ISSN 2225-6016

ВЕСТНИК

Смоленской государственной медицинской академии

Том 19, №2

2020

УДК 519 253

DOI: 10.37903/vsgma.2020:2.7 14.03.06 Фармакология, клиническая фармакология

СПОСОБ ПРИНЯТИЯ СТАТИСТИЧЕСКИХ РЕШЕНИЙ О СОСТОЯНИЯХ, ВЫЗВАННЫХ ФАРМАКОЛОГИЧЕСКИМ ВОЗДЕЙСТВИЕМ, НА ОСНОВЕ КРИТЕРИЯ МИНИМАЛЬНОГО СРЕДНЕГО РИСКА

© Лямец Л.Л., Евсеев А.В., Данилов А.И.

Смоленский государственный медицинский университет, Россия, 214019, Смоленск, ул. Крупской, 28

Резюме

Цель. Разработка способа принятия статистических решений о состояниях организма и биологических систем, вызванных фармакологическим воздействием, на основе критерия минимального ожидаемого среднего риска. При проведении фармакологических исследований может возникать необходимость определения состояния исследуемого объекта в условиях неопределенности, когда прямые признаки, необходимые для идентификации состояния, неизвестны. Способ предназначается для применения в медико-биологических исследованиях, в практической медицинской деятельности, связанной с дифференциальной диагностикой заболеваний, а также для идентификации состояний, вызванных фармакологическим воздействием.

Методика. Разработанный способ состоит из последовательных этапов, выполнение которых направлено на формальное представление информации об исследуемом объекте и последующую обработку этой информации для принятия решения о его состоянии. Для формального описания и представления информации об исследуемом объекте разрабатываются следующие математические конструкции: множество возможных состояний исследуемого объекта; множество допустимых статистических решений; функция потерь; множество реализаций дополнительного эксперимента, необходимого для уточнения знаний об исследуемом объекте; матрица условных вероятностей для возможных реализаций эксперимента, связанных с состояниями исследуемого объекта. Статистическое решение о состоянии исследуемого объекта принимается на основании анализа указанных математических конструкций. Для обработки информации и принятия решения определяются решающие функции и соответствующие им функции риска. Критерием для принимаемого решения является минимальный ожидаемый средний риск. В результате вычислений выбирается такое решение о состоянии исследуемого объекта, которому соответствует решающая функция, обеспечивающая минимальный средний риск.

Результаты. В результате проведенного исследования определен алгоритм формального представления и преобразования априорной информации об объекте исследования в информацию, которая позволит принять обоснованное решение о его состоянии. Разработанный алгоритм составляет сущность способа нахождения статистического решения о состоянии исследуемого объекта на основе критерия минимального среднего риска. Идентификация состояния исследуемого объекта рассмотрена для такой ситуации, когда прямые признаки, точно определяющие состояние объекта исследования, по объективным причинам не могут быть измерены и определены, а косвенные признаки, полученные в результате дополнительного эксперимента и уточняющие знания, имеют вероятностный характер.

Заключение. Проведенное исследование позволило разработать способ обоснования и принятия статистического решения, направленного на идентификацию состояния объекта исследования в условиях неопределенности, когда априорная информация и информация о косвенных признаках, полученная в результате дополнительного эксперимента, имеет вероятностный характер. В данном случае под объектом понимается живой организм или биологическая система, а его возможные состояния, подлежащие идентификации, могут быть вызваны соответствующими заболеваниями или фармакологическими воздействиями. Разработанный способ, кроме основного предназначения, может также применяться в качестве критерия оценки информативности возможных уточняющих экспериментов. Способ идентификации состояния объектов на основе критерия минимального среднего риска может представлять практический интерес для научных работников, осуществляющих исследования в области фармакологии и доказательной медицины, использующих в своей работе статистические методы анализа экспериментальных данных и принятия статистических решений.

Ключевые слова: статистические решения, решающая функция, функция потерь, фармакологическое воздействие, функция риска, средний риск, случайная величина, вероятность события

METHOD FOR MAKING STATISTICAL DECISIONS ABOUT CONDITIONS CAUSED BY PHARMACOLOGICAL EXPOSURE BASED ON THE MINIMUM AVERAGE RISK CRITERION

Lyamets L.L., Evseev A.V., Danilov A.I.
Smolensk State Medical University, 28, Krupskoj St., 214019, Smolensk, Russia

Abstract

Objective. To develop a method for making statistical decisions about the state of the body and biological systems caused by pharmacological effects, based on the criterion of the minimum expected average risk. When conducting pharmacological studies, it may be necessary to determine the state of the studied object under conditions of uncertainty, when the direct signs necessary to identify the state are unknown. The method is intended for use in biomedical research, in practical medical activities related to the differential diagnosis of diseases, as well as for the identification of conditions caused by pharmacological effects.

Method. The developed method consists of successive steps, the implementation of which is aimed at the formal presentation of information about the investigated object and the subsequent processing of this information to make a decision about its condition. For the formal description and presentation of information about the studied object, the following mathematical constructions are developed: the set of possible states of the studied object; many valid statistical decisions; loss function; many implementations of an additional experiment necessary to clarify knowledge about the object under study; matrix of conditional probabilities for possible experimental implementations related to the states of the object under study. A statistical decision on the state of the investigated object is made on the basis of the analysis of these mathematical structures. For information processing and decision making, the decisive functions and the corresponding risk functions are determined. The criterion for the decision is the minimum expected average risk. As a result of the calculations, a decision is made about the state of the object under study, which corresponds to the decisive function that provides the minimum average risk.

Results. As a result of the study, an algorithm for the formal presentation and conversion of a priori information about the object of the study into information that will make it possible to make an informed decision about its condition is determined. The developed algorithm is the essence of the method of finding a statistical decision on the state of the investigated object based on the criterion of minimum average risk. The identification of the state of the object under study is considered for such a situation when direct signs that accurately determine the state of the object of the study cannot be measured and determined for objective reasons, and indirect signs obtained as a result of an additional experiment and clarifying knowledge are probabilistic.

Conclusion. The study made it possible to develop a method for substantiating and making a statistical decision aimed at identifying the state of the object of the study in conditions of uncertainty, when a priori information and information on indirect signs obtained as a result of an additional experiment is probabilistic. In this case, the object is understood as a living organism or biological system, and its possible states to be identified can be caused by the corresponding diseases or pharmacological effects. The developed method, in addition to its main purpose, can also be used as a criterion for evaluating the information content of possible refinement experiments. The method for identifying the state of objects on the basis of the criterion of minimum average risk may be of practical interest for scientists carrying out research in the field of pharmacology and evidence-based medicine, using statistical methods for analyzing experimental data and making statistical decisions in their work.

Keywords: statistical solutions, decision function, loss function, pharmacological impact, risk function, average risk, random variable, event probability

Введение

Определение состояния объекта исследования (организма или биологической системы) при фармакологическом воздействии осуществляется при помощи признаков, измеренных в соответствующих шкалах [1-3]. Аналогично состояние организма при заболевании также определяется посредством патогномоничных признаков (симптомов). Признаки, которые позволяют точно определить состояние организма, будем называть прямыми признаками. При наличии прямых (патогномоничных) признаков принятие решения о состоянии объекта исследования является формальной, нормативно закрепленной процедурой и не представляет каких-либо практических проблем для лица, принимающего решение [6]. Поскольку прямые признаки являются отражением протекающих в организме процессов, то их проявление всегда происходит по истечении некоторого времени, связанного с переходными процессами [10, 11]. По этой причине практический и научный интерес представляет формальное обоснование принятия решения о состоянии организма или биологической системы на момент времени, когда прямые признаки ещё не проявились или по различным объективным причинам не могут быть измерены и определены. При этом объективная реальность такова, что имеется возможность измерять или оценивать признаки, которые в отличие от прямых признаков являются косвенными (непрямыми) [7].

Следует отметить следующую особенность эксперимента с фармакологическим воздействием, которая состоит в том, что косвенные признаки проявляются у объекта исследования раньше прямых признаков, но они имеют вероятностный характер для определяемых состояний. Поэтому косвенные признаки можно рассматривать как случайные величины. Условия, при которых решение о состоянии объекта исследования принимается на основании косвенных признаков при отсутствии прямых признаков, будем называть условиями неопределенности. При принятии решения в условиях неопределенности возможно возникновение ошибок, которые имеют различные по тяжести последствия и соответствующие риски. Для принятия решения о состоянии объекта исследования в условиях неопределенности может быть разработана формальная процедура, которая на основе имеющейся априорной информации позволит определить решающую функцию для минимизации риска принимаемых решений [4, 8].

Формальной основой для построения процедуры принятия решения в условиях неопределенности может служить теория статистических решений, являющаяся одним из направлений теории игр. Предметом исследования теории игр являются методы принятия решений в конфликтных ситуациях, в которых сталкиваются интересы нескольких лиц, преследующих противоположные цели. Специфическим видом конфликтных ситуаций являются статистические игры. Такие игры связаны с практическими ситуациями, когда основным игроком является человек (лицо, принимающее решение), а другой игрок является нейтральным, поскольку не стремится извлечь для себя выгоды и обратить в свою пользу ошибки, совершаемые противником. Подобным игроком может выступать физическая природа, под которой понимается вся совокупность внешних обстоятельств, в условиях которых основному игроку приходится принимать решение. Физическая природа в теории статистических игр не рассматривается как разумный противник, который стремится воспользоваться для достижения своих целей ошибками, совершаемыми человеком, т.е. она не имеет по отношению к нему злого умысла. Она просто развивается и действует в соответствии со своими законами, которые человек может познать и обратить в свою пользу. Если бы лицо, принимающее решение, обладало полными знаниями, оно могло бы использовать их с максимальной для себя выгодой. Однако в большинстве практических случаев законы природы известны недостаточно полно, что и порождает условия неопределенности. В условиях неопределенности нельзя избежать ошибочных решений. Рациональным выходом из этой ситуации является выработка лицом, принимающим решение, такой стратегии, которая, хотя и не исключает возможности принятия ошибочных решений, но позволяет минимизировать связанные с ними нежелательные последствия и риски [5, 9].

Целью исследования являлась разработка способа принятия статистических решений о состояниях организма и биологических систем, вызванных фармакологическим воздействием, на основе критерия минимального ожидаемого среднего риска.

Методика

В результате фармакологического воздействия объект исследования (организм или биологическая система) может оказаться в одном из возможных состояний. Эти состояния точно определяются прямыми признаками. Особенностью данной методики является то, что принятие решения о состоянии исследуемого объекта производится в условиях, когда прямые признаки по объективным причинам определить не представляется возможным. На момент принятия решения можно лишь получить информацию только о косвенных признаках, которые проявляются раньше прямых признаков, но имеют вероятностный характер.

Возможность определения состояния объекта исследования по косвенным признакам соответствует принципу опережающего материального отображения, который лежит в основе прогнозирования и раннего распознавания. Эти признаки являются индикаторами, которые несут превентивную информацию о наиболее вероятном состоянии объекта исследования, в котором он может оказаться после окончания переходных процессов, вызванных фармакологическим воздействием.

Для измерения косвенных признаков проводится дополнительный эксперимент. Группировка и комбинация косвенных признаков позволяют определить их сочетанные проявления – исходы (реализации) дополнительного эксперимента. Другими словами, исходы дополнительного эксперимента – это события, для определения которых используются сочетания косвенных признаков. Поскольку косвенные признаки имеют вероятностную природу, то и исходы дополнительного эксперимента также являются случайными событиями и характеризуются условными вероятностями, зависящими от возможных состояний исследуемого объекта. Для принятия решения о состоянии объекта на основании данных, полученных в результате дополнительного эксперимента, требуется формальное представление и обработка априорной информации. Формализация априорной информации позволяет получить математические конструкции, необходимые для проведения анализа и принятия решения [4].

Основными математическими конструкциями, используемыми в данной методике, являются: множество возможных состояний исследуемого объекта; множество допустимых статистических решений и связанные с этими решениями функции потерь; множество реализаций дополнительного эксперимента, необходимого для уточнения знаний об исследуемом объекте; матрица условных вероятностей для возможных реализаций эксперимента, ассоциируемая с множеством возможных состояний исследуемого объекта. На основании анализа указанных математических конструкций определяются решающие функции и соответствующие им функции риска. Критерием для принимаемого решения является ожидаемый минимальный средний риск. В результате вычислений выбирается такое решение о состоянии исследуемого объекта, которому соответствует решающая функция, обеспечивающая минимальный средний риск. По своей сути методика, положенная в основу принятия статистического решения, является дедуктивным выводом в отношении конкретного объекта исследования, основанным на анализе формализованной априорной информации, имеющей индуктивный характер.

Результаты исследования и их обсуждение

В результате проведенного исследования разработан порядок действий, выражающий сущность способа принятия статистического решения. Действия, необходимые для идентификации состояния исследуемого объекта, вызванного фармакологическим воздействием, удобно представить последовательными этапами. Проведем подробное обсуждение действий, выполняемых на каждом этапе.

На первом этапе производится формальное описание множества всех возможных состояний объекта исследования, в которых он может оказаться в результате фармакологического воздействия. Это множество можно записать в виде $A = \{A_1, A_2, ..., A_m\}$, где A_i, $1 \le i \le m$ – детерминированное состояние объекта исследования. Практический интерес представляют ситуации, в которых число состояний m является конечным и не менее двух. Обычно в большинстве проводимых исследований $m \le 5$. Каждое состояние A_i однозначно определяется множеством прямых признаков V_i, $1 \le i \le m$, т.е. можно сказать, что состояние A_i зависит от множества V_i. Для описания этого факта можно использовать следующую формальную запись $A_i = A_i(V_i)$. Особенность проблемной ситуации, на решение которой направлено данное исследование, состоит в том, что на момент принятия решения о состоянии объекта исследования, прямые признаки, составляющие множества V_i, неизвестны. Они, например, еще не проявились или по различным объективным причинам не могут быть измерены и определены. Следовательно, появление некоторого состояния природы A_i при экспериментальном фармакологическом воздействии можно рассматривать как случайное событие, имеющее вероятность $P(A_i)$. Эта вероятность может быть оценена в результате специально спланированных статистических исследований или определена при помощи метода экспертных оценок. В результате на первом этапе должны быть формально определены следующие конструкции: множество состояний $A = \{A_1, A_2, ..., A_m\}$, в одном из которых может оказаться объект исследования в результате фармакологического воздействия; множества прямых признаков V_i, $1 \le i \le m$, определяющих эти состояния $A_i = A_i(V_i)$; вероятности $P(A_1), P(A_2), ..., P(A_m)$, соответствующие возможным состояниям объекта исследования. Множество возможных состояний $A = \{A_1, A_2, ..., A_m\}$ целесообразно формировать таким образом, чтобы случайные события $A_1, A_2, ..., A_m$ образовывали полную группу событий, для которой выполняется условие $P(A_1) + P(A_2) + ... + P(A_m) = 1$. На этом первый этап можно считать законченным.

На втором этапе определяется множество возможных решений и функции потерь. В общем случае число возможных решений n является конечным и не менее двух. В данной практической задаче можно рассмотреть частный случай, когда число возможных решений n можно принять равным числу состояний природы m, т.е. $n = m$. При этом каждому состоянию A_i будет соответствовать решение E_i, поддерживающее идентификацию этого состояния. Формальная запись множества возможных решений имеет вид $E = \{E_1, E_2, ..., E_m\}$. Каждое принятое решение имеет последствия, которые должны быть количественно оценены. Можно предположить, что решению E_j, $1 \le j \le n$ при состоянии A_i, $1 \le i \le m$ соответствуют потери $L(A_i, E_j)$. Функция $L(A_i, E_j)$, называемая функцией потерь, должна быть заранее определена для всех возможных комбинаций A_i и E_j. Обычно эту функцию удобно задавать в табличном виде (табл. 1).

Таблица 1. Пример табличного представления функции потерь $L(A_i, E_j)$

Состояния A_i	Вероятности $P(A_i)$	Решения E_j			
		E_1	E_2	…	E_n
A_1	$P(A_1)$	$L(A_1, E_1)$	$L(A_1, E_2)$	…	$L(A_1, E_n)$
A_2	$P(A_2)$	$L(A_2, E_1)$	$L(A_2, E_2)$	…	$L(A_2, E_n)$
…	…	…	…	…	…
A_m	$P(A_m)$	$L(A_m, E_1)$	$L(A_m, E_2)$	…	$L(A_m, E_n)$

Количественное выражение функции потерь в случае, когда речь идет о последствиях решений, не поддающихся физическим измерениям, может представлять нетривиальную задачу. Для ее решения, например, можно использовать субъективные измерения на основе метода парных сравнений [4]. Этот метод позволяет проранжировать все возможные исходы от принимаемых решений по степени тяжести последствий на основании обобщения экспертных оценок. В результате можно поручить измерения значений функции потерь в шкале порядка. Очевидно, что правильное решение не влечет за собой потерь, и, следовательно, в этом случае значение функции потерь равно нулю. Относительно этого значения ранжируются остальные значения этой функции. В дальнейшем значения функции потерь используются как весовые коэффициенты при вычислении средних потерь и функции риска.

На третьем этапе заданные значения функции потерь $L(A_i, E_j)$ и известные априорные вероятности состояний природы $P(A_1), P(A_2), ..., P(A_m)$ позволяют вычислить средние потери $F(A, E_j)$ от принимаемого решения E_j. Формула для вычисления имеет вид: $F(A, E_j) = \sum_{i=1}^{m} P(A_i) L(A_i, E_j)$. Средние потери вычисляются для каждого решения E_j. Наилучшим является решение E_j, обеспечивающее минимальные средние потери $F(A, E_j)$. Таким образом, при завершении третьего этапа на основании заданных значений функции потерь $L(A_i, E_j)$ и вероятностей $P(A_1), P(A_2), ..., P(A_m)$, которые соответствуют возможным состояниям объекта исследования, решается предварительная задача по определению наилучшего решения E_j, основанная на минимальном значении средних потерь от принимаемого решения. В данном случае лицо, принимающее решение, не совершает действий по уточнению свих знаний о действительном состоянии объекта исследования путем проведения дополнительного эксперимента.

На четвертом этапе проводится дополнительный эксперимент с целью получения информации и уточнения знаний о состоянии объекта исследования. Под дополнительным экспериментом понимаются продуманные, обоснованные и спланированные действия, для которых должны быть заранее определены возможные исходы (реализации). Исходы эксперимента определяются на основании множества косвенных признаков. Формально результаты эксперимента описываются множеством $Z=\{Z_1, Z_2, ..., Z_k\}$, где Z_u, $1 \le u \le k$ – определенная реализация дополнительного эксперимента. Число возможных исходов эксперимента k является конечным и не менее двух. Очевидно, что реализации дополнительного эксперимента являются случайными событиями и их вероятность зависит от состояния, в котором находится объект исследования. Эта зависимость может быть представлена условными вероятностями $P(Z_u/A_i)$. $1 \le u \le k$, $1 \le i \le m$, которые также определяются на основании статистического анализа результатов проведенного дополнительного эксперимента. Возможные состояния объекта исследования A_i, $1 \le i \le m$, реализации дополнительного эксперимента Z_u, $1 \le u \le k$ и условные вероятности $P(Z_u/A_i)$ удобно представить в табличном виде (табл. 2).

Таблица 2. Пример табличного представления условных вероятностей $P(Z_u/A_i)$

Состояния A_i	Реализации эксперимента Z_u			
	Z_1	Z_2	…	Z_u
A_1	$P(Z_1/A_1)$	$P(Z_2/A_1)$	…	$P(Z_u/A_1)$
A_2	$P(Z_1/A_2)$	$P(Z_2/A_2)$	…	$P(Z_u/A_2)$
…	…	…	…	…
A_m	$P(Z_1/A_m)$	$P(Z_2/A_m)$	…	$P(Z_u/A_m)$

Вычисление и табличное представление условных вероятностей для реализаций дополнительного эксперимента завершают четвертый этап.

При принятии решения на основе результатов дополнительного эксперимента используются следующие математические конструкции: множество возможных состояний объекта исследования $A=\{A_1, A_2, ..., A_m\}$ и соответствующие им вероятности $P(A_1), P(A_2), ..., P(A_m)$; множество возможных решений $E=\{E_1, E_2, ..., E_m\}$ и соответствующие этим решениям функции потерь $L(A_i, E_j)$, $1 \le i \le m$, $1 \le j \le m$; множество исходов дополнительного эксперимента $Z=\{Z_1, Z_2, ..., Z_k\}$; условные вероятности исходов дополнительного эксперимента $P(Z_u/A_i)$, $1 \le u \le k$, $1 \le i \le m$. Для того, чтобы формализовать эту задачу, необходимо определить решающее правило, определяющее, какое решение E_j следует принять при каждом из возможных исходов Z_u дополнительного эксперимента. Правило, определяющее решение E_j при исходе эксперимента Z_u, называется решающей функцией. На пятом этапе определяются все возможные решающие функции. Количество всех возможных решающих функций можно рассматривать как количество размещений с повторениями из n классов по k элементов. Следовательно, общее число решающих функций v можно вычислить по формуле $v=n^k$. Формальное описание множества решающих функций имеет вид $D=\{D_1, D_2, ..., D_v\}$. Каждая решающая функция D_x, $1 \le x \le v$ из множества D представляет собой отображение множества исходов дополнительного эксперимента $Z=\{Z_1, Z_2, ..., Z_k\}$ на множество решений $E=\{E_1, E_2, ..., E_m\}$, т.е. $D: Z \to E$. Понятие решающей функции следует пояснить на примере. Предположим, что множество решений состоит из двух элементов E_1 и E_2. Множество исходов эксперимента включает в себя три элемента Z_1, Z_2, Z_3. В данном примере $n=2$ и $k=3$, следовательно, количество возможных решающих функций равно $v=2^3=8$. Любая решающая функция D_x из множества D может быть задана в виде множества упорядоченных пар индексов (u, j), где u – индекс, определяющий номер исхода дополнительного эксперимента Z_u; j – индекс, определяющий номер возможного решения E_j. Например, одна из возможных решающих функций D_x может определять следующее правило: при исходе дополнительного эксперимента Z_1 принимается решение E_1; при исходе Z_2 принимается решение E_1; при исходе Z_3 принимается решение E_2. Математическая запись этой решающей функции имеет вид $D_x = \{(1, 1), (2, 1), (3, 2)\}$. Для рассматриваемого примера можно записать еще семь решающих функций. Формальным описанием всех возможных решающих функций завершается пятый этап.

На шестом этапе из множества всех возможных решающих функций необходимо выбрать такую функцию, которая позволит принять наиболее адекватное решение. Если, например, выбрана некоторая решающая функция D_x, то тем самым для каждого исхода дополнительного эксперимента Z_u, $1 \le u \le k$ определено решение E_j, $1 \le j \le m$, которому при состоянии объекта исследования A_i будет соответствовать функция потерь $L(A_i, E_j)$. Исход дополнительного эксперимента является случайной величиной с условной вероятностью $P(Z_u / A_i)$, следовательно, с этой же вероятностью будут иметь место и потери $L(A_i, E_j)$. Для выбора оптимальной решающей функции следует учитывать все возможные исходы дополнительного эксперимента. Поэтому необходимо найти потери для каждой решающей функции, которые в данном случае называются функцией риска. Эта функция обозначается $R(D_x, A_i)$ и вычисляется по следующей формуле: $R(D_x, A_i) = \sum_{u=1}^{k} P(Z_u / A_i) L(A_i / E_j)$. Применение приведенной формулы требует пояснения. Для использования формулы необходимо зафиксировать индекс i, определяющий состояние объекта исследования A_i. При изменении значений индекса исхода эксперимента u от 1 до k индекс возможного решения j изменяется в соответствии с правилом, определенным решающей функцией D_x. Например, если $i = 1$, а решающая функция имеет вид $D_1 = \{(1, 1), (2, 1), (3, 2)\}$, то формула для вычисления функции риска $R(D_1, A_1)$ для решающей функции D_1 при состоянии объекта A_1 в развернутом виде будет иметь следующий вид:

$$R(D_1, A_1) = P(Z_1 / A_1) L(A_1, E_1) + P(Z_2 / A_1) L(A_1, E_1) + P(Z_3 / A_1) L(A_1, E_2)$$

Функция риска $R(D_x, A_i)$ учитывает потери, которые дает решающая функция D_x только при одном состоянии объекта исследования A_i. Для определения наилучшей решающей функции используется ожидаемый средний риск, под которым понимается средний риск с учетом всех возможных состояний объекта исследования $A_1, A_2, ..., A_m$ и вероятностей их появления $P(A_1), P(A_2), ..., P(A_m)$. Ожидаемый средний риск обозначается $\overline{R}(D_x, A)$ и вычисляется по следующей формуле: $\overline{R}(D_x, A) = \sum_{i=1}^{m} P(A_i) R(D_x / A_i)$. Ожидаемый средний риск вычисляется для всех возможных решающих функций. В результате анализа полученных значений выбирается такая оптимальная решающая функция $D_{опт}$, которая обеспечивает минимальный ожидаемый средний риск, т.е. $D_{опт} : \overline{R}(D_x, A) \to \min$.

Это правило называется критерием минимального среднего риска, который позволяет вычислить решающую функцию, оптимальную для данного критерия. На этом алгоритм, реализующий способ принятия решения о состоянии объекта исследования на основании критерия минимального среднего риска, можно считать законченным.

После того, как станут доступными для измерения прямые признаки, можно точно идентифицировать состояние объекта исследования. Это позволит сопоставить результат идентификации, сделанный по косвенным признакам с использованием критерия минимального среднего риска, с идентификацией, сделанной по прямым признакам. На этом основании можно судить о качестве дополнительного эксперимента и информативности косвенных признаков, определяющих его исход. Следовательно, разработанный алгоритм также может служить основанием для оценки информативности дополнительного эксперимента и его исходов [2, 8].

Пример практического применения критерия минимального среднего риска

Рассмотрим эксперимент, в котором на объект исследования оказывается детерминированное фармакологическое воздействие. В результате объект может оказаться в одном из трех возможных состояний. На первом этапе для множества возможных состояний $A=\{A_1, A_2, A_3\}$ определяются вероятности $P(A_1), P(A_2), P(A_3)$. Эти вероятности могут быть определены на основании теоретических положений или в результате специально спланированных статистических исследований. Для примера примем, что вероятности состояний объекта исследования после фармакологического воздействия равны $P(A_1)=0{,}3$; $P(A_2)=0{,}4$; $P(A_3)=0{,}3$. Очевидно, что события A_1, A_2, A_3 образуют полную группу событий, поскольку выполняется условие $P(A_1)+P(A_2)+P(A_3)=0{,}3+0{,}4+0{,}3=1$.

На втором этапе необходимо определить множество возможных решений и функции потерь. Возможные решения составляют множество $E=\{E_1, E_2, E_3\}$, в котором: решение E_1 принимается в пользу состояния A_1; решение E_2 принимается в пользу состояния A_2; решение E_3 принимается в пользу состояния A_3. Для рассматриваемого примера предположим, что отношения между функциями потерь $L(A_i, E_j)$ определены в результате субъективных измерений на основе метода парных сравнений для всех возможных комбинаций A_i и E_j [1]. Если состояния A_1, A_2, A_3 расположены в порядке возрастания исследуемого свойства, то можно воспользоваться простым правилом. Для правильного решения функция потерь минимальна и равна нулю, т.е. $L(A_1, E_1)=L(A_2, E_2)=L(A_3, E_3)=0$. При ошибке на две градации функция потерь больше, чем при ошибке на одну градацию, т.е., например, $L(A_1, E_3)>L(A_1, E_2)>L(A_1, E_1)=0$. На основании этого правила ранжированный ряд для функций потерь имеет следующий вид:

$$L(A_1, E_3)=L(A_3, E_1)>L(A_1, E_2)=L(A_2, E_1)=L(A_3, E_2)=L(A_2, E_3)>L(A_1, E_1)=L(A_2, E_2)=L(A_3, E_3).$$

Ранжированным функциям потерь $L(A_i, E_j)$ без нарушения записанных выше отношений между ними можно поставить в соответствие следующие числовые значения: $L(A_1, E_3)=L(A_3, E_1)=2$; $L(A_1, E_2)=L(A_2, E_1)=L(A_3, E_2)=L(A_2, E_3)=1$; $L(A_1, E_1)=L(A_2, E_2)=L(A_3, E_3)=0$. Значения функции потерь для рассматриваемого примера представлены в табл. 3.

Таблица 3. Пример табличного представления функции потерь $L(A_i, E_j)$

Состояния A_i	Вероятности $P(A_i)$	Решения E_j		
		E_1	E_2	E_3
A_1	$P(A_1) = 0{,}3$	$L(A_1, E_1) = 0$	$L(A_1, E_2) = 1$	$L(A_1, E_3) = 2$
A_2	$P(A_2) = 0{,}4$	$L(A_2, E_1) = 1$	$L(A_2, E_2) = 0$	$L(A_2, E_3) = 1$
A_3	$P(A_3) = 0{,}3$	$L(A_3, E_1) = 2$	$L(A_3, E_2) = 1$	$L(A_3, E_3) = 0$

На третьем этапе проведем вычисление средних потерь $F(A, E_j)$, $1\le j\le m$ для каждого из решений E_1, E_2, E_3. Расчёт производится по формуле $F(A, E_j)=\sum_{i=1}^{3}P(A_i)L(A_i, E_j)$, $1\le j\le m$. В этой формуле значения функции потерь $L(A_i, E_j)$, приведенные в таблице 3, используются в качестве весовых коэффициентов. Произведем вычисления. Запишем формулы для средних потерь в развернутом виде.

$$F(A, E_1)=P(A_1)\,L(A_1, E_1)+P(A_2)\,L(A_2, E_1)+P(A_3)\,L(A_3, E_1)=0{,}3\cdot 0+0{,}4\cdot 1+0{,}3\cdot 2=1.$$

$$F(A, E_2)=P(A_1)\,L(A_1, E_2)+P(A_2)\,L(A_2, E_2)+P(A_3)\,L(A_3, E_2)=0{,}3\cdot 1+0{,}4\cdot 0+0{,}3\cdot 1=0{,}6.$$

$$F(A, E_3)=P(A_1)\,L(A_1, E_3)+P(A_2)\,L(A_2, E_3)+P(A_3)\,L(A_3, E_3)=0{,}3\cdot 2+0{,}4\cdot 1+0{,}3\cdot 0=1.$$

На третьем этапе решается предварительная задача по определению наилучшего решения, основанная на минимальном значении средних потерь от принимаемого решения. В результате вычислений установлено, наименьшие средние потери $F(A, E_2)=0{,}6$ соответствуют решению E_2 в пользу состояния A_2. В данном случае не совершаются действия по уточнению свих знаний о действительном состоянии объекта исследования путем проведения дополнительного эксперимента. На этом третий этап можно считать выполненным.

На четвертом этапе проводится дополнительный эксперимент. Его целью является уточнение знаний об объекте исследования. В результате дополнительного эксперимента измеряются косвенные признаки, из которых в результате их группировки и сочетания формируются и определяются события, называемые в дальнейшем исходами (реализациями). В рассматриваемом примере будем полагать, что дополнительный эксперимент имеет три исхода $Z = \{Z_1, Z_2, Z_3\}$. Каждый исход характеризуется условной вероятностью $P(Z_u/A_i)$, $1 \le u \le 3$, $1 \le i \le 3$, которая зависит от состояния, в котором находится объект исследования. Условные вероятности удобно представить в виде таблицы. Значения условных вероятностей для рассматриваемого примера приведены в табл. 4.

Таблица 4. Значения условных вероятностей $P(Z_u/A_i)$, $1 \le u \le 3$, $1 \le i \le 3$ для рассматриваемого примера

Состояния A_i	Исходы дополнительного эксперимента Z_u		
	Z_1	Z_2	Z_3
A_1	$P(Z_1/A_1) = 0{,}6$	$P(Z_2/A_1) = 0{,}3$	$P(Z_3/A_1) = 0{,}1$
A_2	$P(Z_1/A_2) = 0{,}1$	$P(Z_2/A_2) = 0{,}6$	$P(Z_3/A_2) = 0{,}3$
A_3	$P(Z_1/A_3) = 0{,}1$	$P(Z_2/A_3) = 0{,}2$	$P(Z_3/A_3) = 0{,}7$

Следует отметить, что условные вероятности исходов определяются либо на основании теоретических положений, либо на основании специально спланированных статистических исследований. На этом четвертый этап исследования можно считать законченным.

На пятом этапе необходимо определить и формально записать все возможные решающие функции. Поскольку число возможных решений состоит из элементов E_1, E_2, E_3, т.е. $n = 3$, а множество исходов эксперимента включает в себя три элемента Z_1, Z_2, Z_3, т.е. $k = 3$, то количество возможных решающих функций равно $v = 3^3 = 27$. Формальное представление всех 27 решающих функций имеет вид:

$$D_1 = \{(1,1),(2,1),(3,1)\};\ D_2 = \{(1,1),(2,1),(3,2)\};\ D_3 = \{(1,1),(2,1),(3,3)\};\ D_4 = \{(1,1),(2,2),(3,1)\};$$
$$D_5 = \{(1,1),(2,2),(3,2)\};\ D_6 = \{(1,1),(2,2),(3,3)\};\ D_7 = \{(1,1),(2,3),(3,1)\};\ D_8 = \{(1,1),(2,3),(3,2)\};$$
$$D_9 = \{(1,1),(2,3),(3,3)\};\ D_{10} = \{(1,2),(2,1),(3,1)\};\ D_{11} = \{(1,2),(2,1),(3,2)\};\ D_{12} = \{(1,2),(2,1),(3,3)\};$$
$$D_{13} = \{(1,2),(2,2),(3,1)\};\ D_{14} = \{(1,2),(2,2),(3,2)\};\ D_{15} = \{(1,2),(2,2),(3,3)\};\ D_{16} = \{(1,2),(2,3),(3,1)\};$$
$$D_{17} = \{(1,2),(2,3),(3,2)\};\ D_{18} = \{(1,2),(2,3),(3,3)\};\ D_{19} = \{(1,3),(2,1),(3,1)\};\ D_{20} = \{(1,3),(2,1),(3,2)\};$$
$$D_{21} = \{(1,3),(2,1),(3,3)\};\ D_{22} = \{(1,3),(2,2),(3,1)\};\ D_{23} = \{(1,3),(2,2),(3,2)\};\ D_{24} = \{(1,3),(2,2),(3,3)\};$$
$$D_{25} = \{(1,3),(2,3),(3,1)\};\ D_{26} = \{(1,3),(2,3),(3,2)\};\ D_{27} = \{(1,3),(2,3),(3,3)\}.$$

На этом пятый этап вычислений можно считать завершенным.

На шестом заключительном этапе необходимо вычислить функции риска $R(D_x, A_i)$ для каждой решающей функции. Вычисления производятся по следующей формуле: $R(D_x, A_i) = \sum_{u=1}^{3} P(Z_u/A_i) L(A_i/E_j)$. Для проведения вычислений необходимо: выбрать одну из решающих функций; зафиксировать индекс i, определяющий состояние объекта исследования A_i; значения индекса u, определяющего исход дополнительного эксперимента Z_u, изменять в пределах от 1 до 3, а значения индекса j, определяющего решение E_j и соответствующее значение функции потерь $L(A_i/E_j)$, выбирается в соответствии с решающей функцией D_x.

Выполним необходимые вычисления. Для решающей функции $D_1 = \{(1,1),(2,1),(3,1)\}$ и состояния A_1 вычисления функции риска $R(D_1, A_1)$, записанные подробно, имеют вид:

$$R(D_1, A_1) = P(Z_1/A_1)L(A_1, E_1) + P(Z_2/A_1)L(A_1, E_1) + P(Z_3/A_1)L(A_1, E_1) = 0{,}6 \cdot 0 + 0{,}3 \cdot 0 + 0{,}1 \cdot 0 = 0.$$

Для состояния A_2:

$$R(D_1, A_2) = P(Z_1/A_2)L(A_2, E_1) + P(Z_2/A_2)L(A_2, E_1) + P(Z_3/A_2)L(A_2, E_1) = 0{,}1 \cdot 1 + 0{,}6 \cdot 1 + 0{,}3 \cdot 1 = 1.$$

Для состояния A_3:
$R(D_1, A_3) = P(Z_1/A_3)L(A_3, E_1) + P(Z_2/A_3)L(A_3, E_1) + P(Z_3/A_3)L(A_3, E_1) = 0{,}1 \cdot 2 + 0{,}2 \cdot 2 + 0{,}7 \cdot 2 = 2$.
Поскольку число решающих функций 27, а число возможных состояний равно 3, то общее число функций риска, вычисляемых для всех состояний A_1, A_2, A_3, равно $27 \cdot 3 = 81$.
Для нахождения наилучшей решающей функции вычислим ожидаемый средний риск по формуле: $\overline{R}(D_x, A) = \sum_{i=1}^{3} P(A_i)R(D_x/A_i)$. Так, например, для решающей функции $D_1 = \{(1,1), (2,1), (3,1)\}$ при известных априорных вероятностях $P(A_1) = 0{,}3$; $P(A_2) = 0{,}4$; $P(A_3) = 0{,}3$ подробно записанные вычисления ожидаемого среднего риска имеют вид:

$$\overline{R}(D_1, A) = \sum_{i=1}^{3} P(A_i)R(D_1/A_i) = P(A_1)R(D_1/A_1) + P(A_2)R(D_1/A_2) + P(A_3)R(D_1/A_3) = 0{,}3 \cdot 0 + 0{,}4 \cdot 1 + 0{,}3 \cdot 2 = 1$$

.
Аналогичным образом просчитываются значения ожидаемого среднего риска для 26 оставшихся решающих функций. Очевидно, что для автоматизации расчетов удобно использовать вычислительные возможности современных информационных технологий, например, табличного процессора Excel.
Для рассматриваемого примера наилучшей будет решающая функция $D_6 = \{(1,1), (2,2), (3,3)\}$. Для нее функции риска $R(D_x, A_i)$ принимают следующие значения:

$$R(D_6, A_1) = P(Z_1/A_1)L(A_1, E_1) + P(Z_2/A_1)L(A_1, E_2) + P(Z_3/A_1)L(A_1, E_3) = 0{,}6 \cdot 0 + 0{,}3 \cdot 1 + 0{,}1 \cdot 2 = 0{,}5.$$

$$R(D_6, A_2) = P(Z_1/A_2)L(A_2, E_1) + P(Z_2/A_2)L(A_2, E_2) + P(Z_3/A_2)L(A_2, E_3) = 0{,}1 \cdot 1 + 0{,}6 \cdot 0 + 0{,}3 \cdot 1 = 0{,}4.$$

$$R(D_6, A_3) = P(Z_1/A_3)L(A_3, E_1) + P(Z_2/A_3)L(A_3, E_2) + P(Z_3/A_3)L(A_3, E_3) = 0{,}1 \cdot 2 + 0{,}2 \cdot 1 + 0{,}7 \cdot 0 = 0{,}4.$$

Средний ожидаемый риск $\overline{R}(D_6, A)$ для решающей функции $D_6 = \{(1,1), (2,2), (3,3)\}$ имеет вид:

$$\overline{R}(D_6, A) = P(A_1)R(D_6/A_1) + P(A_2)R(D_6/A_2) + P(A_3)R(D_6/A_3) = 0{,}3 \cdot 0{,}5 + 0{,}4 \cdot 0{,}4 + 0{,}3 \cdot 0{,}4 = 0{,}43.$$

Решающая функция $D_6 = \{(1,1), (2,2), (3,3)\}$ имеет наименьший ожидаемый средний риск и по этому критерию является наилучшей из 27 возможных решающих функций. В соответствии с этой решающей функцией формулируется следующее решающее правило, по которому на основании результатов дополнительного эксперимента определяется состояние объекта исследования: при исходе дополнительного эксперимента Z_1 принимается решение E_1 в пользу состояния A_1; при исходе Z_2 принимается решение E_2 в пользу состояния A_2; при исходе Z_3 принимается решение E_3 в пользу состояния A_3. Полученное правило является объективным формальным основанием для принятия решения о состоянии объекта исследования в фармакологическом эксперименте на основании косвенных признаков [4, 6].

Заключение

В результате проведенного исследования разработан способ обоснования статистического решения, направленного на определение состояния объекта исследования в условиях неопределенности, когда прямые признаки, определяющие состояние объекта исследования по объективным причинам неизвестны, а информация о косвенных признаках, полученная в результате дополнительного эксперимента, имеет вероятностный характер. Разработанный способ, кроме основного предназначения, может также применяться в качестве критерия оценки информативности возможных дополнительных экспериментов. Способ идентификации состояния объектов на основе критерия минимального среднего риска может представлять практический интерес для научных работников, осуществляющих исследования в области фармакологии и доказательной медицины, использующих в своей работе статистические методы анализа экспериментальных данных и принятия статистических решений.

Литература (references)

1. Боровков А.А. Математическая статистика. Оценка параметров. Проверка гипотез. – М.: Физматлит, 2007. – 704 с. [Borovkov A.A. *Matematicheskaja statistika. Ocenka parametrov. Proverka gipotez.* Mathematical statistics. Parameter estimation. Verification of hypotheses. – Moscow: Fizmatlit, 2007. – 704 p. (in Russian)]
2. Волкова П.А., Шипунов А.Б. Статистическая обработка данных в учебно-исследовательских работах. – М.: Форум, 2012. – 96 с. [Volkova P.A., Shipunov A.B. *Statisticheskaja obrabotka dannyh v uchebno-issledovatel'skih rabotah.* Statistical data processing in educational and research works. – Moscow: Forum, 2012. – 96 p. (in Russian)]
3. Гланц С. Медико-биологическая статистика. – М.: Практика, 1998. – 459 с. [Glanc S. *Mediko-biologicheskaja statistika.* Medico-biological statistics. – Moscow: Praktika, 1998. – 459 p. (in Russian)]
4. Евланов Л.Г. Основы теории принятия решений. – М.: Москва, 1979. – 212 с. [Evlanov L.G. *Osnovy teorii prinjatija reshenij.* Fundamentals of decision-making theory. – Moscow, 1979. – 212 p. (in Russian)]
5. Петри А., Сэбин К. Наглядная статистика в медицине. – М.: Гэотар-Мед, 2002. – 144 с. [Petri A., Sjebin K. Nagljadnaja statistika v medicine. Visual statistics in medicine. – Moscow: GEOTAR-Med, 2002. – 144 p. (in Russian)]
6. Плавинский С.Л. Введение в биостатистику для медиков. – М.: Акварель, 2011. – 584 с. [Plavinskij S.L. *Vvedenie v biostatistiku dlja medikov.* Introduction to biostatistics for physicians. – Moscow: Akvarel, 2011. – 584 p. (in Russian)]
7. Платонов А.Е. Статистический анализ в медицине и биологии: задачи, терминология, логика, компьютерные методы. – М.: Изд-во РАМН, 2000. – 52 с. [Platonov A.E. *Statisticheskij analiz v medicine i biologii: zadachi, terminologija, logika, komp'juternye metody.* Statistical analysis in medicine and biology: problems, terminology, logic, computer methods. – Moscow: publishing house of the Russian Academy of medical Sciences, 2000. – 52 p. (in Russian)]
8. Шаропин К.А., Берестнева О.Г., Шкатова Г.И. Визуализация результатов экспериментальных исследований // Известия Томского политехнического университета. – 2010 – Т.316, №5. – С. 172-176. [Sharopin K.A., Berestneva O.G., Shkatova G.I. *Izvestija Tomskogo politehnicheskogo universiteta.* Izvestiya Tomsk Polytechnic University, 2010. – V.316, N5. – P. 172-176. (in Russian)]
9. McKillup S. Statistics explained. An introductory guide for life scientists. – England: Cambridge University Press, 2005. – 267 p. [(in Russian)]
10. Rowntree D. Statistics without tears. – England: Clays, 2000. – 195 p. [(in Russian)]
11. van Emden H. Statistics for terrified biologists. – USA: Blackwell Publishing, 2008. – 343 p. [(in Russian)]

Информация об авторах

Лямец Леонид Леонидович – кандидат технических наук, доцент, заведующий кафедрой физики, математики и медицинской информатики ФГБОУ ВО «Смоленский государственный медицинский университет» Минздрава России. E-mail: LLL190965@yandex.ru

Евсеев Андрей Викторович – доктор медицинских наук, профессор, заведующий кафедрой нормальной физиологии, заведующий научно-исследовательским центром ФГБОУ ВО «Смоленский государственный медицинский университет» Минздрава России. E-mail: hypoxia@yandex.ru

Данилов Андрей Игоревич – кандидат медицинских наук, ассистент кафедры клинической фармакологии ФГБОУ ВО «Смоленский государственный медицинский университет» Минздрава России. E-mail: dr.danandr@yandex.ru