é possível determinar quais são os números que irão ser gerados) no intervalo real $[0,1]$, que no *software* utilizado corresponde ao comando *runif(1,0,1)*. Deste modo, cada lançamento de uma moeda será definido através de um número aleatório sendo, por exemplo, associado ao lançamento a face nacional quando o valor for inferior a 0.5 e a face europeia comum caso contrário. Assim, o número de vezes que sai face nacional pode ser contabilizado o que permitirá determinar a frequência relativa. A frequência relativa (tendo em consideração os resultados de convergência estocástica, nomeadamente a Lei dos Grandes Números que garante a convergência das frequências relativas para a correspondente probabilidade) irá aproximar-se da probabilidade pretendida à medida que aumentamos o número de lançamentos. Assim, a repetição da experiência aleatória permite a determinação da frequência relativa do acontecimento em análise, determinando um valor aproximado da probabilidade desse acontecimento. Neste exemplo elementar de lançamento de uma moeda já conhecemos o valor que pretendemos obter através da simulação, pelo que será fácil aferir da distância entre os valores obtidos pela simulação e o valor teórico. Na aplicação de testes compostos que iremos simular desconhecemos o valor teórico (caso contrário não seria necessário realizar as simulações) pelo que teremos de determinar qualquer medida que nos permita

identificar se o valor simulado está perto ou longe do valor teórico. Como tal, e como numa simulação o valor obtido terá igual probabilidade de ser superior ou inferior ao valor teórico, repetimos diversas vezes a simulação (denominadas habitualmente por réplicas) e analisamos os valores obtidos. Na Figura 2 repetimos 20 vezes 20000 lançamentos, isto é, realizamos 20 réplicas de 20000 lançamentos através das quais obtivemos 20 valores aproximados do valor teórico. Quando desconhecemos o valor teórico utilizamos a informação de todas as réplicas, recorrendo à **média** como valor aproximado e, para medir a variabilidade dos valores obtidos, o **desvio padrão** dos valores obtidos nas diferentes réplicas. Assim sendo, caso o desvio padrão seja diminuto significará que as distâncias entre os valores obtidos nas diferentes simulações são pequenas pelo que a aproximação é boa; caso contrário significará que nas simulações foram obtidos valores com uma distância considerável pelo que na utilização dessa metodologia poderemos obter performances distintas em diferentes aplicações.

Deste modo, supondo conhecidas as condições da infeção (taxa de prevalência, sensibilidade e especificidade) podemos simular a realização de análises clínicas de forma a classificar cada indivíduo da população, primeiro simulado quem está ou não infetado (recorrendo à geração de um número aleatório e utilizando a taxa de prevalência p , pois se $runif(1,0,1) < p$ o indivíduo está infetado, caso contrário estará saudável) e depois simulando o resultado da análise clínica através da geração de um novo número aleatório e utilizando os valores da sensibilidade caso o indivíduo esteja infetado (se $runif(1,0,1) < S$ o resultado será positivo pelo que corresponderá a um verdadeiro positivo, caso contrário o resultado será negativo e corresponderá a um falso negativo) e recorrendo ao valor da especificidade caso esteja saudável (se $runif(1,0,1) < E$ o resultado será negativo pelo que corresponderá a um verdadeiro negativo, caso contrário o resultado será positivo e corresponderá a um falso positivo). Se repetirmos esta prova muitas vezes (considerando uma população com dimensão elevada) e determinarmos as frequências relativas de cada um dos 4 possíveis resultados (verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos) podemos determinar valores aproximados das respetivas medidas de fiabilidade (nomeadamente a sensibilidade e a especificidade da metodologia) permitindo, desta forma, conhecer as probabilidades de ocorrência de má classificação. Conforme previamente referido, por forma a podermos ter uma medida para analisar a variabilidade destes valores, podemos, em vez de simularmos uma única população, simular

diversas populações (em todas as simulações que realizamos utilizamos 1000 populações) e depois analisar a variabilidade dos valores obtidos (recorrendo ao desvio padrão). Deste modo, como pretendemos um único valor, iremos utilizar a média dos valores obtidos nas 1000 populações (por conseguinte, os valores apresentados em todas as tabelas referentes a simulações foram obtidos determinando a média do que foi obtido em 1000 populações cada uma das quais com pelo menos 10000 indivíduos) e para analisar a sua variabilidade, isto é, se existe uma grande disparidade entre os valores observados nas distintas populações, determinou-se o desvio padrão (que é apresentado entre parêntesis nas tabelas). A partir desta informação é uma questão de elaboração do algoritmo adequado para a descrição de testes compostos e, posteriormente, das metodologias que serão descritas ao longo do capítulo 3. Os principais programas elaborados para a realização das simulações são apresentados em Anexo. Com estes programas podemos analisar o comportamento de cada uma das metodologias, nomeadamente no que se refere a alterações da dimensão dos grupos (ver o que acontece ao custo relativo e à fiabilidade da classificação se para os mesmos valores da taxa de prevalência, sensibilidade e especificidade alterarmos a dimensão do grupo). Deste modo, para cada situação, conhecendo estas características, podemos analisar a performance das diferentes metodologias de forma a conseguirmos tomar uma decisão informada acerca da metodologia a adotar para a infeção em estudo.

Em suma, o processo de simulação será utilizado na execução e análise das metodologias de classificação para retratar as metodologias baseadas em testes clínicos compostos a fim de identificar os indivíduos infetados numa determinada população. Deste modo, a simulação irá recriar as metodologias de classificação com base nos parâmetros de taxa de prevalência (p), sensibilidade (S), especificidade (E) e dimensão do grupo (n) permitindo, após as execuções dos procedimentos, identificar o custo relativo (CR) bem como a sensibilidade e especificidade da metodologia (SM e EM). Estas simulações serão executadas através do *software* gratuito R, pois trata-se uma ferramenta que possibilita o tratamento de dados contendo diversas funções de probabilidades e estatística já previamente definidas, além de ser eventualmente o *software* mais utilizado em investigação em Estatística (o que comprova claramente a sua qualidade) e de estar acessível a qualquer indivíduo ou empresa que o pretenda utilizar (uma vez que é *freeware*).

3 – Metodologias de classificação

Os testes compostos permitem identificar os grupos infetados (grupos com pelo menos um indivíduo infetado) mas não identificam os indivíduos infetados nem permitem saber quantos indivíduos estão infetados dentro do grupo. Deste modo, para a classificação dos indivíduos ter-se-á de realizar outros testes aos elementos do grupo infetado de forma a podermos classificar todos os indivíduos da população. Desta forma, neste capítulo serão apresentadas as diferentes metodologias de classificação que permitem a utilização de testes conjuntos para a classificação de todos os indivíduos de uma dada população. A eficiência (custo relativo da metodologia – CR) bem como a fiabilidade (mensurada através da especificidade da metodologia – EM – e da sensibilidade da metodologia – SM) de cada uma das metodologias apresentadas depende de diversos fatores, nomeadamente da dimensão dos grupos (n), da taxa de prevalência da infeção em estudo (p) e da fiabilidade dos testes individuais e compostos (mensurada através da especificidade individual e composta – E e EC – e da sensibilidade individual e composta – S e SC). Por conseguinte, estas metodologias devem ser investigadas sob diferentes condições, o que irá permitir identificar, através de simulações, a dimensão ótima do grupo para cada caso dentro de cada metodologia, isto é, a dimensão que permite obter a melhor eficiência (ao minimizar o custo relativo e, por conseguinte, o número esperado de testes a serem realizados para a identificação de todos os indivíduos infetados de uma população). A programação necessária para a obtenção das simulações em cada metodologia é apresentada em Anexo. A informação da dimensão ótima para cada metodologia será fundamental para a comparação das diferentes metodologias (que será realizada no capítulo seguinte, referente à discussão dos resultados obtidos por simulação) uma vez que nessa comparação só será tida em consideração o ótimo (isto é, vamos comparar o melhor que cada metodologia consegue obter) para cada cenário em cada metodologia, o que corresponde ao principal objetivo do presente trabalho, isto é, conhecendo a taxa de prevalência e a fiabilidade dos testes (individuais e compostos) identificar a metodologia mais adequada a ser aplicada através da análise da eficiência e fiabilidade das diferentes metodologias.

Deste modo, este capítulo pretende apresentar, de forma resumida, a aplicação de

metodologias baseadas em análises compostas com o objetivo de classificação, nomeadamente para testes qualitativos quer na ausência quer na presença de problemas de má classificação. Iremos começar por analisar a metodologia de Dorfman (primeira metodologia e, como tal, a mais elementar), um caso particular de uma metodologia hierárquica e um caso particular de uma metodologia baseada em *arrays* (restrita a duas dimensões e igual número de linha e colunas) com e sem a utilização prévia de *master pool* (teste composto a todos os n^2 indivíduos que pertencem à mesma matriz).

3.1 – *M1* – Metodologia de Dorfman

Os testes compostos foram introduzidos na Estatística por Dorfman (1943), durante a segunda guerra mundial, com o objetivo de obter uma identificação mais eficiente dos soldados com sífilis, comparativamente com a realização de testes individuais. Segundo a metodologia de Dorfman, a população é dividida em grupos de n indivíduos, sendo retirada uma amostra de sangue (fluido no qual será analisada a presença ou ausência da bactéria *Treponema pallidum* – que permite a identificação da sífilis) a cada um dos indivíduos de cada grupo. Todas as amostras de sangue de cada grupo de indivíduos são combinadas (misturadas até ser obtida uma mistura homogénea do sangue dos n indivíduos que compõem cada grupo) e submetidas à realização de teste qualitativos (que identificam a presença *versus* ausência da referida bactéria). Se o resultado do teste qualitativo composto for negativo (ausência da bactéria no sangue composto) nenhum indivíduo do grupo está infetado; se o resultado do teste qualitativo composto for positivo (presença da bactéria no sangue composto) implicará que há pelo menos um indivíduo do grupo infetado.

Neste caso, em que o grupo é classificado como infetado, Dorfman propôs a realização de testes individuais a todos os elementos do grupo de forma a serem identificados todos os indivíduos infetados. Deste modo, podemos resumir a utilização da metodologia de Dorfman (que representaremos por *M1*) do seguinte modo:

- Seleção da população e subdivisão da população em grupos de n indivíduos. Para cada grupo é obtida uma mistura homogénea do sangue dos indivíduos que compõem cada grupo (sangue composto) e realiza-se um teste composto:

- Se o resultado do teste conjunto for negativo então todos os indivíduos do grupo são classificados como saudáveis;
- Se o resultado do teste conjunto for positivo então realizam-se testes individuais de forma a classificar individualmente todos os elementos.

Notemos que num grupo não infetado bastará a realização de um teste para classificar n indivíduos, pelo que há uma poupança de $n-1$ testes em comparação com a realização de testes individuais. No caso de o grupo estar infetado, além da realização do teste composto há a realização de n testes individuais, pelo que houve a realização de $n+1$ testes, isto é, de mais um teste do que se fossem logo efetuados testes individuais. Deste modo, haverá poupança na utilização desta metodologia no caso em que os grupos estão maioritariamente não infetados (o que unicamente ocorre se a taxa de prevalência for reduzida). Dorfman (1943) demonstrou que a utilização desta metodologia é mais eficiente do que a realização de teste individuais se a taxa de prevalência for inferior a 30 por cento, sendo a eficiência maior quanto menor for a taxa de prevalência da infeção. Na sua aplicação, identificação de soldados infetados com sífilis, como em algumas regiões (e considerando algumas etnias) a percentagem de soldados infetados era próxima de 1 por cento (taxa de prevalência igual a 0.01, $p = 0.01$) Dorfman conseguiu reduzir o número de testes em 80 por cento (isto é, utilizar uma metodologia cujo custo relativo é 20 por cento, $CR = 0.20$).

Contudo, a análise que Dorfman é restrita aos casos em que não há erros de classificação (especificidade e sensibilidade iguais a um, ou seja, iguais a 100 por cento), uma simplificação que poderá, em algumas situações, não corresponder à realidade pelo que analisaremos igualmente situações com sensibilidade e especificidade distintas.

No Quadro 2 são apresentados os resultados obtidos em simulações realizadas sobre a aplicação da metodologia de Dorfman, supondo ausência de erros de classificação, nas quais foi utilizada a programação apresentada no Anexo 7, utilizando diferentes taxas de prevalência (p assume os valores 10, 5, 1, 0.5 e 0.1 por cento) bem como diferentes dimensões para o grupo (n assume os valores 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 15, 25, e 50) de forma a podermos observar o comportamento destas metodologias nas variações destes parâmetros. Notemos que unicamente foram utilizadas taxas de prevalência baixas uma vez

que a utilização de testes compostos unicamente tem vantagens, em relação à utilização de testes individuais, quando tal se verifica.

	$p = 0,1$	$p = 0,05$	$p = 0,01$	$p = 0,005$	$p = 0,001$
$n = 2$	68,99 (0,89)	59,76 (0,79)	51,97 (0,39)	50,96 (0,22)	50,20 (0,11)
$n = 3$	60,15 (1,41)	47,68 (1,02)	36,31 (0,52)	34,81 (0,42)	33,59 (0,16)
$n = 4$	59,28 (1,90)	43,48 (1,52)	29,05 (0,77)	26,93 (0,49)	25,39 (0,21)
$n = 5$	60,75 (1,86)	42,67 (1,66)	24,71 (0,82)	22,56 (0,66)	20,47 (0,26)
$n = 6$	63,54 (2,29)	43,47 (1,66)	22,68 (0,97)	19,62 (0,75)	17,25 (0,32)
$n = 7$	66,44 (2,34)	44,43 (1,97)	21,04 (1,27)	17,59 (0,92)	14,95 (0,41)
$n = 8$	69,83 (2,16)	46,04 (2,44)	19,98 (1,39)	16,36 (0,95)	13,33 (0,47)
$n = 9$	72,34 (2,38)	48,11 (2,80)	19,54 (1,60)	15,38 (1,14)	12,00 (0,51)
$n = 10$	75,03 (3,01)	49,83 (2,69)	19,43 (1,71)	14,65 (1,37)	11,03 (0,68)
$n = 11$	77,29 (3,01)	52,69 (3,28)	19,20 (1,91)	14,34 (1,14)	10,19 (0,66)
$n = 12$	79,87 (3,07)	53,65 (3,25)	19,99 (2,05)	14,13 (1,43)	9,70 (0,78)
$n = 15$	86,54 (2,73)	60,94 (3,46)	20,54 (2,71)	13,71 (1,80)	8,03 (0,88)
$n = 25$	96,90 (2,36)	76,66 (5,32)	26,30 (3,56)	15,53 (2,88)	6,56 (1,24)
$n = 50$	101,55 (0,91)	94,97 (3,23)	41,11 (5,68)	24,20 (5,19)	7,21 (2,92)

Quadro 2: Custos relativos obtidos na simulação da metodologia de Dorfman sem erros de classificação

Recordamos, para a correta interpretação dos valores obtidos, que foram utilizadas 1000 réplicas de populações constituídas por 10000 indivíduos, sendo apresentado o valor da média (que corresponde à estimativa do custo relativo) e o valor do desvio padrão (que é apresentado entre parêntesis) e que permite avaliar a variabilidade dos resultados obtidos nas diferentes populações. Deste modo, a análise aos valores obtidos (Quadro 2) permite tirar as seguintes conclusões:

- Os valores dos desvios padrões são, regra geral, diminutos pelo que os valores das estimativas são fiáveis (isto é, os valores obtidos nas 1000 populações distintas são próximos uns dos outros);
- A eficiência aumenta à medida que diminui a taxa de prevalência (este facto é notório se observarmos os valores situados na mesma linha do Quadro 2);
- Para as taxas de prevalências analisadas a eficiência aumenta até determinada dimensão do grupo (que corresponderá à dimensão ótima) e depois começa a diminuir (este facto pode ser observado analisando os valores situados na mesma coluna do Quadro 2);
- Em todos os casos analisados (exceto para grupo de dimensão 50 quando a taxa de prevalência é igual a 10 por cento) é sempre preferível (mais eficiente) utilizar a metodologia de Dorfman do que testes individuais (o que só acontece porque restringimos a análise aos casos em que é vantajoso o recurso a testes compostos, isto é, a taxas de prevalência baixas, nomeadamente a valores menores ou iguais a 10 por cento).

Deste modo concluímos (através dos resultados obtidos pelas simulações realizadas, que não foram todos apresentados no trabalho de forma a não serem apresentadas tabelas demasiado extensas) que os valores ótimos para a dimensão do grupo em termos de eficiência na metodologia de Dorfman são os apresentados no Quadro 3.

Nesta metodologia, no caso de ausência de erros de classificação, Dorfman (1943) deduziu a fórmula que permite deduzir o custo relativo para cada cenário, correspondendo a

$$CR = \frac{n+1}{n} - (1-p)^n,$$

o que permite determinar os valores teóricos (apresentados no Quadro 3). Notemos que as distâncias entre os dois valores são diminutas, o que notoriamente demonstra a fiabilidade dos valores aproximados que foram obtidos. Por outro lado, saliente-se que nas restantes metodologias, e mesmo nesta metodologia de Dorfman quando se inclui a possibilidade de ocorrência de erros de classificação, não há fórmulas matemáticas disponíveis (como a previamente apresentada) que nos indiquem os valores das variáveis que se pretende analisar, isto é, do custo relativo, da sensibilidade e da especificidade das metodologias investigadas. Caso tais fórmulas estivessem disponíveis, o recurso à simulação seria desnecessário (porque havemos de obter valores aproximados se podemos ter valores exatos). É nesta ausência de fórmulas para a obtenção dos valores das variáveis cruciais para a tomada de decisão que está o fundamento do recurso à simulação no estudo das metodologias propostas neste trabalho.

	$p = 0,1$	$p = 0,05$	$p = 0,01$	$p = 0,005$	$p = 0,001$
n	4	5	11	15	32
Custo relativo (valor obtido por simulação)	59,28	42,67	19,20	13,71	6,22
Custo relativo (valor teórico)	59,39	42,62	19,55	13,91	6,28

Quadro 3: Valores ótimos utilizando a metodologia de Dorfman

Com base no Quadro 3, podemos concluir relativamente à metodologia de Dorfman, nomeadamente aos valores mais baixos do custo relativo para as diferentes taxas de prevalência, as seguintes afirmações:

- Para uma dimensão de grupos constituído por quatro indivíduos e com uma taxa de prevalência de 0.1, o custo relativo resultou num valor de 59.28% o que significa que foram necessários 59.28% dos testes para identificar os indivíduos infetados (comparativamente com a realização os testes individuais) sendo conseguido um ganho de 40.72% (100%-59.28%) dos testes na execução dos testes compostos;
- Para uma dimensão de grupos constituído por cinco indivíduos e com uma taxa de prevalência de 0.05, o custo relativo resultou num valor de 42.67% o que significa

que foram necessários 42.67% dos testes para identificar os indivíduos infetados (comparativamente com a realização os testes individuais) sendo conseguido um ganho de 57.33% (100%-42.67%) dos testes na execução dos testes compostos;

- Para uma dimensão de grupos constituído por onze indivíduos e com uma taxa de prevalência de 0.01, o custo relativo resultou num valor de 19.20% o que significa que foram necessários 19.20% dos testes para identificar os indivíduos infetados (comparativamente com a realização os testes individuais) sendo conseguido um ganho de 80.80% (100%-19.20%) dos testes na execução dos testes compostos;
- Para uma dimensão de grupos constituído por quinze indivíduos e com uma taxa de prevalência de 0.005, o custo relativo resultou num valor de 13.71% o que significa que foram necessários 13.71% dos testes para identificar os indivíduos infetados (comparativamente com a realização os testes individuais) sendo conseguido um ganho de 86.29% (100%-13.71%) dos testes na execução dos testes compostos;
- Para uma dimensão de grupos constituído por trinta e dois indivíduos e com uma taxa de prevalência de 0.001, o custo relativo resultou num valor de 6.22% o que significa que foram necessários 6.22% dos testes para identificar os indivíduos infetados (comparativamente com a realização os testes individuais) sendo conseguido um ganho de 93.78% (100%-6.22%) dos testes na execução dos testes compostos.

Na análise efetuada no Quadro 4 incluímos os erros de classificação, considerando que a sensibilidade e a especificidade (dos testes individuais e nos testes compostos) assumem o valor 0.95 (quando os erros de classificação são iguais a 5 por cento). Neste caso, além de ser considerada a eficiência do processo (medida pelo custo relativo – *CR*), serão igualmente apresentadas a sensibilidade (*SM*) e a especificidade (*EM*) da metodologia de Dorfman. De referir que no Quadro 4, com o objetivo de evitar a apresentação de tabelas demasiado extensas, só foram incluídos alguns dos resultados obtidos nas simulações realizadas, pelo que as conclusões que iremos apresentar (nomeadamente em relação à dimensão ótima dos grupos para cada taxa de prevalência) não são só fundamentadas nos valores apresentados. Assim, os resultados obtidos nestas simulações permitem tirar as seguintes ilações em relação à metodologia de Dorfman:

n		$p = 0,1$	$p = 0,05$	$p = 0,01$	$p = 0,005$	$P = 0,001$
3	<i>CR</i>	62,73 (1,40)	51,26 (1,10)	40,92 (0,83)	39,83 (0,77)	38,63 (0,70)
	<i>SM</i>	90,25 (1,55)	90,60 (2,69)	90,39 (5,15)	89,66 (7,78)	90,17 (6,02)
	<i>EM</i>	98,86 (0,22)	99,31 (0,15)	99,65 (0,10)	99,71 (0,10)	99,73 (0,09)
5	<i>CR</i>	62,01 (2,03)	45,40 (1,97)	29,48 (1,18)	27,21 (0,98)	25,52 (0,99)
	<i>SM</i>	90,24 (1,80)	89,92 (2,81)	89,89 (5,98)	90,49 (8,02)	89,86 (6,52)
	<i>EM</i>	98,17 (0,28)	98,91 (0,24)	99,59 (0,14)	99,67 (0,10)	99,73 (0,10)
10	<i>CR</i>	73,77 (2,50)	51,33 (2,63)	23,64 (1,78)	19,37 (1,56)	15,97 (1,53)
	<i>SM</i>	90,32 (1,80)	90,23 (2,93)	89,59 (5,82)	90,51 (7,73)	90,92 (5,66)
	<i>EM</i>	96,96 (0,34)	98,06 (0,31)	99,37 (0,17)	99,55 (0,15)	99,71 (0,12)
15	<i>CR</i>	83,21 (2,78)	59,92 (3,40)	24,89 (2,51)	17,92 (2,24)	13,41 (1,78)
	<i>SM</i>	90,17 (2,03)	90,26 (3,08)	91,28 (4,92)	90,48 (8,30)	90,86 (5,23)
	<i>EM</i>	96,30 (0,42)	97,46 (0,34)	99,13 (0,19)	99,44 (0,16)	99,64 (0,12)
25	<i>CR</i>	92,69 (2,79)	73,80 (4,27)	29,27 (3,93)	19,66 (3,54)	11,10 (2,46)
	<i>SM</i>	90,36 (2,16)	89,75 (2,81)	89,99 (5,38)	88,52 (8,01)	90,73 (5,18)
	<i>EM</i>	95,62 (0,42)	96,61 (0,41)	98,76 (0,26)	99,25 (0,24)	99,64 (0,13)

Quadro 4: Eficiência e fiabilidade obtidas na simulação da metodologia de Dorfman considerando erros de classificação de 5 por cento

- Os valores ótimos (em termos de eficiência) da metodologia são (quase sempre) iguais quer seja considerada a possibilidade de erros de classificação quer não o seja;
- A eficiência diminui ligeiramente com a possibilidade de ocorrência de erros de classificação (considerando a mesma dimensão do grupo, isto é, para o mesmo n);
- A sensibilidade da metodologia de Dorfman é sempre inferior à sensibilidade individual e composta considerada. De facto, e considerando o caso retratado no qual a sensibilidade e a especificidade (individual e composta) assumem o valor 0.95, para um indivíduo infetado ser classificado como infetado, terá de obter um teste composto positivo (que ocorre com probabilidade por 0.95) e depois no teste individual terá de obter igualmente um teste positivo (o que ocorrerá igualmente com probabilidade igual a 0.95). Assim sendo, e considerando que o resultado do primeiro teste não influencia o resultado do posterior teste individual, a probabilidade de um indivíduo infetado ser classificado como infetado será 0.95×0.95 que é igual a 0.9025 (valor que se obtém para a sensibilidade da metodologia de Dorfman, pelo que os valores obtidos por simulação são sempre próximos deste);
- A especificidade da metodologia de Dorfman é sempre superior à especificidade individual e composta utilizada. A obtenção do valor teórico da especificidade da metodologia de Dorfman não é tão acessível como o da sensibilidade, mas pode igualmente ser deduzida (consulte-se, por exemplo, Santos *et al.*, 2013). Contudo, o valor da especificidade da metodologia de Dorfman depende da taxa de prevalência e da dimensão do grupo, sendo maior quanto menor for o valor de p e/ou menor for o valor de n ;
- Em termos gerais no que concerne à fiabilidade, há uma perda devido à diminuição da sensibilidade e um ganho relativo ao aumento da especificidade, sendo habitualmente a diminuição da sensibilidade, em termos absolutos, superior ao aumento da especificidade. Todavia, como a sensibilidade refere-se unicamente aos indivíduos infetados (que são uma minoria uma vez que estamos a considerar taxas de prevalência baixas) e a especificidade aos indivíduos saudáveis (que são a clara

maioria) o número de classificações corretas (verdadeiros positivos mais verdadeiros negativos) será superior na utilização da metodologia de Dorfman. Contudo, esta análise deve ser realizada com precaução, pois o erro associado à sensibilidade (falsos negativos) pode ser, nomeadamente em casos de possível contágio, muito mais grave do que o erro associado à especificidade (falsos positivos).

Através das simulações realizadas para a apresentação do Quadro 4 e de outras análogas (recorrendo à programação apresentada no Anexo 7), concluímos (via simulação) que, considerando erros de classificação a 5 por cento (isto é, considerando a sensibilidade e especificidade individuais e compostas iguais a 95 por cento) os valores ótimos para a dimensão do grupo em termos de eficiência na aplicação da metodologia de Dorfman são os expostos no Quadro 5.

	$p = 0,1$	$p = 0,05$	$p = 0,01$	$p = 0,005$	$p = 0,001$
n	4	5	11	15	32
CR	60,93	45,18	23,46	18,23	10,95
SM	90,31	90,32	90,15	90,40	90,51
EM	98,54	98,74	99,33	99,44	99,61

Quadro 5: Valores ótimos, com erros de classificação de 5%, utilizando a metodologia de Dorfman

Os valores patentes no Quadro 5, referentes à aplicação da metodologia de Dorfman (MI) podem ser interpretados do seguinte modo:

- Para uma taxa de prevalência igual a 10 por cento, a dimensão ótima do grupo é 4, realizando-se (em média) 60.93 testes para classificar 100 indivíduos, sendo 90.31 por cento dos infetados e 98.54 por cento dos saudáveis bem classificados;
- Para uma taxa de prevalência igual a 5 por cento, a dimensão ótima do grupo é 5, realizando-se (em média) 45.18 testes para classificar 100 indivíduos, sendo 90.32 por cento dos infetados e 98.74 por cento dos saudáveis bem classificados;
- Para uma taxa de prevalência igual a 1 por cento, a dimensão ótima do grupo é 11, realizando-se (em média) 23.46 testes para classificar 100 indivíduos, sendo 90.15 por cento dos infetados e 99.33 por cento dos saudáveis bem classificados;

- Para uma taxa de prevalência igual a 0.5 por cento, a dimensão ótima do grupo é 15, realizando-se (em média) 18.23 testes para classificar 100 indivíduos, sendo 90.40 por cento dos infetados e 99.44 por cento dos saudáveis bem classificados;
- Para uma taxa de prevalência igual a 0.1 por cento, a dimensão ótima do grupo é 32, realizando-se (em média) 10.95 testes para classificar 100 indivíduos, sendo 90.51 por cento dos infetados e 99.61 por cento dos saudáveis bem classificados.

Em resumo, a utilização da metodologia de Dorfman na presença de erros de classificação tem a mesma dimensão ótima para os grupos que a mesma metodologia na ausência de problemas de classificação, obtendo-se um custo relativo semelhante (mas ligeiramente superior), sendo a sensibilidade da metodologia inferior à sensibilidade individual e a especificidade da metodologia superior à especificidade individual.

3.2 – M2 – Metodologias hierárquicas

A metodologia hierárquica corresponde a uma generalização da metodologia de Dorfman de forma a serem obtidos métodos mais eficientes. Esta metodologia é idêntica à metodologia de Dorfman pois apresenta o mesmo comportamento exceto quando o resultado do teste composto é positivo. Desta forma, para a metodologia hierárquica, a população é dividida em grupos de n indivíduos sendo retirada uma amostra de sangue a cada um dos indivíduos de cada grupo. Todas as amostras de sangue de cada grupo de indivíduos são combinadas e submetidas ao teste qualitativo composto (conforme previamente descrito). Se o resultado do teste qualitativo composto for negativo nenhum dos indivíduos do grupo está infetado (e tudo decorre tal como analisado na metodologia de Dorfman); se o resultado do teste qualitativo composto for positivo implicará que há pelo menos um indivíduo do grupo que está infetado, sendo então esse grupo subdividido em novos grupos de dimensão mais pequena para a realização de testes compostos em cada um desses subgrupos. Deste modo, voltam a ser realizados as misturas homogêneas sobre os quais são realizados novos testes compostos até se conseguir classificar todos os indivíduos da população. Assim podemos estruturar esta metodologia hierárquica do seguinte modo:

- Seleção da população e subdivisão da população em grupos de n indivíduos. Para cada grupo é obtida uma mistura homogênea do sangue dos indivíduos que compõem cada grupo (sangue composto) e realiza-se um teste composto:
 - Se o teste composto for negativo então todos os indivíduos são classificados como não infectados;
 - Se o teste composto for positivo então subdividem-se o grupo em subgrupos e volta-se a repetir o processo até, caso seja necessário, realizarem-se testes individuais.

Assim sendo, vamos repetir a divisão dos grupos infectados até termos testes individuais (únicos que permitem identificar se um indivíduo está realmente infectado). Notemos ainda que a metodologia de Dorfman corresponde a um caso particular, no qual, num teste composto positivo, o grupo de n indivíduos é dividido em n grupos de um indivíduo (passamos a ter logo testes individuais). No caso que vamos analisar (uma vez que há muitas formas de subdividir os grupos positivos em subgrupos, pois não é obrigatório que a divisão seja feita em unicamente dois grupos) será dividir em dois grupos com a mesma dimensão (se n for ímpar um dos subgrupos terá mais um indivíduo do que o outro subgrupo). Este caso particular das metodologias hierárquicas será representado por $M2$ e corresponderá a uma das metodologias em estudo via simulação.

O Quadro 6 apresenta os valores obtidos (via simulação utilizando 1000 populações com 10000 indivíduos cada, conforme programação apresentada no Anexo 8) na aplicação da metodologia hierárquica $M2$ para o custo relativo utilizando diferentes valores da taxa de prevalência bem como da dimensão do grupo. São apresentados os valores da média e do desvio padrão (que é exposto entre parêntesis).

Deste modo, podemos apresentar as seguintes conclusões relativas à utilização da metodologia hierárquica $M2$ com base nos resultados obtidos por simulação apresentados no Quadro 6 (que são muito semelhantes ao que tínhamos observado na metodologia de Dorfman):

- Os valores dos desvios padrões são igualmente diminutos pelo que as estimativas obtidas são fiáveis (os valores obtidos nas 1000 populações são próximos);

	$p = 0,1$	$p = 0,05$	$p = 0,01$	$p = 0,005$	$p = 0,001$
$n = 2$	79,03 (2,05)	64,84 (1,30)	53,03 (0,68)	51,44 (0,51)	50,27 (0,21)
$n = 3$	67,36 (1,99)	50,99 (1,50)	39,93 (0,64)	35,10 (0,42)	33,68 (0,19)
$n = 4$	61,30 (1,69)	44,02 (1,40)	28,93 (0,60)	26,88 (0,45)	25,38 (0,20)
$n = 5$	60,09 (1,62)	41,58 (1,14)	24,58 (0,56)	22,26 (0,49)	20,46 (0,20)
$n = 6$	59,31 (1,42)	39,77 (1,30)	21,54 (0,56)	19,21 (0,45)	17,16 (0,21)
$n = 7$	60,59 (1,36)	39,75 (1,22)	19,80 (0,71)	16,99 (0,43)	14,83 (0,18)
$n = 8$	61,32 (1,54)	39,31 (1,30)	18,23 (0,70)	15,50 (0,52)	13,11 (0,20)
$n = 9$	62,84 (1,41)	40,05 (1,27)	17,50 (0,62)	14,33 (0,47)	11,75 (0,21)
$n = 10$	63,79 (1,42)	40,31 (1,24)	16,93 (0,71)	13,40 (0,47)	10,67 (0,20)
$n = 11$	65,63 (1,40)	41,63 (1,33)	16,40 (0,74)	12,78 (0,59)	9,84 (0,26)
$n = 12$	67,30 (1,36)	42,41 (1,23)	15,99 (0,66)	12,27 (0,58)	9,09 (0,21)
$n = 15$	72,03 (1,09)	45,80 (1,18)	15,77 (0,71)	11,34 (0,57)	7,64 (0,22)
$n = 25$	84,76 (0,96)	57,35 (1,38)	17,54 (0,82)	10,95 (0,65)	5,44 (0,25)
$n = 50$	98,75 (0,58)	78,10 (1,05)	25,87 (0,99)	14,69 (0,73)	4,66 (0,40)

Quadro 6: Custos relativos obtidos na simulação da metodologia hierárquica sem erros de classificação

- O custo relativo diminui à medida que diminui a taxa de prevalência (observação dos valores situados na mesma linha do Quadro 6);
- Para cada taxa de prevalência a eficiência aumenta até certo ponto à medida que vamos aumentando a dimensão do grupo (ponto que corresponde à dimensão ótima) e depois começa a diminuir (análise dos valores situados na mesma coluna do Quadro 6);

- Todos os custos relativos determinados são inferiores a 100 por cento, pelo que é (para os valores dos parâmetros analisados) sempre preferível utilizar a metodologia *M2* do que testes individuais (tendo em consideração que só foram utilizadas taxas de prevalência baixas).

Deste modo concluímos (por simulação) que utilizando o valor ótimo para a dimensão do grupo em termos de eficiência na metodologia hierárquica *M2* obtemos os valores patentes no Quadro 7.

	$p = 0,1$	$p = 0,05$	$p = 0,01$	$p = 0,005$	$p = 0,001$
n	6	8	16	22	45
Custo relativo	59,31	39,31	15,73	10,77	4,60

Quadro 7: Valores ótimos utilizando a metodologia hierárquica

Desta forma, a interpretação dos valores patentes no Quadro 7 pode ser realizada da seguinte forma:

- Para uma dimensão de grupos constituído por seis indivíduos e com uma taxa de prevalência de 0.1, o custo relativo resultou num valor de 59.31% o que significa que foram necessários 59.31% dos testes para identificar os indivíduos infetados (comparativamente com a realização os testes individuais) sendo conseguido um ganho de 40.69% (100%-59.31%) dos testes na execução dos testes compostos;
- Para uma dimensão de grupos constituído por oito indivíduos e com uma taxa de prevalência de 0.05, o custo relativo resultou num valor de 39.31% o que significa que foram necessários 39.31% dos testes para identificar os indivíduos infetados (comparativamente com a realização os testes individuais) sendo conseguido um ganho de 60.69% (100%-39.31%) dos testes na execução dos testes compostos;
- Para uma dimensão de grupos constituído por dezasseis indivíduos e com uma taxa de prevalência de 0.01, o custo relativo resultou num valor de 15.73% o que significa que foram necessários 15.73% dos testes para identificar os indivíduos infetados (comparativamente com a realização os testes individuais) sendo conseguido um ganho de 84.27% (100%-15.73%) dos testes na execução dos testes compostos;

- Para uma dimensão de grupos constituído por vinte e dois indivíduos e com uma taxa de prevalência de 0.005, o custo relativo resultou num valor de 10.77% o que significa que foram necessários 10.77% dos testes para identificar os indivíduos infetados (comparativamente com a realização os testes individuais) sendo conseguido um ganho de 89.23% (100%-10.77%) dos testes na execução dos testes compostos;
- Para uma dimensão de grupos constituído por quarenta e cinco indivíduos e com uma taxa de prevalência de 0.001, o custo relativo resultou num valor de 4.60% o que significa que foram necessários 4.60% dos testes para identificar os indivíduos infetados (comparativamente com a realização os testes individuais) sendo conseguido um ganho de 95.40% (100%-4.60%) dos testes na execução dos testes compostos.

No Quadro 8 são apresentados os valores obtidos por simulação considerando a possibilidade de ocorrência de erros de classificação, supondo que a especificidade e a sensibilidade individual e composta assumem o valor 0.95 (erros de classificação de 5 por cento). Assim, os resultados obtidos nestas simulações permitem tirar as seguintes conclusões em relação à metodologia hierárquica *M2*:

- O valor ótimo para a dimensão do grupo (em termos de eficiência) na aplicação da metodologia hierárquica *M2* é (quase sempre) igual na presença e na ausência de erros de classificação;
- A eficiência é muito próxima com e sem erros de classificação (havendo casos com maior eficiência e casos com menor eficiência);
- Tal como acontecia na metodologia de Dorfman, a sensibilidade da metodologia hierárquica *M2* (próxima dos 85 por cento) é sempre inferior à sensibilidade individual e composta considerada e a especificidade da metodologia hierárquica *M2* é sempre superior à especificidade individual e composta utilizada. Saliente-se que, enquanto na metodologia de Dorfman há fórmulas que permitiriam obter estes valores, nesta metodologia (bem como nas retratadas nos subcapítulos 3.3 e 3.4) não há, pelo que a única forma é obter valores aproximados via simulação;

N		$p = 0,1$	$p = 0,05$	$p = 0,01$	$p = 0,005$	$p = 0,001$
3	<i>CR</i>	67,88 (1,82)	53,10 (1,61)	40,29 (0,80)	38,62 (0,66)	37,28 (0,46)
	<i>SM</i>	85,73 (1,81)	85,42 (2,74)	86,04 (6,37)	87,02 (8,73)	86,73 (9,47)
	<i>EM</i>	99,65 (0,10)	99,83 (0,07)	99,96 (0,04)	99,97 (0,03)	99,98 (0,03)
5	<i>CR</i>	59,11 (1,56)	41,99 (1,27)	26,40 (0,75)	24,39 (0,52)	22,65 (0,36)
	<i>SM</i>	85,80 (1,74)	85,56 (2,25)	85,77 (5,28)	84,50 (7,47)	86,94 (10,47)
	<i>EM</i>	99,23 (0,14)	99,60 (0,88)	99,91 (0,05)	99,95 (0,03)	99,98 (0,02)
10	<i>CR</i>	61,00 (1,39)	39,60 (1,22)	17,67 (0,71)	14,47 (0,51)	11,88 (29,76)
	<i>SM</i>	85,51 (1,31)	85,53 (1,78)	86,70 (3,33)	85,17 (5,45)	87,56 (11,64)
	<i>EM</i>	98,40 (0,14)	99,10 (0,09)	99,80 (0,05)	99,88 (0,04)	99,97 (0,02)
15	<i>CR</i>	67,36 (1,35)	43,74 (1,13)	16,14 (0,79)	11,89 (0,55)	8,49 (0,29)
	<i>SM</i>	85,73 (1,15)	85,99 (1,62)	85,63 (2,74)	85,55 (4,30)	85,29 (10,09)
	<i>EM</i>	97,72 (0,15)	98,67 (0,11)	99,69 (0,06)	99,83 (0,04)	99,96 (0,02)
25	<i>CR</i>	78,20 (1,15)	53,66 (1,28)	17,48 (0,81)	11,16 (0,07)	5,99 (0,36)
	<i>SM</i>	85,80 (0,98)	85,69 (1,38)	86,00 (2,08)	85,67 (3,03)	85,45 (7,64)
	<i>EM</i>	96,77 (0,11)	97,92 (0,11)	99,48 (0,06)	99,73 (0,04)	99,93 (0,02)

Quadro 8: Eficiência e fiabilidade obtidas na simulação da metodologia hierárquica considerando erros de classificação de 5 por cento

- Em termos gerais a metodologia hierárquica *M2* tem um comportamento semelhante à da metodologia de Dorfman, tendo uma perda significativa em termos da sensibilidade e um ganho ligeiro na especificidade.

Através das simulações realizadas para a apresentação do Quadro 8 e de outras análogas (programação apresentada no Anexo 8), concluímos, para a aplicação da metodologia hierárquica *M2*, a dimensão ótima para cada uma das taxas de prevalência investigadas nos casos em que os erros de classificação são iguais a 5 por cento (sensibilidade e especificidade individuais e compostas iguais a 95 por cento). Estes valores são os expostos no Quadro 9.

	$p = 0,1$	$p = 0,05$	$p = 0,01$	$p = 0,005$	$p = 0,001$
n	6	8	16	22	45
CR	58,21	38,85	16,16	11,18	4,95
SM	85,64	85,54	85,88	85,99	86,11
EM	98,88	99,30	99,65	99,75	99,89

Quadro 9: Valores ótimos, com erros de classificação de 5%, utilizando a metodologia hierárquica

Os valores patentes no Quadro 9, referentes à aplicação da metodologia hierárquica (*M2*) podem ser interpretados do seguinte modo:

- Para uma taxa de prevalência igual a 10 por cento, a dimensão ótima do grupo é 6, realizando-se (em média) 58.21 testes para classificar 100 indivíduos, sendo 85.64 por cento dos infetados e 98.88 por cento dos saudáveis bem classificados;
- Para uma taxa de prevalência igual a 5 por cento, a dimensão ótima do grupo é 8, realizando-se (em média) 38.85 testes para classificar 100 indivíduos, sendo 85.54 por cento dos infetados e 99.30 por cento dos saudáveis bem classificados;
- Para uma taxa de prevalência igual a 1 por cento, a dimensão ótima do grupo é 16, realizando-se (em média) 16.16 testes para classificar 100 indivíduos, sendo 85.88 por cento dos infetados e 99.65 por cento dos saudáveis bem classificados;
- Para uma taxa de prevalência igual a 0.5 por cento, a dimensão ótima do grupo é 22, realizando-se (em média) 11.18 testes para classificar 100 indivíduos, sendo 85.99 por cento dos infetados e 99.75 por cento dos saudáveis bem classificados;

- Para uma taxa de prevalência igual a 0.1 por cento, a dimensão ótima do grupo é 45, realizando-se (em média) 4.95 testes para classificar 100 indivíduos, sendo 86.11 por cento dos infetados e 99.89 por cento dos saudáveis bem classificados.

Em suma, através da aplicação da metodologia hierárquica *M2* na presença de erros de classificação obtemos a mesma dimensão ótima para os grupos que no caso da inexistência de problemas de classificação, obtendo-se um custo relativo semelhante, uma sensibilidade razoavelmente inferior à sensibilidade individual e uma especificidade ligeiramente superior à especificidade individual.

3.3 – *M3* – Metodologias baseadas em *arrays* sem *master pool*

A metodologia baseada em *arrays* que iremos utilizar é baseada no recurso a matrizes de dimensão $n \times n$ (com n linhas e n colunas) que iremos representar por *M3* (contudo salientemos que na aplicação desta metodologia não é obrigatório ser utilizado o mesmo número de linhas e de colunas, isto é, não é restrito a matrizes quadradas). Deste modo, os indivíduos que formam o grupo são colocados em matrizes sendo cada posição ocupada por um indivíduo (e, portanto, teremos n^2 indivíduos na matriz). Após colocados todos os indivíduos em matrizes, são realizados testes compostos aos grupos de dimensão n que são definidos pelos indivíduos que estão na mesma linha (e teremos aqui n grupos, cada um constituído por n indivíduos) bem como aos indivíduos que pertencem à mesma coluna (onde teremos mais n grupos constituídos por n indivíduos cada). Deste modo, vão ser realizados $2n$ testes compostos. Após terem sido realizados todos os $2n$ testes compostos (de todas as linhas e colunas da matriz), conseguimos identificar as linhas e/ou colunas nas quais nenhum indivíduo está infetado (correspondente teste negativo) ou se há pelo menos um indivíduo infetado (correspondente teste positivo). Notemos que cada indivíduo pertence a dois grupos, ao da linha correspondente e ao da coluna correspondente. Sendo desta forma possível identificar as linhas e colunas positivas que identificam (na sua intersecção) potenciais indivíduos infetados (pois mesmo que a linha e coluna correspondente sejam positivas poderá não ser suficiente para garantir que o indivíduo esteja infetado). Nestes casos, terão de ser realizados testes individuais de forma a ser

possível classificar o indivíduo. De forma a clarificar esta ideia, seguem a ilustração de dois casos distintos. No primeiro exemplo considerou-se uma matriz com um único indivíduo infetado (o 0 representa um indivíduo saudável e o 1 um indivíduo infetado).

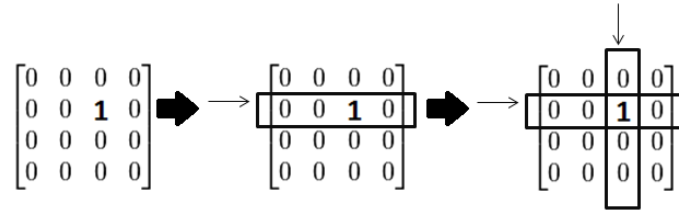


Figura 3: Metodologia baseada em *arrays* com um indivíduo infetado

Neste caso (considerando que não há problemas de má classificação) unicamente o grupo da segunda linha e o grupo da terceira coluna serão positivos (todas as restantes linhas e colunas terão testes compostos negativos), pelo que não haverá dúvidas que o indivíduo situado simultaneamente na linha 2 e na coluna 3 está infetado. No segundo exemplo colocamos uma matriz com dois indivíduos infetados.

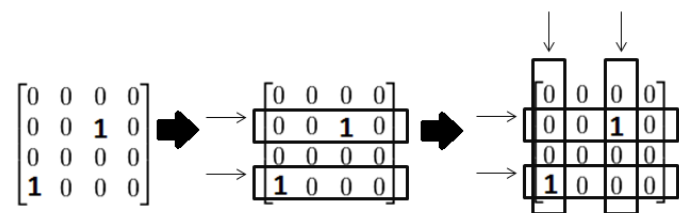


Figura 4: Metodologia baseada em *arrays* com dois indivíduos infetado

Neste último caso (supondo igualmente que não há erros de má classificação) teremos os grupos da segunda e da quarta linha bem como os grupos da primeira e da terceira coluna com testes compostos positivos (as restantes linhas e colunas terão testes compostos negativos), pelo que há quatro indivíduos situados nas intersecções destas linhas e colunas não havendo a possibilidade de classificar os quatro indivíduos com esta informação (nem sequer haverá informação suficiente para podermos concluir o número exato de indivíduos infetados na matriz, pois poderiam existir 2, 3 ou 4 indivíduos infetados). Por este motivo, nestas situações, há a necessidade de realizar testes individuais aos quatro indivíduos que estão nas localizações consideradas de potenciais indivíduos infetados. Assim podemos estruturar esta metodologia baseada em *arrays* do seguinte modo:

- Seleção da população e subdivisão da população em grupos de n^2 indivíduos sendo cada grupo distribuído por uma matriz $n \times n$ (cada indivíduo ocupa um lugar da

matriz). Para cada matriz são obtidos os n sangues compostos dos grupos constituídos pelos elementos da mesma linha e os n sangues compostos dos grupos constituídos pelos elementos da mesma coluna e realizam-se os $2n$ testes compostos:

- Todos os indivíduos situados em linhas com resultado do teste composto negativo ou em colunas com resultado do teste composto negativo são classificados como saudáveis;
- Realizam-se testes individuais a todos os indivíduos potencialmente infectados (com a correspondente linha e a correspondente coluna com testes compostos positivos).

No Quadro 10 são apresentados os resultados obtidos nas simulações realizadas através da aplicação da metodologia baseada em *arrays* (programação apresentada no Anexo 9), considerando 1000 populações constituídas por 1000 matrizes (isto é, a dimensão da população dependeu da dimensão a ser utilizada nas matrizes, de forma a termos o mesmo número de matrizes em cada uma das simulações), diferentes taxas de prevalência bem como dimensões para as matrizes. Tal como nas tabelas anteriores, são apresentados os valores da média e do desvio padrão (entre parêntesis).

Através da análise aos valores expostos no Quadro 10, podemos tirar as seguintes conclusões em relação à aplicação da metodologia baseada em *arrays* de dimensão $n \times n$:

- Os valores dos desvios padrões são mais baixos que os obtidos anteriormente, garantindo a fiabilidade das estimativas obtidas uma vez que os valores obtidos nas 1000 populações são quase iguais (a diminuição do desvio padrão está associado ao aumento da dimensão da população uma vez que, neste caso, considerou-se populações de dimensão $1000 \times n^2$);
- Tal como nas metodologias anteriormente analisadas, o custo relativo diminui à medida que diminui a taxa de prevalência (observação dos valores situados na mesma linha do Quadro 10), o que corresponde a uma das características conhecidas deste tipo de metodologias baseadas em testes compostos, quanto menor for a taxa de prevalência maior será a poupança que pode ser realizada;

	$p = 0,1$	$p = 0,05$	$p = 0,01$	$p = 0,005$	$p = 0,001$
$n = 2$	110,86 (2,16)	105,24 (1,40)	101,03 (0,50)	100,57 (0,39)	100,09 (0,15)
$n = 3$	79,64 (1,51)	72,52 (1,09)	67,72 (0,36)	67,15 (0,24)	66,76 (0,11)
$n = 4$	66,88 (1,98)	56,92 (1,06)	51,07 (0,29)	50,52 (0,21)	55,12 (0,09)
$n = 5$	60,72 (1,78)	48,43 (1,12)	41,14 (0,29)	40,54 (0,17)	40,10 (0,07)
$n = 6$	58,41 (1,61)	43,23 (1,07)	34,53 (0,24)	33,91 (0,16)	33,44 (0,06)
$n = 7$	58,38 (1,65)	40,13 (1,18)	29,98 (0,29)	29,13 (0,13)	28,67 (0,04)
$n = 8$	59,41 (1,64)	38,46 (1,11)	26,45 (0,26)	25,60 (0,14)	25,10 (0,05)
$n = 9$	61,60 (1,64)	38,08 (1,06)	23,82 (0,24)	22,89 (0,13)	22,33 (0,04)
$n = 10$	63,54 (1,85)	37,94 (1,15)	21,74 (0,24)	20,71 (0,14)	20,12 (0,04)
$n = 11$	66,51 (1,64)	48,49 (1,23)	20,07 (0,24)	18,92 (0,14)	18,30 (0,03)
$n = 12$	69,04 (1,71)	39,36 (1,34)	18,72 (0,26)	17,44 (0,15)	16,78 (0,03)
$n = 15$	76,99 (1,25)	43,02 (1,34)	16,01 (0,31)	14,3 (0,15)	13,45 (0,04)
$n = 25$	94,20 (0,71)	40,51 (0,99)	13,57 (0,36)	9,81 (0,16)	8,16 (0,03)
$n = 50$	102,97 (0,15)	89,26 (0,56)	19,97 (0,54)	9,21 (0,28)	4,33 (0,04)

Quadro 10: Custos relativos obtidos na simulação da metodologia baseada em *arrays* sem master pool nem erros de classificação

- Também de forma semelhante à das metodologias anteriormente analisadas, para uma taxa de prevalência fixa a eficiência vai aumentando com a dimensão do grupo até certo ponto (até atingirmos a dimensão ótima) e depois começa a diminuir (análise dos valores situados na mesma coluna do Quadro 10);
- Para os valores baixos da taxa de prevalência (valores analisados) é mais eficiente utilizar a metodologia *M3* do que testes individuais.

Para cada uma das taxas de prevalência em análise apresentamos no Quadro 11 os valores do custo relativo para a dimensão de grupo que maximiza a eficiência na aplicação da metodologia *M3*.

	$p = 0,1$	$p = 0,05$	$p = 0,01$	$p = 0,005$	$p = 0,001$
n	7	10	25	38	106
Custo relativo	58,38	37,94	13,57	8,62	2,97

Quadro 11: Valores ótimos utilizando a metodologia baseada em *arrays* sem master pool

Os valores no Quadro 11 podem ser interpretados do seguinte modo:

- Para uma dimensão de grupos constituído por sete indivíduos e com uma taxa de prevalência de 0.1, o custo relativo resultou num valor de 58.38% o que significa que foram necessários 58.38% dos testes para identificar os indivíduos infetados (comparativamente com a realização os testes individuais) sendo conseguido um ganho de 41.62% (100%-58.38%) dos testes na execução dos testes compostos;
- Para uma dimensão de grupos constituído por dez indivíduos e com uma taxa de prevalência de 0.05, o custo relativo resultou num valor de 37.94% o que significa que foram necessários 37.94% dos testes para identificar os indivíduos infetados (comparativamente com a realização os testes individuais) sendo conseguido um ganho de 62.06% (100%-37.94%) dos testes na execução dos testes compostos;
- Para uma dimensão de grupos constituído por vinte e cinco indivíduos e com uma taxa de prevalência de 0.01, o custo relativo resultou num valor de 13.57% o que significa que foram necessários 13.57% dos testes para identificar os indivíduos infetados (comparativamente com a realização os testes individuais) sendo conseguido um ganho de 86.43% (100%-13.57%) dos testes na execução dos testes compostos;
- Para uma dimensão de grupos constituído por trinta e oito indivíduos e com uma taxa de prevalência de 0.005, o custo relativo resultou num valor de 8.62% o que significa que foram necessários 8.62% dos testes para identificar os indivíduos infetados (comparativamente com a realização os testes individuais) sendo

conseguido um ganho de 91.38% (100%-8.62%) dos testes na execução dos testes compostos;

- Para uma dimensão de grupos constituído por cento e seis indivíduos e com uma taxa de prevalência de 0.001, o custo relativo resultou num valor de 2.97% o que significa que foram necessários 2.97% dos testes para identificar os indivíduos infetados (comparativamente com a realização os testes individuais) sendo conseguido um ganho de 97.03% (100%-2.97%) dos testes na execução dos testes compostos.

No Quadro 12 são apresentados os valores obtidos na aplicação da metodologia *M3* via simulação (recorrendo ao programa apresentado no Anexo 9) considerando a possibilidade de ocorrência de erros de classificação (erros de classificação de 5 por cento, isto é, especificidade e sensibilidade individual e composta iguais a 0.95). De referir que com a inclusão da possibilidade de erros de classificação a interpretação dos resultados dos testes conjuntos aos indivíduos situados na mesma linha e dos testes conjuntos aos indivíduos localizados na mesma coluna da matriz tornam-se ainda mais complexas, pois neste caso poderá ocorrer situações nas quais nenhuma linha apresenta um teste composto positivo mas há uma coluna com um teste composto positivo (neste caso teremos a certeza de que houve pelo menos um erro de classificação: ou houve uma linha que deveria ter resultado positivo e não o teve, ou a coluna que obteve teste composto positivo corresponde a um falso positivo). Independentemente de qual foi o erro que ocorreu, tal resultado (ou outros semelhantes) fazem a definição de indivíduos potencialmente infetados mais dúbia. Neste trabalho optamos por apostar em tentar ganhar um pouco de fiabilidade à custa de perder um pouco da eficiência considerando, no caso descrito, que todos os elementos da coluna que obteve um teste composto positivo como potenciais infetados, pelo que todos esses indivíduos serão testados individualmente.

Nestas simulações, além de ser considerada a eficiência do processo, serão igualmente apresentadas a sensibilidade e a especificidade da metodologia *M3*. Deste modo, os resultados obtidos nestas simulações (Quadro 12) permitem tirar as seguintes conclusões em relação à metodologia baseada em *arrays M3*:

n		$p = 0,1$	$p = 0,05$	$p = 0,01$	$p = 0,005$	$p = 0,001$
3	CR	84,39 (1,70)	79,57 (1,74)	75,99 (1,40)	75,68 (1,61)	75,62 (1,51)
	SM	90,27 (3,07)	92,38 (4,05)	92,78 (8,69)	92,25 (9,36)	92,42 (10,26)
	EM	99,57 (0,25)	99,56 (0,24)	99,58 (0,20)	99,56 (0,22)	99,55 (0,22)
5	CR	61,27 (1,80)	51,77 (1,07)	48,01 (0,96)	48,04 (1,10)	48,06 (1,14)
	SM	86,74 (2,37)	88,82 (2,62)	91,28 (5,89)	91,33 (8,51)	92,03 (9,45)
	EM	99,33 (0,18)	99,61 (0,14)	99,63 (0,12)	99,62 (0,14)	99,57 (0,15)
10	CR	60,85 (1,54)	38,13 (1,20)	25,09 (0,55)	24,92 (0,56)	25,99 (0,76)
	SM	85,85 (1,31)	85,97 (1,85)	88,12 (3,69)	89,39 (4,37)	91,30 (9,62)
	EM	98,23 (0,16)	99,28 (0,10)	99,79 (0,05)	99,77 (0,05)	99,71 (0,07)
15	CR	71,64 (1,19)	41,95 (1,13)	17,52 (0,37)	16,73 (0,35)	17,63 (0,53)
	SM	85,81 (1,01)	85,69 (1,14)	86,84 (2,23)	87,13 (2,80)	89,52 (5,93)
	EM	97,27 (0,13)	98,72 (0,09)	99,84 (0,03)	99,85 (0,03)	99,79 (0,04)
25	CR	86,13 (0,87)	56,77 (0,98)	14,55 (0,41)	10,84 (0,24)	10,14 (0,35)
	SM	85,81 (0,75)	85,74 (0,85)	85,95 (1,64)	86,12 (1,85)	86,94 (4,21)
	EM	96,16 (0,09)	97,67 (0,08)	99,71 (0,03)	99,88 (0,02)	99,90 (0,02)

Quadro 12: Eficiência e fiabilidade obtidas na simulação da metodologia baseada em *arrays* sem master pool considerando erros de classificação de 5 por cento

- Uma vez mais, o valor ótimo (em termos de eficiência) não parecer ser influenciado pela presença de erros de classificação;
- Os valores observados para a eficiência são muito próximos aos verificados na ausência de erros de classificação (sendo na maioria dos casos ligeiramente

inferior);

- Em termos gerais a metodologia hierárquica *M3* tem um comportamento semelhante às anteriores metodologias (*M1* e *M2*), tendo uma perda significativa em termos da sensibilidade (dos 95 por cento para cerca de 85 por cento) e um ganho ligeiro na especificidade (ganho que depende da dimensão do grupo e da taxa de prevalência).

Através das simulações realizadas recorrendo à programação apresentada no Anexo 9 (sendo que alguns dos resultados estão apresentados do Quadro 12) concluímos a dimensão ótima para cada uma das taxas de prevalência na aplicação da metodologia baseada em *arrays M3*, quando os erros de classificação são iguais a 5 por cento (sensibilidade e especificidade individuais e compostas iguais a 95 por cento). Para estas dimensões apresentamos no Quadro 13 os valores das medidas associadas à eficiência e à fiabilidade da metodologia *M3*.

	$p = 0,1$	$p = 0,05$	$p = 0,01$	$p = 0,005$	$p = 0,001$
n	7	10	25	38	106
CR	57,11	38,13	14,55	9,64	3,87
SM	85,95	85,97	85,95	86,03	85,84
EM	98,90	99,28	99,71	99,80	99,91

Quadro 13: Valores ótimos, com erros de classificação de 5%, utilizando a metodologia baseada em *arrays*

Os valores patentes no Quadro 13, referentes à aplicação da metodologia baseada em *arrays* sem *master pool (M3)* podem ser interpretados do seguinte modo:

- Para uma taxa de prevalência igual a 10 por cento, a dimensão ótima da matriz é 7×7 , realizando-se (em média) 57.11 testes para classificar 100 indivíduos, sendo 85.95 por cento dos infetados e 98.90 por cento dos saudáveis bem classificados;
- Para uma taxa de prevalência igual a 5 por cento, a dimensão ótima da matriz é 10×10 , realizando-se (em média) 38.13 testes para classificar 100 indivíduos, sendo 85.97 por cento dos infetados e 99.28 por cento dos saudáveis bem classificados;

- Para uma taxa de prevalência igual a 1 por cento, a dimensão ótima da matriz é 25×25 , realizando-se (em média) 14.55 testes para classificar 100 indivíduos, sendo 85.95 por cento dos infetados e 99.71 por cento dos saudáveis bem classificados;
- Para uma taxa de prevalência igual a 0.5 por cento, a dimensão ótima da matriz é 38×38 , realizando-se (em média) 9.64 testes para classificar 100 indivíduos, sendo 86.03 por cento dos infetados e 99.80 por cento dos saudáveis bem classificados;
- Para uma taxa de prevalência igual a 0.1 por cento, a dimensão ótima da matriz é 106×106 , realizando-se (em média) 3.87 testes para classificar 100 indivíduos, sendo 85.84 por cento dos infetados e 99.91 por cento dos saudáveis bem classificados.

Deste modo, podemos uma vez mais concluir que na aplicação da metodologia baseada em *arrays* sem a utilização do *master pool* (metodologia *M3*), a dimensão do grupo não é influenciada pela presença de erros de classificação obtendo-se igualmente um custo relativo semelhante. No que concerne à fiabilidade, a sensibilidade é claramente inferior à sensibilidade individual e a especificidade é ligeiramente superior à especificidade individual.

3.4 – *M4* – Metodologias baseadas em *arrays* com *master pool*

A metodologia baseada *arrays* com (a utilização de) *master pool* corresponde a uma generalização da metodologia baseada em *arrays* (explicada na subsecção anterior) na qual é ainda realizado um teste composto prévio aos n^2 indivíduos que compõem a matriz (teste que é denominado na bibliografia da área por *master pool*). Esta metodologia irá ser representada por *M4*. Assim, caso o *master pool* resulte negativo todos os n^2 indivíduos são classificados com saudáveis; caso o *master pool* tenha um resultado positivo os n^2 indivíduos são posicionados numa matriz $n \times n$ e o processo decorre conforme a metodologia baseada em *arrays* apresentada na secção anterior. Deste modo podemos esquematizar esta metodologia do seguinte modo:

- Seleção da população e subdivisão da população em grupos de n^2 indivíduos. Para

cada grupo é obtida uma mistura homogênea do sangue dos indivíduos que compõem cada grupo (sangue composto) e realiza-se um teste composto:

- Se o teste composto for negativo então todos os indivíduos são classificados como saudáveis (não infectados);
- Se o teste composto for positivo então o grupo é distribuído numa matriz $n \times n$ (cada indivíduo ocupa um lugar da matriz). Para cada matriz são obtidos os n sangues compostos dos grupos constituídos pelos elementos da mesma linha e os n sangues compostos dos grupos constituídos pelos elementos da mesma coluna e realizam-se os $2n$ testes compostos:
 - Todos os indivíduos situados em linhas com teste composto negativo ou em colunas com teste composto negativo são classificados como saudáveis;
 - Realizam-se testes individuais a todos os indivíduos potencialmente infectados (com a correspondente linha e a correspondente coluna com testes compostos positivos).

No caso de existência de erros de classificação, este esquema deverá ser alterado devido à ambiguidade existente em algumas situações (como, por exemplo, a existência de linhas com testes compostos positivos e simultaneamente a ausência de colunas com testes compostos positivos). Tal como no subcapítulo anterior, nestes casos optamos por alterar estes dois últimos itens, considerando que em caso de ambiguidade, todos os indivíduos situados nessas linhas serão testados individualmente. Todavia, apesar de estarmos a privilegiar a fiabilidade podendo perder um pouco de eficiência, estes casos muito provavelmente serão residuais, pelo que qualquer que seja a opção tomada esta não deverá ter um impacto substancial nos resultados finais.

No Quadro 14 são apresentados os resultados obtidos nas simulações realizadas através da aplicação da metodologia baseada em *arrays* com a utilização de *master pool* (metodologia *M4* cuja programação é apresentada no Anexo 10), considerando 1000 populações constituídas por 1000 matrizes (tal como na metodologia anterior, fixamos o número de matrizes em cada uma das simulações fazendo a dimensão da população depender da dimensão utilizada nas matrizes), diferentes taxas de prevalência bem como dimensões

para as matrizes. Nesta tabela serão apresentados os valores da média e do desvio padrão (apresentado entre parêntesis).

	$p = 0,1$	$p = 0,05$	$p = 0,01$	$p = 0,005$	$p = 0,001$
$n = 2$	70,44 (1,99)	48,47 (4,70)	29,96 (2,13)	27,53 (1,99)	25,55 (0,78)
$n = 3$	64,91 (1,37)	41,33 (4,37)	18,09 (2,39)	14,58 (1,67)	11,92 (0,80)
$n = 4$	63,51 (0,94)	41,22 (3,13)	14,01 (1,84)	10,57 (1,43)	7,11 (0,67)
$n = 5$	61,70 (0,64)	41,08 (2,35)	14,09 (1,64)	9,38 (1,42)	5,18 (0,77)
$n = 6$	60,47 (0,60)	40,89 (1,87)	14,26 (1,54)	8,85 (1,40)	4,02 (0,68)
$n = 7$	60,20 (0,58)	39,96 (1,41)	14,57 (1,58)	8,89 (1,27)	3,51 (0,68)
$n = 8$	61,12 (0,56)	39,17 (1,44)	14,89 (1,15)	9,09 (1,15)	3,27 (0,64)
$n = 9$	62,60 (0,49)	38,86 (1,11)	15,16 (1,32)	9,15 (1,04)	3,04 (0,62)
$n = 10$	64,76 (0,51)	38,81 (1,12)	15,29 (1,07)	9,60 (1,13)	3,02 (0,60)
$n = 11$	67,25 (0,47)	39,25 (1,16)	15,66 (1,02)	9,76 (1,06)	3,00 (0,64)
$n = 12$	69,63 (0,45)	40,08 (1,18)	15,52 (0,85)	10,10 (0,87)	3,04 (0,63)
$n = 15$	77,35 (0,39)	43,77 (1,28)	15,06 (0,56)	10,26 (0,63)	3,31 (0,48)
$n = 25$	94,25 (0,75)	60,78 (1,16)	13,69 (0,42)	9,56 (0,29)	4,03 (0,44)
$n = 50$	103,01 (0,13)	89,23 (0,61)	19,99 (0,55)	9,28 (0,25)	4,04 (0,12)

Quadro 14: Custos relativos obtidos na simulação da metodologia baseada em *arrays* com master pool sem erros de classificação

Os valores obtidos permitem tirar as seguintes conclusões em relação à aplicação da metodologia *M4* baseada em *arrays* $n \times n$ com a utilização prévia de *master pool*:

- Os valores dos desvios padrões são, regra geral, razoavelmente mais elevados que

os obtidos sem a utilização do *master pool* (*M3*), o que significa que há distintas performances na aplicação desta metodologia em 1000 populações com características idênticas (mesma dimensão, taxa de prevalência, sensibilidade e especificidade). Este aumento do desvio padrão estará ligado à utilização do *master pool*, uma vez que sendo este teste negativo toda a matriz fica classificada (todos os indivíduos são classificados como saudáveis) mas se o *master pool* for positivo ter-se-ão que realizar pelo menos $2n$ testes;

- Da observação dos valores situados na mesma linha do Quadro 14, a evolução do custo relativo é bastante semelhante à das metodologias previamente analisadas uma vez que o custo relativo diminui à medida que diminui a taxa de prevalência (verificando-se, uma vez mais, uma das usuais características das metodologias de classificação baseadas em testes compostos);
- Através da análise dos valores situados na mesma coluna do Quadro 14, podemos concluir que para uma taxa de prevalência fixa a evolução da eficiência é distinta da verificada nas restantes metodologias. Nesta metodologia o custo relativo vai diminuindo com a dimensão do grupo até certo ponto (até atingirmos uma dimensão ótima “local”) e depois começa a aumentar até determinada altura e, por fim, volta a diminuir para valores mais baixos que no primeiro ótimo local (mas continua a diminuir à medida que aumentamos o valor de n , mas esta diminuição a certa altura é praticamente insignificante). Por este motivo, no Quadro 15 são analisadas duas dimensões para o grupo para algumas das taxas de prevalência, nas quais o valor de n mais elevado obtém um custo relativo ligeiramente menor;
- Para os valores baixos da taxa de prevalência (valores analisados) é mais eficiente utilizar a metodologia *M4* do que testes individuais (no quadro há uma única exceção quando são utilizados grupos de dimensão 50 e a taxa de prevalência é igual a 10 por cento).

No Quadro 15 são apresentados, para cada taxa de prevalência, os valores do custo relativo para a dimensão de grupo ótima (ou, em alguns casos, foram consideradas duas dimensões, a primeira ótima “local” e outra consideravelmente mais elevada) na aplicação da metodologia *M4*, baseada em *arrays* com *master pool*.

	$p = 0,1$	$p = 0,05$	$p = 0,01$		$p = 0,005$		$p = 0,001$	
n	7	10	4	25	6	38	11	106
Custo relativo	60,20	38,81	14,01	13,69	8,85	8,67	3,00	2,99

Quadro 15: Valores ótimos utilizando a metodologia baseada em *arrays* com *master pool*

Numa visão geral sobre a metodologia baseada em *arrays* com *master pool*, sem a ocorrência de erros de classificação, podemos interpretar os valores apresentados no Quadro 15 do seguinte modo:

- Para uma dimensão de grupos constituído por sete indivíduos e com uma taxa de prevalência de 0.1, o custo relativo resultou num valor de 60.20% o que significa que foram necessários 60.20% dos testes para identificar os indivíduos infetados (comparativamente com a realização os testes individuais) sendo conseguido um ganho de 39.80% (100%-60.20%) dos testes na execução dos testes compostos;
- Para uma dimensão de grupos constituído por dez indivíduos e com uma taxa de prevalência de 0.05, o custo relativo resultou num valor de 38.81% o que significa que foram necessários 38.81% dos testes para identificar os indivíduos infetados (comparativamente com a realização os testes individuais) sendo conseguido um ganho de 61.19% (100%-38.81%) dos testes na execução dos testes compostos;
- Para uma dimensão de grupos entre os quatro e os vinte e cinco indivíduos com uma taxa de prevalência de 0.01, o custo relativo resultou num intervalo de 14.01% e 13.69% que significa que foram necessários entre 14.01% e 13.69% dos testes para identificar os indivíduos infetados (comparativamente com a realização os testes individuais) sendo conseguido um ganho entre 85.99% e 86.31% dos testes na execução dos testes compostos;
- Para uma dimensão de grupos entre os seis e os trinta e oito indivíduos com uma taxa de prevalência de 0.005, o custo relativo resultou num intervalo de 8.85% e 8.67% que significa que foram necessários entre 8.85% e 8.67% dos testes para identificar os indivíduos infetados (comparativamente com a realização os testes individuais) sendo conseguido um ganho entre 91.15% e 91.33% dos testes na execução dos testes compostos;
- Para uma dimensão de grupos entre os onze e os cento e seis indivíduos com uma taxa de prevalência de 0.001, o custo relativo resultou num intervalo de 3.00% e

2.99% que significa que foram necessários entre 3.00% e 2.99% dos testes para identificar os indivíduos infetados (comparativamente com a realização os testes individuais) sendo conseguido um ganho entre 97.00% e 97.01% dos testes na execução dos testes compostos.

<i>n</i>		<i>p</i> = 0,1	<i>p</i> = 0,05	<i>p</i> = 0,01	<i>p</i> = 0,005	<i>p</i> = 0,001
3	<i>CR</i>	65,04 (4,16)	44,01 (4,55)	20,76 (2,72)	17,82 (1,88)	15,83 (1,84)
	<i>SM</i>	86,01 (4,18)	87,60 (5,46)	89,74 (13,83)	88,49 (13,48)	88,34 (12,54)
	<i>EM</i>	99,73 (0,18)	99,88 (0,13)	99,97 (0,05)	99,97 (0,07)	99,96 (0,06)
5	<i>CR</i>	59,21 (2,71)	41,46 (2,94)	16,06 (2,09)	11,62 (1,62)	7,62 (1,21)
	<i>SM</i>	82,93 (3,33)	84,75 (3,62)	87,04 (6,9)	87,20 (10,02)	86,93 (11,49)
	<i>EM</i>	99,38 (0,16)	99,75 (0,11)	99,94 (0,06)	99,96 (0,04)	99,97 (0,03)
10	<i>CR</i>	58,99 (1,83)	37,14 (1,46)	16,06 (1,14)	10,47 (1,20)	4,21 (0,90)
	<i>SM</i>	81,86 (2,20)	82,09 (2,67)	83,02 (4,49)	84,92 (5,89)	86,41 (10,88)
	<i>EM</i>	98,31 (0,15)	99,32 (0,09)	99,91 (0,03)	99,95 (0,02)	99,98 (0,02)
15	<i>CR</i>	68,90 (1,92)	40,45 (1,52)	15,32 (0,67)	10,95 (0,77)	4,03 (0,66)
	<i>SM</i>	81,98 (1,76)	81,99 (2,52)	82,36 (2,91)	83,57 (3,55)	85,05 (7,39)
	<i>EM</i>	97,37 (0,14)	98,79 (0,10)	99,86 (0,03)	99,93 (0,02)	99,98 (0,01)
25	<i>CR</i>	82,37 (2,24)	54,20 (1,68)	13,86 (0,52)	10,05 (0,37)	4,46 (0,44)
	<i>SM</i>	81,66 (2,20)	81,55 (2,02)	81,32 (2,71)	82,02 (3,12)	82,76 (5,34)
	<i>EM</i>	96,37 (0,14)	97,78 (0,10)	99,73 (0,03)	99,89 (0,01)	99,98 (0,01)

Quadro 16: Eficiência e fiabilidade obtidas na simulação da metodologia baseada em *arrays* com *master pool* considerando erros de classificação de 5 por cento

No Quadro 16 incluímos os erros de classificação no modelo, considerando que a sensibilidade e a especificidade assumem o valor 0.95 (considerando assim erros de classificação de 5 por cento). Neste caso, além de ser considerada a eficiência do processo, serão igualmente apresentadas a sensibilidade e a especificidade da metodologia (valores obtidos via simulação recorrendo ao programa apresentado no Anexo 10). Deste modo, os resultados obtidos nestas simulações (Quadro 16) permitem tirar as seguintes conclusões em relação à metodologia baseada em *arrays* com *master pool M4*:

- O valor ótimo (em termos de eficiência) não parecer ter diferenças significativas entre os casos de presença e ausência de erros de classificação, apesar de existir algumas oscilações nos valores obtidos para o custo relativo;
- Os valores observados para a eficiência parecem ter diferenças significativas entre os casos de presença e ausência de erros de classificação, contudo não há uma tendência (não é sempre maior num caso em relação ao outro);
- Todavia, em termos de fiabilidade, esta metodologia tem um comportamento análogo às restantes metodologias (*M1*, *M2* e *M3*), tendo igualmente uma perda considerável em termos da sensibilidade e um ténue ganho em termos de especificidade.

Com recurso à programação apresentada no Anexo 10 e aos resultados apresentados do Quadro 16, concluímos a dimensão ótima para cada uma das taxas de prevalência na aplicação da metodologia baseada em *arrays* com utilização de *master pool (M4)*, quando os erros de classificação são iguais a 5 por cento (sensibilidade e especificidade individuais e compostas iguais a 95 por cento). Nesta metodologia, ao contrário das anteriores, após o primeiro valor ótimo (ótimo “local”) o custo relativo volta a diminuir para valores de n elevados, pelo que apresentamos ainda outra dimensão nos casos em que a taxa de prevalência assume os valores 1, 0.5 e 0.1 por cento. Para estas dimensões apresentamos no Quadro 17 os valores das medidas associadas à eficiência e à fiabilidade da metodologia *M4*.

p	0,1	0,05	0,01		0,005		0,001	
n	7	10	4	25	6	38	6	106
CR	56,16	36,99	17,29	13,95	10,67	9,21	6,05	3,67
SM	81,67	81,64	88,17	81,63	87,28	81,45	88,10	81,66
EM	98,97	99,32	99,95	99,73	99,96	99,81	99,98	99,91

Quadro 17: Valores ótimos, com erros de classificação de 5%, utilizando a metodologia baseada em *arrays* com *master pool*

Os valores patentes no Quadro 17, referentes à aplicação da metodologia baseada em *arrays* com a utilização de *master pool* (M4) podem ser interpretados do seguinte modo:

- Para uma taxa de prevalência igual a 10 por cento, a dimensão ótima da matriz é 7×7, realizando-se (em média) 56.16 testes para classificar 100 indivíduos, sendo 81.67 por cento dos infetados e 98.97 por cento dos saudáveis bem classificados;
- Para uma taxa de prevalência igual a 5 por cento, a dimensão ótima da matriz é 10×10, realizando-se (em média) 36.99 testes para classificar 100 indivíduos, sendo 81.64 por cento dos infetados e 99.32 por cento dos saudáveis bem classificados;
- Para uma taxa de prevalência igual a 1 por cento, a dimensão ótima da matriz é 4×4, realizando-se (em média) 17.29 testes para classificar 100 indivíduos, sendo 88.17 por cento dos infetados e 99.95 por cento dos saudáveis bem classificados;
- Para uma taxa de prevalência igual a 1 por cento, a dimensão ótima da matriz é 25×25, realizando-se (em média) 13.95 testes para classificar 100 indivíduos, sendo 81.63 por cento dos infetados e 99.73 por cento dos saudáveis bem classificados;
- Para uma taxa de prevalência igual a 0.5 por cento, a dimensão ótima da matriz é 6×6, realizando-se (em média) 10.67 testes para classificar 100 indivíduos, sendo 87.28 por cento dos infetados e 99.96 por cento dos saudáveis bem classificados;
- Para uma taxa de prevalência igual a 0.5 por cento, a dimensão ótima da matriz é 38×38, realizando-se (em média) 9.21 testes para classificar 100 indivíduos, sendo 81.45 por cento dos infetados e 99.81 por cento dos saudáveis bem classificados;

- Para uma taxa de prevalência igual a 0.1 por cento, a dimensão ótima da matriz é 6×6 , realizando-se (em média) 6.05 testes para classificar 100 indivíduos, sendo 88.10 por cento dos infetados e 99.98 por cento dos saudáveis bem classificados;
- Para uma taxa de prevalência igual a 0.1 por cento, a dimensão ótima da matriz é 106×106 , realizando-se (em média) 3.67 testes para classificar 100 indivíduos, sendo 81.66 por cento dos infetados e 99.91 por cento dos saudáveis bem classificados.

A grande diferença na análise desta metodologia é a possibilidade de conseguirmos, através de um aumento significativo da dimensão da amostra, um resultado mais eficiente (com um custo relativo ligeiramente inferior). Todavia, notemos que tal situação não é adequada, uma vez que para conseguirmos diminuir ligeiramente o custo relativo estamos a perder cerca de 6 por cento na sensibilidade, pelo que a perda em termos de percentagem de fiabilidade é demasiado grande em relação ao que se ganha em termos de eficiência. Por outro lado, a utilização de dimensões demasiado elevadas poderá ser impossível, pois ao considerarmos uma matriz de dimensão 106 estamos a supor que colocamos numa matriz 11236 indivíduos e que realizamos um teste composto com todos eles (*master pool*) o que não é realista. Deste modo, nos casos em que apresentamos dois valores, a opção pela dimensão correspondente ao primeiro ótimo “local” será mais adequada (notemos que o segundo ótimo corresponde ao ótimo da metodologia *M3* baseada em *arrays* mas sem a utilização do *master pool*).

4 – Comparação e discussão de Resultados

Para cada metodologia de classificação foram elaboradas as diferentes técnicas de simulação através do *software* R. Uma vez concluída a apresentação de cada metodologia, descritas no capítulo anterior, na discussão dos resultados será realizada uma análise conjunta, comparando os resultados obtidos em cada metodologia de classificação onde são utilizados os valores (para a dimensão dos grupos) ótimos de cada metodologia conforme foram deduzidos no capítulo anterior (comparação do melhor que se consegue obter em cada metodologia).

A discussão dos resultados terá como base a análise da eficiência (e, consequentemente, o custo) e da fiabilidade de utilização de cada metodologia (sensibilidade e a especificidade de cada metodologia) da execução das diferentes metodologias baseadas em testes compostos sob diferentes cenários relativos à dimensão dos grupos (n), à taxa de prevalência (p) bem como à ausência ou presença de erros de classificação (e, neste último caso, com diferentes sensibilidades e especificidades).

Tendo em consideração a existência de uma infinidade de possíveis cenários, restringimos a nossa análise a três casos específicos: ausência de erros de classificação no teste individual (sensibilidade e especificidade iguais a 100 por cento, isto é, $S = E = 100$ que será analisado no Quadro 18); sensibilidade e especificidade individuais iguais a 99 por cento ($S = E = 99$, conforme apresentado no Quadro 19); e sensibilidade e especificidade iguais a 95 por cento ($S = E = 95$, cujos resultados são apresentados no Quadro 20). Em cada situação foram utilizadas as quatro metodologias previamente descritas ($M1$, $M2$, $M3$ e $M4$) recorrendo a simulação (programas apresentados nos Anexos 7, 8, 9 e 10), tendo sido aplicadas diversas taxas de prevalência (10, 5, 1, 0.5 e 0.1 por cento) e para cada taxa de prevalência foi utilizada a dimensão ótima de cada metodologia (que foi determinada no subcapítulo anterior).

Desta forma, tal como nas simulações anteriores, estamos a considerar três quantidades fixas (a taxa de prevalência, a sensibilidade e a especificidade) para as quais comparamos

os resultados obtidos pela aplicação de cinco metodologias (as quatro baseadas em testes conjuntos, descritas previamente, e a utilização de testes individuais). Deste modo, numa infecção concreta, conhecendo as três quantidades que foram fixadas, podemos através da consulta das tabelas concluir qual a melhor metodologia a aplicar. Sublinhemos, desde já, que as duas medidas em avaliação (eficiência *versus* fiabilidade) terão por vezes sentidos opostos, isto é, para diminuir os custos (aumentar a eficiência) teremos de ceder na fiabilidade (diminuindo a sensibilidade e/ou a especificidade). Esta decisão, que é estratégica (qualidade do serviço *versus* custo associado ao serviço), não será aqui discutida uma vez que o único objetivo do trabalho é fornecer ferramentas que permitam aos decisores obter a informação necessária para tomarem a decisão. Por outro lado, salientemos igualmente que os valores aqui apresentados são unicamente um exemplo (a maioria dos valores não estará contemplada nas tabelas que apresentamos). Todavia, os programas necessários para as simulações serão apresentados em anexo e, como tal, as performances destas metodologias para quaisquer valores distintos dos analisados podem ser facilmente obtidas através das correspondentes simulações realizadas recorrendo aos referidos programas e ao *software* gratuito R.

4.1 – Simulação sem erros de classificação

No Quadro 18 apresentamos os resultados das diferentes metodologias de classificação na ausência de erros de classificação.

	<i>p</i>	0,1	0,05	0,01		0,005		0,001	
<i>M1</i>	<i>n</i>	4	5	11		15		32	
	<i>CR</i>	59,28	42,67	19,20		13,71		6,22	
<i>M2</i>	<i>n</i>	6	8	16		22		45	
	<i>CR</i>	59,31	39,31	15,73		10,77		4,60	
<i>M3</i>	<i>n</i>	7	10	25		38		106	
	<i>CR</i>	58,38	37,94	13,57		8,62		2,97	
<i>M4</i>	<i>n</i>	7	10	4	25	6	38	11	106
	<i>CR</i>	60,20	38,81	14,01	13,69	8,85	8,67	3,00	2,99

Quadro 18: Simulação sem erros de classificação: Especificidade e Sensibilidade a 100%

Numa visão geral, da análise aos valores obtidos através das simulações realizadas para as quatro metodologias, sem erros de classificação, constatamos o seguinte:

- Para os mesmos cenários (mesmos valores de taxa de prevalência) todas as quatro metodologias apresentam valores de eficiência muito semelhantes. A exceção verifica-se no caso em que a taxa de prevalência é bastante baixa, igual a 0.1 por cento, na qual há algumas diferenças significativas. Contudo, estas diferenças devem-se sobretudo à utilização de dimensões muito elevadas para os grupos o que, muito provavelmente, tornarão o teste composto de impossível realização, nomeadamente a realização do *master pool* na metodologia *M4* (mesmo recorrendo à ajuda de máquinas);
- Há um ganho mais evidente quando estamos perante taxas de prevalência mais baixas, onde a utilização de testes compostos permite atingir uma maior eficiência, sendo que nestes casos a poupança de recursos pode atingir valores bastante significativos.

4.2 – Simulação com erros de classificação a 99 por cento

Apresentamos, no Quadro 19, os resultados obtidos (via simulação) relativos à aplicação das diferentes metodologias quando a sensibilidade e a especificidade assumem o valor 0.99 (99 por cento). Da análise a estes resultados podemos tirar as seguintes conclusões:

- Para a mesma taxa de prevalência todas as quatro metodologias apresentam valores de eficiência, sensibilidade e especificidade muito semelhantes;
- Em relação à aplicação de testes individuais, efetivamente há um ganho significativo em termos de eficiência, mas há também uma ligeira perda em termos de fiabilidade devido à diminuição da sensibilidade;
- Há igualmente um ganho mais significativo quando estamos perante taxas de prevalência mais baixas, sendo que nestes casos existe uma considerável poupança de recursos.

	<i>p</i>	0,1	0,05	0,01		0,005		0,001	
<i>M1</i>	<i>n</i>	4	5	11		15		32	
	<i>CR</i>	59,70	43,58	20,34		14,77		7,22	
	<i>SM</i>	98	98,02	98,03		98,01		97,08	
	<i>EM</i>	99,72	99,77	99,90		99,92		99,96	
<i>M2</i>	<i>N</i>	7	8	16		22		45	
	<i>CR</i>	59,82	30,40	15,84		10,85		4,68	
	<i>SM</i>	97,08	97,15	97,09		96,82		97,15	
	<i>EM</i>	99,77	99,86	99,30		99,95		99,98	
<i>M3</i>	<i>n</i>	7	10	25		38		106	
	<i>CR</i>	58,15	37,99	13,73		8,80		3,13	
	<i>SM</i>	97,02	97,06	97,05		97,03		97,04	
	<i>EM</i>	99,78	99,86	99,95		99,97		99,99	
<i>M4</i>	<i>n</i>	7	10	4	25	6	38	11	106
	<i>CR</i>	59,27	38,5	15,22	13,74	9,14	8,78	3,2	3,11
	<i>SM</i>	96,02	96,03	97,70	96,03	97,64	96,13	97,67	96,01
	<i>EM</i>	99,78	99,86	100	99,95	100	99,97	100	99,99

Quadro 19: Simulação sem erros de classificação: Especificidade e Sensibilidade a 99%

4.3 – Simulação com erros de classificação a 95 por cento

Por fim, no Quadro 20 estão expostos os valores obtidos nas simulações realizadas com a aplicação das diferentes metodologias quando a sensibilidade e a especificidade são iguais a 95 por cento. Desta análise salientamos as seguintes conclusões:

- Para a mesma taxa de prevalência todas as quatro metodologias apresentam valores de eficiência (custos relativos) e fiabilidade (sensibilidade e especificidade da metodologia) muito parecidos;
- Há igualmente um ganho mais significativo quando estamos perante taxas de prevalência mais baixas, sendo que nestes casos a poupança de recursos pode ser bastante elevada;

	<i>p</i>	0,1	0,05	0,01		0,005		0,001	
<i>M1</i>	<i>n</i>	4	5	11		15		32	
	<i>CR</i>	60,93	45,18	23,46		18,23		10,95	
	<i>SM</i>	90,31	90,32	90,15		90,40		90,51	
	<i>EM</i>	98,54	98,74	99,33		99,44		99,61	
<i>M2</i>	<i>N</i>	7	8	16		22		45	
	<i>CR</i>	58,21	38,85	16,16		11,18		4,95	
	<i>SM</i>	85,64	85,54	85,88		85,99		86,11	
	<i>EM</i>	98,88	99,30	99,65		99,75		99,89	
<i>M3</i>	<i>n</i>	7	10	25		38		106	
	<i>CR</i>	57,11	38,13	14,55		9,64		3,87	
	<i>SM</i>	85,95	85,97	85,95		86,03		85,84	
	<i>EM</i>	98,90	99,28	99,71		99,80		99,91	
<i>M4</i>	<i>n</i>	7	10	4	25	6	38	6	106
	<i>CR</i>	56,16	36,99	17,29	13,95	10,67	9,21	6,05	3,67
	<i>SM</i>	81,67	81,64	88,17	81,63	87,28	81,45	88,10	81,66
	<i>EM</i>	98,97	99,32	99,95	99,73	99,96	99,81	99,98	99,91

Quadro 20: Simulação sem erros de classificação: Especificidade e Sensibilidade a 95%

- Em relação à aplicação de testes individuais, efetivamente há um ganho significativo em termos de eficiência, mas há também uma perda igualmente significativa em termos de fiabilidade devido a uma diminuição considerável da sensibilidade (apesar de existir um aumento da especificidade, mas que é sempre menor que a diminuição da sensibilidade).

Desta forma, há um conflito entre a eficiência e a fiabilidade dos resultados. Por conseguinte, cabe a quem toma a decisão sobre a metodologia a aplicar o papel de procurar o equilíbrio entre o custo que está disposto a investir e a fiabilidade que está disposto a abdicar, pois não há metodologias que consigam melhorar estas duas características em simultâneo.

5 – Conclusão

Este trabalho teve como principal objetivo comparar a performance de cinco metodologias de classificação (através de testes individuais ou da aplicação de quatro algoritmos distintos baseados em testes compostos) sob diferentes cenários, nomeadamente no que se refere ao custo relativo, à sensibilidade e à especificidade de cada metodologia, de forma a tornar fácil a identificação da metodologia mais adequada na classificação de indivíduos relativamente à infeção em análise. Esta análise foi efetuada através de simulações realizadas no *software* gratuito R. Este processo permitiu analisar, no capítulo 3, a performance de cada metodologia, de forma a permitir a escolha da dimensão do grupo n mais adequada, isto é, que permite minimizar o custo de classificação da população. Todavia, além de procurarmos diminuir o custo, é igualmente objetivo manter a fiabilidade das classificações em níveis aceitáveis, nomeadamente no que se refere à sensibilidade e à especificidade da metodologia aplicada. Nesta secção foi permitido concluir que efetivamente, sob determinadas condições, todas as 4 metodologias descritas conseguem ter um elevada eficiência (custo relativo baixo) mantendo (em alguns casos) a fiabilidade em níveis razoáveis. Por outro lado, no capítulo 4 comparamos as metodologias, sendo que nenhuma é claramente preferível às restantes, pois dependerá dos valores assumidos pela taxa de prevalência, sensibilidade e especificidade. Foi possível identificar que, havendo ou não interferência ao nível de erros de classificação, existem sempre ganhos de eficiência na execução dos testes compostos (se a taxa de prevalência for relativamente baixa, única situação aqui analisada pois caso contrário os testes individuais são preferíveis) sendo esses ganhos mais significativos nas situações em que a taxa de prevalência é bastante baixa, podendo-se recorrer a grupos de maiores dimensões. Todavia, há que ter atenção na escolha da dimensão do grupo, pois em alguns casos a perda na sensibilidade é relevante, podendo ser aconselhável a não utilização de testes compostos apesar da obtenção de um custo relativo bastante baixo.

As simulações realizadas mostram claramente que os testes compostos só podem ser recomendados em casos com baixas taxas de prevalência (situações que permitem um ganho em termos de eficiência) e baixas probabilidades de erros de classificação (pois,

caso contrário, a diminuição da sensibilidade da metodologia é demasiado significativa colocando em causa a fiabilidade dos resultados) sendo possível identificar a metodologia mais adequada para cada caso em função da sua taxa de prevalência, sensibilidade e especificidade (apesar de, em muitos casos, termos observado performances semelhantes nas quatro metodologias baseadas em testes compostos que neste trabalho foram investigadas). Além disso, sempre que a taxa de prevalência, a sensibilidade e a especificidade são conhecidos (ou, pelo menos, estimativas razoáveis estão disponíveis uma vez que pequenas diferenças nestes parâmetros não alteram muito os resultados), simulações podem ser realizadas para identificar a metodologia mais adequada, e, deste modo, encontrar um ponto de equilíbrio entre o custo e a fiabilidade da classificação.

As medidas de performance de cada uma das metodologias sob cenários não contemplados nas tabelas apresentadas ao longo do presente trabalho podem facilmente ser obtidas recorrendo aos programas de *software* R apresentados nos Anexos. Deste modo, o trabalho desenvolvido permite a obtenção da performance das metodologias apresentadas, para quaisquer valores da taxa de prevalência, da sensibilidade e da especificidade, informação necessária para a adequada tomada de decisão.

Os resultados obtidos através das simulações, cada uma delas considerando 1000 populações, apresentam na sua maioria desvios padrões relativamente baixos, o que significa a obtenção de performances semelhantes quando a mesma metodologia é aplicada em diferentes populações sob as mesmas características. Deste modo, os valores aproximados fornecidos pelas simulações são fiáveis e podem ser tidos em consideração como informação para a adequada tomada de decisão. Todavia, e tendo em consideração que não era objetivo do presente trabalho apresentar aproximações inequivocamente próximas do valor teórico, podem ser obtidas aproximações ainda com uma maior precisão do que as apresentadas, bastando para tal aumentar a dimensão da população utilizada nas simulações (e ter um computador com um excelente processador ou, em alternativa, ter a disponibilidade temporal necessária para esperar pelos resultados uma vez que, mesmo com os valores utilizados no presente trabalho, houve diversas simulações que demoraram diversas horas até o computador disponibilizar o resultado obtido).

Esperamos que com este trabalho tenhamos conseguido dar um pequeno contributo para a divulgação destas metodologias, bem como para a análise das suas performances, nomeadamente nos níveis de eficiências que com elas podemos atingir. Por outro lado, pretendemos igualmente alertar para a importância do conhecimento da sensibilidade e da especificidade, pois em casos em que os testes individuais já evidenciem problemas com o valor da sensibilidade, a utilização destas metodologias compostas é totalmente desaconselhada, principalmente em situações de possível contágio, pois a sensibilidade destas metodologias será ainda mais diminuta (o que pode levar à existência de um grande número de falsos negativos – indivíduos que foram classificados como saudáveis mas que efetivamente estão infetados).

É certo que, além das metodologias que neste trabalho foram aplicadas, outras (mais complexas) existem que poderão ter resultados ainda mais eficientes e/ou mais fiáveis. No entanto, e a concluir, valerá a pena reiterar a performance dos que foram focados no presente estudo, dado que, por apresentarem resultados no campo da aceitabilidade em termos de fiabilidade, e bastante auspiciosos em termos de eficiência, indiciam a sua inequívoca utilidade em abordagens futuras, se estes forem adequadamente divulgados.

Bibliografia

Berger, T., Mandell, J.W., Subrahmanya, P. (2000). Maximally efficient two-stage screening. *Biometrics* **56**, 833-840.

Bilder, C.R., Zhang, B., Schaarschmidt, F., Tebbs, J. (2010). binGroup: a package for group testing, *The R Journal* **2**, 56-60.

Boswell, M.T., Gore, S.D., Lovison, G., Patil, G.P. (1996). Annotated bibliography of composite sampling, Part A: 1936-92, *Environmental and Ecological Statistics* **3**, 1-50.

Chen, C.L., Swallow, W.H. (1990). Using group testing to estimate a proportion, and to test the binomial model, *Biometrics* **46**, 1035-1046.

Dorfman, R (1943). The detection of defective members in large populations, *Ann. Math. Statistics* **14**, 436-440.

Finucan, H.M. (1964). The blood testing problem, *Applied Statistics* **13**, 43-50.

Garner, F.C., Stapanian, M.A., Yfantis, E.A., Williams, L.R. (1989). Probability estimation with sample compositing techniques, *J. Off. Stat.* **5**, 365-374.

Granado A.C. e Santos, R. (2014). Compound clinical trials: a comparative simulation study of its accuracy and cost, Abstracts of the 2nd IPLeiria Internacional Health Congress, Challenges & Innovation in Health, *Rev. Saúde Pública* **48**, 138.

Hughes-Oliver, J.M. (2006). *Pooling experiments for blood screening and drug discovery, Screening: Methods for Experimentation in Industry, Drug Discovery, and Genetics*, Springer, 48-68.

Hung, M., Swallow, W.H. (1999). Robustness of group testing in the estimation of proportions, *Biometrics* **55**, 231-237.

Hwang, F.K. (1976). Group testing with a dilution effect, *Biometrika* **63**, 671-673.

Johnson, N.L., Kotz, S., and Wu, X. (1991). *Inspection errors for attributes in quality control*, Chapman and Hall.

Kim, H., Hudgens, M., Dreyfuss, J., Westreich, D., Pilcher, C. (2007). Comparison of group testing algorithms for case identification in the presence of testing errors, *Biometrics* **63**, 1152-1163.

Lancaster, V.A., Keller-McNulty, S. (1998). A review of composite sampling methods, *Journal of the American Statistical Association* **93**, 1216-1230.

Liu, S.C., Chiang, K.S., Lin, C.H., Chung, W.C., Lin, S.H., Yang, T.C. (2011). Cost analysis in choosing group size when group testing for potato virus Y in the presence of classification errors, *Ann. Appl. Biology* **159**, 491-502

Loyer, M.W. (1983). Bad probability, good statistics, and group testing for binomial estimation, *The American Statistician* **37**, 57-59.

Phatarfod, R.M., Sudbury, A. (1994). The use of a square array scheme in blood testing. *Statistics in Medicine* **13**, 2337-2343.

Roederer, M. and Koup, R A. (2003). Optimized determination of T cell epitope responses. *Journal of Immunological Methods* **274**, 221-228.

Ross, S. M. (2006). *Simulation*, 4th edition, Elsevier Academic Press.

Santos, R., Pestana, D., Martins, J.P. (2013). Extensions of Dorfman's Theory. In Oliveira, P.E. et al. (eds.): *Studies in Theoretical and Applied Statistics, Recent Developments in Modeling and Applications in Statistics*, 179-189, Springer.

Santos, R., Felgueiras, M., and Martins, J.P. (2014). Known mean, unknown maxima? Testing the maximum knowing only the mean, *Communications in Statistics - Simulation and Computation* (publicado online: 23 Jan 2014).

Sobel, K.M., Elashoff, R.M. (1975). Group testing with a new goal, estimation, *Biometrika* **62**, 181-193.

Sobel, M., Groll, P.A. (1959) Group testing to eliminate efficiently all defectives in a binomial sample, *Bell System Technical Journal* **38**, 1179-1252

Sousa, R. (2012). *Testes Conjuntos: Extensões da Teoria de Dorfman*, Tese de Doutoramento, Universidade de Lisboa.

Sterret, A. (1957). On the detection of defective members of large populations, *Ann. Math. Statistics* **28**, 1033-1036.

Tu, X.M., Litvak, E., Pagano, M. (1994). Studies of AIDS and HIV surveillance, screening tests: can we get more by doing less?, *Statistics in Medicine* **13**, 1905-1919.

Tu, X.M., Litvak, E., Pagano, M. (1995). On the informativeness and accuracy of pooled testing in estimating prevalence of a rare disease: application to HIV screening, *Biometrika* **82**, 287-297.

Wein, L.M., Zenios, S.A. (1996). Pooled testing for HIV screening: capturing the dilution effect, *Operations Research* **44**, 543-569.

Woodbury, C.P., Fitzloff, J.F., Vincent, S.S. (1995). Sample multiplexing for greater throughput in HPLC and related methods, *Analytical Chemistry* **67**, 885-890.

Zenios, S., Wein, L. (1998). Pooled testing for HIV prevalence estimation exploiting the dilution effect, *Stat. Med.* **17**, 1447-1467.

Anexos

Anexo 1: Programação elaborada através do *software* R para simular se um indivíduo está infectado ou saudável

```
p=0.05 # p – taxa de prevalência  
  
if (runif(1,0,1)<p) {indivíduo = 1}  
else {indivíduo=0}  
  
indivíduo
```

Anexo 2: Programação elaborada através do *software* R para a realização de um teste individual (com inclusão de erros através da sensibilidade e especificidade)

```
p=0.05
s=0.95      # s – sensibilidade
e=0.95      # e – especificidade

if (runif(1,0,1)<p) {individuo = 1}
else {individuo=0}

if (individuo == 1 & runif(1,0,1)<s)
  {resultado="verdadeiro positivo"}

else if (individuo == 0 & runif(1,0,1)>e)
  {resultado="falso positivo"}

else if (individuo == 1 & runif(1,0,1)>=s)
  {resultado="falso negativo"}

else {resultado="verdadeiro negativo"}

resultado
```

Anexo 3: Programação elaborada através do *software* R para determinar o número de indivíduos infectados numa população de dimensão npop.

```
p=0.05
npop = 10^5    # npop – dimensão da população
NII=0          # NII – Número de Indivíduos Infectados

for (n in 1:npop)
  {if (runif(1,0,1)<p)
    {NII=NII+1}
   else {NII=NII}}
```

Anexo 4: Programação elaborada através do *software* R para a contabilização de verdadeiros positivos, verdadeiros negativos, falsos positivos, falsos negativos e número de testes na realização de testes individuais numa população de dimensão npop (com inclusão de erros através da sensibilidade e especificidade)

```
p=0.05
s=0.95
e=0.95
npop = 10^5
NII=0      # NII – Número de Indivíduos Infetados
NIIP=0     # NIIP – Número de Indivíduos Infetados com resultado Positivo
NINIP=0    # NINIP – Número de Indivíduos Não Infetados com resultado Positivo
NT=0       # NT – Número de testes realizados

for (n in 1:npop)
  {if (runif(1,0,1)<p)
    {individuo = 1; NII=NII+1}
  else {individuo=0};

  NT=NT+1;
  if (individuo == 1 & runif(1,0,1)<s)
    {NIIP=NIIP+1}
  else if (individuo == 0 & runif(1,0,1)>e)
    {NINIP=NINIP+1}
  }

paste ("Percentagem de doentes = ",NII/npop*100, ";")
paste ("Sensibilidade = ", NIIP/NII*100, ";")
paste ("Especificidade = ", (npop-NII-NINIP)/(npop-NII)*100, ";")
paste ("Custo relativo = ", NT/npop*100)
```

Anexo 5: Programação elaborada através do *software* R para a contabilização de verdadeiros positivos, verdadeiros negativos, falsos positivos, falsos negativos e número de testes na realização de testes individuais em nsim populações de dimensão npop (com inclusão de erros através da sensibilidade e especificidade)

```
p=0.05
s=0.95
e=0.95
npop = 10^3
nsim = 10^3      # nsim – número de simulações (réplicas)
NII<-rep(0, nsim) # NII – vetor com o n.º de indivíduos infetados em cada réplica
NIIP<-rep(0, nsim) # NIIP – vetor com o n.º de verdadeiros positivos
NINIP<-rep(0, nsim) # NINIP – vetor com o n.º de falsos positivos
NT<-rep(0, nsim)  # NT – vetor com o n.º de número de testes realizados

for(sim in 1:nsim)
  {for (n in 1:npop)
    {if (runif(1,0,1)<p)
      {indivíduo = 1; NII[sim]=NII[sim]+1}
    else {indivíduo=0};
    NT[sim]=NT[sim]+1;
    if (indivíduo == 1 & runif(1,0,1)<s)
      {NIIP[sim]=NIIP[sim]+1}
    else if (indivíduo == 0 & runif(1,0,1)>e)
      {NINIP[sim]=NINIP[sim]+1}
    }
  }

paste("Percentagem de doentes (média) = ", mean(NII/npop)*100, " com desvio padrão
=", sd(NII/npop*100) )
paste("Sensibilidade (média) = ", mean(NIIP/NII)*100, " com desvio padrão =",
sd(NIIP/NII*100) )
paste("Especificidade (média) = ", mean((npop-NII-NINIP)/(npop-NII))*100, " com desvio
padrão =", sd((npop-NII-NINIP)/(npop-NII)*100) )
paste("Custo relativo (média) = ", mean(NT/npop)*100, " com desvio padrão =",
sd(NT/npop*100) )
```

Anexo 6: Programação elaborada através do *software* R para a contabilização de verdadeiros positivos, verdadeiros negativos, falsos positivos, falsos negativos e número de testes na aplicação da metodologia de Dorfman numa população de dimensão npop (com inclusão de erros através da sensibilidade e especificidade)

```
npop = 10^4
n=5          # n – dimensão do grupo
p=0.05
s=0.95       # s – sensibilidade (individual e composta)
e=0.95       # e – especificidade (individual e composta)
NII=0
NIIP=0
NINIP=0
NT=0

ngrupos=npop/n
for(ng in 1:ngrupos)
  {grupo=rbinom(n,1,p);
  nd=sum(grupo);
  NII=NII+nd;
  NT=NT+1;
  if (nd>0 & runif(1,0,1)<s)
    {NIIP=NIIP+sum(rbinom(nd,1,s));
    NINIP=NINIP+sum(rbinom(n-nd,1,1-e));
    NT=NT+n}
  else if (nd==0 & runif(1,0,1)>e)
    {NINIP=NINIP+sum(rbinom(n,1,1-e));
    NT=NT+n}
  }

paste("Percentagem de doentes = ",NII/npop*100, ";")
paste("Sensibilidade Composta = ", NIIP/NII*100, ";")
paste("Especificidade composta = ", (npop-NII-NINIP)/(npop-NII)*100, ";")
paste("Custo relativo = ", NT/npop*100)
```

Anexo 7: Programação elaborada através do *software* R para a contabilização de verdadeiros positivos, verdadeiros negativos, falsos positivos, falsos negativos e número de testes na aplicação da metodologia de Dorfman (*MI*) em *nsim* populações de dimensão *npop* (com inclusão de erros através da sensibilidade e especificidade)

```

nsim = 10^3
npop = 10^4
n=5
p=0.05
s=0.95
e=0.95
NII<-rep(0, nsim)
NIIP<-rep(0, nsim)
NINIP<-rep(0, nsim)
NT<-rep(0, nsim)

ngrupos=npop/n
for(sim in 1:nsim)
  {for(ng in 1:ngrupos)
    {grupo=rbinom(n,1,p);
      nd=sum(grupo); # nd é o número de doentes do grupo
      NII[sim]=NII[sim]+nd;
      NT[sim]=NT[sim]+1;
      if (nd>0 & runif(1,0,1)<s)
        {NIIP[sim]=NIIP[sim]+sum(rbinom(nd,1,s));
          NINIP[sim]=NINIP[sim]+sum(rbinom(n-nd,1,1-e));
          NT[sim]=NT[sim]+n}
      else if (nd==0 & runif(1,0,1)>e)
        {NINIP[sim]=NINIP[sim]+sum(rbinom(n,1,1-e));
          NT[sim]=NT[sim]+n}
    }
  }

paste("Percentagem de doentes (média) = ", mean(NII/npop)*100, " com desvio padrão
=", sd(NII/npop*100) )
paste("Sensibilidade Composta (média) = ", mean(NIIP/NII)*100, " com desvio padrão =",
sd(NIIP/NII*100) )
paste("Especificidade composta (média) = ", mean((npop-NII-NINIP)/(npop-NII))*100, "
com desvio padrão =", sd((npop-NII-NINIP)/(npop-NII)*100) )
paste("Custo relativo (média) = ", mean(NT/npop)*100, " com desvio padrão =",
sd(NT/npop*100) )

```

Anexo 8: Programação elaborada através do *software* R para a contabilização de verdadeiros positivos, verdadeiros negativos, falsos positivos, falsos negativos e número de testes na aplicação da metodologia hierárquica (*M2*) em *nsim* populações de dimensão *npop* (com inclusão de erros através da sensibilidade e especificidade)

```

nsim = 10^3;
n=5
ngrupos=10^4
npop = n*ngupos;
p=0.05
s=.95
e=.95
NII<-rep(0, nsim)
NIIP<-rep(0, nsim)
NINIP<-rep(0, nsim)
NT<-rep(0, nsim)

for(sim in 1:nsim) {
  for(ng in 1:ngrupos)
    {grupo=rbinom(n,1,p);
     nd=sum(grupo);
     NII[sim]=NII[sim]+nd;
     if ((nd>0 & runif(1,0,1)<s) | (nd==0 & runif(1,0,1)>e))
       {grupo1=grupo[1:(floor(n/2))];
        nd1=sum(grupo1);
        if ((nd1>0 & runif(1,0,1)<s) | (nd1==0 & runif(1,0,1)>e))
          {if (nd1>0) {NIIP[sim]=NIIP[sim]+sum(rbinom(nd1,1,s))};
           NINIP[sim]=NINIP[sim]+sum(rbinom(floor(n/2)-nd1,1,1-e));
           NT[sim]=NT[sim]+floor(n/2)}
         grupo2=grupo[(floor(n/2)+1):n];
         nd2=sum(grupo2);
         if ((nd2>0 & runif(1,0,1)<s) | (nd2==0 & runif(1,0,1)>e))
           {if (nd2>0) {NIIP[sim]=NIIP[sim]+sum(rbinom(nd2,1,s))};
            NINIP[sim]=NINIP[sim]+sum(rbinom(ceiling(n/2)-nd2,1,1-e));
            NT[sim]=NT[sim]+ceiling(n/2)}
          NT[sim]=NT[sim]+3}
     else {NT[sim]=NT[sim]+1}}}

paste("Porcentagem de doentes (média) = ", mean(NII/npop)*100, " com desvio padrão =",
sd(NII/npop*100) )
paste("Sensibilidade Composta (média) = ", mean(NIIP/NII)*100, " com desvio padrão =",
sd(NIIP/NII*100) )
paste("Especificidade composta (média) = ", mean((npop-NII-NINIP)/(npop-NII))*100, " com
desvio padrão =", sd((npop-NII-NINIP)/(npop-NII)*100) )
paste("Custo relativo (média) = ", mean(NT/npop)*100, " com desvio padrão =",
sd(NT/npop*100) )

```


Anexo 9: Programação elaborada através do *software* R para a contabilização de verdadeiros positivos, verdadeiros negativos, falsos positivos, falsos negativos e número de testes na aplicação da metodologia baseada em *arrays* sem aplicação do *master pool* (M3) em *nsim* populações de dimensão *npop* (com inclusão de erros através da sensibilidade e especificidade)

```

nsim = 10^3;
n=5                                # Dimensão da matriz utilizada é n×n
ngrupos=10^3
npop = n^2*ngrupos;
p=0.05
s=.95
e=.95
NII<-rep(0, nsim)
NIIP<-rep(0, nsim)
NINIP<-rep(0, nsim)
NT<-rep(0, nsim)

for(sim in 1:nsim) {
  for(ng in 1:ngrupos)
    {grupo=rbinom(n^2,1,p);
     nd=sum(grupo);
     NII[sim]=NII[sim]+nd;
     {MT<-array(0, c(n,n));
      for(l in 1:n)
        {for(c in 1:n)
          {MT[l,c]=grupo[(l-1)*n+c]}
        };
      linhas<-rep(0,n);
      colunas<-rep(0,n);
      linhas_positivas<-rep(0,0);
      colunas_positivas<-rep(0,0);
      for(l in 1:n)
        {if ((sum(MT[l,])>0 & runif(1,0,1)<s)
          | (sum(MT[l,])==0 & runif(1,0,1)>e))
          {linhas[l]=1;linhas_positivas=c(linhas_positivas,l)}
        if ((sum(MT[l,])>0 & runif(1,0,1)<s)
          | (sum(MT[l,])==0 & runif(1,0,1)>e))
          {colunas[l]=1;colunas_positivas=c(colunas_positivas,l)}
        }
      if (max(sum(linhas),sum(colunas))==0)
        {TI=0}                                # TI – Número de testes individuais
      else if (sum(linhas)==0) {TI=n*sum(colunas);
        NIIP[sim]=NIIP[sim]+
          sum(rbinom(sum(MT[,colunas_positivas]),1,s));
        NINIP[sim]=NINIP[sim]+sum(rbinom(n*sum(colunas)-

```

```

sum(MT[,colunas_positivas]),1,1-e))}
else if (sum(colunas)==0){TI=n*sum(linhas);
  NIIP[sim]=NIIP[sim]+sum(rbinom(sum(MT[linhas_positivas,]),1,s));
  NINIP[sim]=NINIP[sim]+sum(rbinom(n*sum(linhas)-
    sum(MT[linhas_positivas,]),1,1-e))}
else {TI=sum(linhas)*sum(colunas);
  for(ti in linhas_positivas) {for(tic in colunas_positivas)
    if (MT[ti,tic]==1 & runif(1,0,1)<s) {NIIP[sim]=NIIP[sim]+1}
    else if (MT[ti,tic]==0 & runif(1,0,1)>e) {NINIP[sim]=NINIP[sim]+1}
  }}
NT[sim]=NT[sim]+2*n+TI}}}

```

```

paste("Percentagem de doentes (média) = ", mean(NII/npop)*100, " com desvio padrão
=", sd(NII/npop*100) )
paste("Sensibilidade Composta (média) = ", mean(NIIP/NII)*100, " com desvio padrão =",
sd(NIIP/NII*100) )
paste("Especificidade composta (média) = ", mean((npop-NII-NINIP)/(npop-NII))*100, "
com desvio padrão =", sd((npop-NII-NINIP)/(npop-NII)*100) )
paste("Custo relativo (média) = ", mean(NT/npop)*100, " com desvio padrão =",
sd(NT/npop*100) )

```

Anexo 10: Programação elaborada através do *software* R para a contabilização de verdadeiros positivos, verdadeiros negativos, falsos positivos, falsos negativos e número de testes na aplicação da metodologia baseada em *arrays* com aplicação do *master pool* (M4) em *nsim* populações de dimensão *npop* (com inclusão de erros através da sensibilidade e especificidade)

```

nsim = 10^3;
n=5                                # Dimensão da matriz utilizada é n×n
ngrupos=10^3
npop = n^2*ngrupos;
p=0.05
s=.95
e=.95
NII<-rep(0, nsim)
NIIP<-rep(0, nsim)
NINIP<-rep(0, nsim)
NT<-rep(0, nsim)

#teste previo a todos os indivíduos na matriz#
for(n in 1:npop) {
  if (runif(1,0,1)>p) {populacao='populacao com todos classificados como não
infectados';
  NT = NT+1}
  else {for(sim in 1:nsim) {
    for(ng in 1:ngrupos){grupo=rbinom(n^2,1,p);
      nd=sum(grupo);
      NII[sim]=NII[sim]+nd;
      { MT<-array(0, c(n,n));
        for(l in 1:n){for(c in 1:n){MT[l,c]=grupo[(l-1)*n+c]}};
        linhas<-rep(0,n);
        colunas<-rep(0,n);
        linhas_positivas<-rep(0,0);
        colunas_positivas<-rep(0,0);
        for(l in 1:n){
          if ((sum(MT[l,])>0 & runif(1,0,1)<s)
            | (sum(MT[l,])==0 & runif(1,0,1)>e))
            {linhas[l]=1;linhas_positivas=c(linhas_positivas,l)}
          if ((sum(MT[,l])>0 & runif(1,0,1)<s)
            | (sum(MT[,l])==0 & runif(1,0,1)>e))
            {colunas[l]=1;
              colunas_positivas=c(colunas_positivas,l)}}
          if (max(sum(linhas),sum(colunas))==0) {TI=0}
          else if (sum(linhas)==0) {TI=n*sum(colunas);
            NIIP[sim]=NIIP[sim]+
              sum(rbinom(sum(MT[,colunas_positivas]),1,s));
            NINIP[sim]=NINIP[sim]+

```

```

sum(rbinom(n*sum(colunas)-
sum(MT[,colunas_positivas]),1,1-e)))
else if (sum(colunas)==0){TI=n*sum(linhas);
NIIP[sim]=NIIP[sim]+
sum(rbinom(sum(MT[linhas_positivas,]),1,s));
NINIP[sim]=NINIP[sim]+sum(rbinom(n*sum(linhas)-
sum(MT[linhas_positivas,]),1,1-e))}
else {TI=sum(linhas)*sum(colunas);
for(ti in linhas_positivas) {for(tic in colunas_positivas)
if (MT[ti,tic]==1 & runif(1,0,1)<s)
{NIIP[sim]=NIIP[sim]+1}
else if (MT[ti,tic]==0 & runif(1,0,1)>e)
{NINIP[sim]=NINIP[sim]+1}}}
NT[sim]=NT[sim]+2*n+TI}}}}

```

```

paste("Percentagem de doentes (média) = ", mean(NII/npop)*100, " com desvio padrão
=", sd(NII/npop*100) )
paste("Sensibilidade Composta (média) = ", mean(NIIP/NII)*100, " com desvio padrão =",
sd(NIIP/NII*100) )
paste("Especificidade composta (média) = ", mean((npop-NII-NINIP)/(npop-NII))*100, "
com desvio padrão =", sd((npop-NII-NINIP)/(npop-NII)*100) )
paste("Custo relativo (média) = ", mean(NT/npop)*100, " com desvio padrão =",
sd(NT/npop*100) )

```