

# Інформаційні технології в медицині

УДК 519.22: 612.17-017.26:618.3

В.Ю. Дубницький<sup>1</sup>, І.Д. Дужий<sup>2</sup>, С.В. Харченко<sup>2</sup>, А.І. Ходырев<sup>1</sup>

<sup>1</sup> Харківський учебно-научний інститут банківського дела  
ГВУЗ «Університет банківського дела», Харків

<sup>2</sup> Медичинський інститут Сумського державного університета, Суми

## ОЦЕНКА СТАТИСТИЧЕСКОЙ СВЯЗИ МЕЖДУ ПОКАЗАТЕЛЯМИ ЧАСТОТЫ ГОСПИТАЛИЗАЦИИ И ТЕРРИТОРИАЛЬНОЙ РАСПРОСТРАНЁННОСТЬЮ ЯЗВЕННОЙ БОЛЕЗНИ

*Установлено наличие статистической связи между показателями частоты госпитализации и территориальной распространённостью язвенной болезни. Источником данных для анализа были сведения государственной статистики о госпитализации больных с диагнозом язвенной болезни, осложненной кровотечением, для различных уровней административно-территориального деления Украины. Уровень статистической связи оценивали методами непараметрической статистики. В работе использованы критерии Колмогорова-Смирнова, Крамера-Мизеса, Андерсона-Дарлинга, Шапиро-Уилка, Шапиро-Френсиса, Краскелла-Уоллеса, критерия для медиан Муда, коэффициент ранговой корреляции Спирмена.*

**Ключевые слова:** язвенная болезнь, непараметрическая статистика, критерии Колмогорова-Смирнова, Крамера-Мизеса, Андерсона-Дарлинга, Шапиро-Уилка, Шапиро-Френсиса, Краскелла-Уоллеса, критерий для медиан Муда, коэффициент ранговой корреляции Спирмена.

### Введение

В условиях реформы здравоохранения в Украине, особенно при нестабильном экономическом и демографическом развитии, определение областей или районов Украины, насыщенных жителями, болеющими осложненной язвенной болезнью (ЯБ), является важной задачей, требующей комплексного подхода. Важность этой задачи возрастает в связи с переходом к страховой медицине. Одним из возможных способов решения этой задачи может быть применение популяционного эпидемиологического анализа, дополненного современными статистическими методами.

**Характеристика объекта исследования.** С течением времени во всем мире количество ежегодно госпитализируемых больных с язвенными кровотечениями уменьшается [1]. Следует принять во внимание, что если анализировать территориальные особенности стран, то тотального уменьшения госпитализаций нет, поскольку для одних областей госпитализации по поводу этой патологии не изменяется, а для других – уменьшается или растет [2 – 6]. Более того, существуют местные «лидеры» в госпитализациях. Например, в Греции наблюдается 160 случаев кровотечения на 100 000 взрослого населения ежегодно [7]. Такие эпидемиологические вариации побуждают к поиску географических областей во всем мире, в нашем случае, в Украине, которые должны быть зачислены к рисковому местностям.

В дальнейшем это приведет к профилактико-организационным изменениям в процессе медико-экономического планирования. Поэтому определение «безопасных» местностей ведет к исследованию местных организационных особенностей медицинской помощи таким больным. Изучение тенденций в изменении количества госпитализаций позволит дать рекомендации по рациональной структуре и месторасположению единого центра желудочно-кишечных кровотечений [8, 9]. Изучение статистики госпитализаций отображает нагрузку на инфраструктуру лечебно-профилактических учреждений и позволяет выделить административно-территориальные единицы, наиболее восприимчивые к ЯБ.

**Анализ литературы.** В данной работе приняты следующие термины. Для диагноза «Желудочно-кишечное кровотечение» использованы коды K25 – K28 Международной классификации болезней 10-го пересмотра Всемирной Организации Здравоохранения (Женева, Швейцария, 1989 г.).

В рамках данной работы использованы такие относительные показатели заболеваемости, как  $N_p$  – количество госпитализированных на 10 000 чел. и  $Q_s$  – количество госпитализированных на 1000 кв.км. Для этих показателей составлены временные ряды, характеризующие тенденцию относительного изменения заболеваемости ЯБ во времени и пространстве [10]. Связь между этими показателями и особенностями их совместного использования для изучения

такого явления, как госпитализационная заболеваемость, насколько известно авторам данного сообщения, в литературе не рассмотрена.

**Постановка задачи.** Целью работы является изучение статистической взаимосвязи между показателями заболеваемости для различных уровней административно-территориального деления (АТД).

### Полученные результаты и дискуссия

В работе использованы данные, полученные Государственной службой статистики Украины (<http://www.ukrstat.gov.ua/>), Центром медицинской статистики Министерства здравоохранения Украины (<http://medstat.gov.ua/>), Государственным земельным агентством Украины (<http://land.gov.ua/>) с января 1999г. по январь 2014 г.г. Эти данные относятся к Украине в целом, одной из её областей и административному району выбранной для изучения области. В обработанном виде они приведены в табл. 1.

Таблица 1

Исходные данные  
о количестве госпитализированных больных  
с диагнозом «Желудочно-кишечное кровотечение»

| Годы | Количество случаев госпитализации |          |       |                 |          |       |
|------|-----------------------------------|----------|-------|-----------------|----------|-------|
|      | На 10 000 человек                 |          |       | на 1 000 кв.км. |          |       |
|      | Ук-раина                          | Об-ласть | Район | Ук-раина        | Об-ласть | Район |
|      | Npu                               | Nps      | Npg   | Qsu             | Qss      | Qsg   |
| 1998 | 3,91                              | 4,18     | 4,13  | 32,50           | 24,18    | 16,63 |
| 1999 | 3,99                              | 4,52     | 3,90  | 32,91           | 25,85    | 15,48 |
| 2000 | 5,09                              | 6,14     | 5,86  | 41,57           | 34,59    | 22,94 |
| 2001 | 5,08                              | 6,91     | 3,56  | 41,14           | 38,44    | 13,76 |
| 2002 | 4,28                              | 6,12     | 2,84  | 34,34           | 33,55    | 10,89 |
| 2003 | 4,21                              | 6,09     | 4,85  | 33,57           | 32,91    | 18,35 |
| 2004 | 4,51                              | 6,30     | 4,91  | 35,66           | 33,55    | 18,35 |
| 2005 | 4,56                              | 5,62     | 2,34  | 35,83           | 29,52    | 8,60  |
| 2006 | 4,55                              | 5,90     | 2,06  | 35,48           | 30,60    | 7,45  |
| 2007 | 4,47                              | 6,44     | 1,60  | 34,72           | 33,00    | 5,73  |
| 2008 | 4,55                              | 6,19     | 1,79  | 35,16           | 31,37    | 6,31  |
| 2009 | 4,67                              | 7,59     | 1,65  | 35,91           | 38,12    | 5,73  |
| 2010 | 4,61                              | 7,14     | 1,00  | 35,31           | 35,54    | 3,44  |
| 2011 | 4,57                              | 6,10     | 2,19  | 34,89           | 30,11    | 7,45  |
| 2012 | 4,80                              | 6,05     | 3,06  | 36,63           | 29,65    | 10,32 |
| 2013 | 4,80                              | 6,21     | 3,10  | 36,56           | 30,15    | 10,32 |

Для решения поставленной задачи был выбран метод однофакторного дисперсионного анализа. Уровням факторов соответствовали элементы административно-территориального деления Украины. Классические методы дисперсионного анализа будут давать корректные результаты только в том случае, когда анализируемая исходная совокупность распределена нормально [11]. В том случае, когда

исходные данные распределены ненормально, рекомендуется для решения поставленной задачи использовать методы непараметрической статистики [12]. Результаты проверки справедливости гипотезы нормального распределения показателя Np приведены в табл. 2. Результаты проверки справедливости гипотезы нормального распределения показателя Qs приведены в табл. 3.

В этих таблицах приняты следующие обозначения: Pф – фактическое значение соответствующего критерия, Pкр – двустороннее критическое значение критерия, соответствующее доверительной вероятности  $\alpha=0,05$ , «Да» – принимается гипотеза нормального распределения исходных данных, «Нет» – эта же гипотеза отвергается. Для выполнения этих и последующих вычислений использована система AtteStat 2013 [13].

Для визуализации данных использованы так называемые усаые ящик Тьюки, показывающие значение медиан ряда распределения и его размах, показанные на оси абсцисс и наименование ряда, показанное на оси ординат. Для показателя Np этот график показан на рис. 1. Для показателя Qs этот график показан на рис. 2.

Значения медиан, соответствующих различным рядам распределений изучаемых показателей, приведены в табл. 4. Из этой таблицы следует, что наибольшая медиана количества госпитализаций соответствует областному уровню АТД (6,13 госпитализированных на 10 000 чел. населения), наименьшая – районному уровню АТД (2,95 госпитализированных на 10 000 чел. населения). Для медиан количества госпитализаций на 1 000 кв.км. ситуация иная: наибольший показатель на уровне страны – 35,56; наименьший на районном уровне АТД – 10,32 соответственно.

Анализируя эти данные, можно выдвинуть следующие нестатистических гипотезы, относящиеся к приведенному массиву данных.

1. Медианы рядов показателей, характеризующих различные уровни АТД, существенно различны не только в целом, но и попарно.

2. Существует статистическая зависимость между численными значениями изучаемых показателей для различных уровней АТД. Для проверки первой гипотезы использовали тест Краскелла-Уоллеса и тест медиан Муда. Результаты приведены в табл. 5.

Результаты, приведенные в табл. 5, позволяют утвердительно ответить на первую часть гипотезы, сформулированной в п.1, но не дают ответа на вопрос: «Какие именно медианы различны попарно?». Для ответа на этот вопрос использован критерий Коновера, результаты применения которого приведены в табл. 6. В табл. 6 «Да» означает, что гипотеза о попарном различии медиан принимается. Из этого следует, что показатели Np и Qs различаются на всех уровнях АТД.

Таблица 2

Проверка справедливости гипотезы нормального распределения показателя  
 Np – количества госпитализированных больных на 10 000 чел.

| Критерий            | Регион  |       |       |         |       |       |       |       |       |
|---------------------|---------|-------|-------|---------|-------|-------|-------|-------|-------|
|                     | Украина |       |       | Область |       |       | Район |       |       |
|                     | Рф      | Ркр   | Вывод | Рф      | Ркр   | Вывод | Рф    | Ркр   | Вывод |
| Колмогорова         | 0,169   | 0,508 | Да    | 0,231   | 0,065 | Да    | 0,133 | 0,805 | Да    |
| Смирнова            | 0,109   | 0,710 | Да    | 0,154   | 0,431 | Да    | 0,133 | 0,758 | Да    |
| Крамера-Мизеса      | 0,072   | 0,506 | Да    | 0,181   | 0,017 | Нет   | 0,044 | 0,758 | Да    |
| Андерсона-Дарлингга | 0,398   | 0,707 | Да    | 0,921   | 0,041 | Нет   | 0,275 | 0,667 | Да    |
| Шапиро-Уилка        | 0,948   | 0,900 | Да    | 0,889   | 0,109 | Да    | 0,958 | 0,799 | Да    |
| Шапиро-Френсиса     | 0,951   | 0,877 | Да    | 0,875   | 0,069 | Да    | 0,965 | 0,669 | Да    |

Таблица 3

Проверка справедливости гипотезы нормального распределения показателя  
 Qs – количества госпитализированных больных на 1 000 кв.км.

| Критерий            | Регион  |       |       |         |       |       |       |       |       |
|---------------------|---------|-------|-------|---------|-------|-------|-------|-------|-------|
|                     | Украина |       |       | Область |       |       | Район |       |       |
|                     | Рф      | Ркр   | Вывод | Рф      | Ркр   | Вывод | Рф    | Ркр   | Вывод |
| Колмогорова         | 0,237   | 0,049 | Нет   | 0,140   | 0,954 | Да    | 0,158 | 0,688 | Да    |
| Смирнова            | 0,237   | 0,021 | Нет   | 0,089   | 0,299 | Да    | 0,158 | 0,377 | Да    |
| Крамера-Мизеса      | 0,137   | 0,023 | Нет   | 0,037   | 0,560 | Да    | 0,069 | 0,561 | Да    |
| Андерсона-Дарлингга | 1,038   | 0,023 | Нет   | 0,253   | 0,537 | Да    | 0,406 | 0,673 | Да    |
| Шапиро-Уилка        | 0,838   | 0,016 | Нет   | 0,965   | 0,531 | Да    | 0,940 | 0,691 | Да    |
| Шапиро-Френсиса     | 0,833   | 0,019 | Нет   | 0,966   | 0,604 | Да    | 0,947 | 0,782 | Да    |

Таблица 4

Значения медиан для рядов распределений количества госпитализированных

| Вид критерия                  | Медианы рядов распределений показателей госпитализации больных |                 |                 |  |                 |                 |
|-------------------------------|--|-----------------|-----------------|--|-----------------|-----------------|
|                               | Np – количества госпитализированных на 10 000чел.              |                 |                 | Qs – количества госпитализированных на 1 000 кв.км |                 |                 |
| Характеристики по уровням АТД | N <sub>pu</sub>  | N <sub>ps</sub> | N <sub>pg</sub> | Q <sub>pu</sub>                                    | Q <sub>pu</sub> | Q <sub>pg</sub> |
| Значения медиан               | 4,55   | 6,13            | 2,95            | 35,59  | 32,14           | 10,32           |

Таблица 5

Результаты проверки гипотезы несущественном различии медиан

| Вид критерия      | Показатели госпитализации больных                 |                         |       |  |                         |       |
|-------------------|---|-------------------------|-------|--|-------------------------|-------|
|                   | Np – количество госпитализированных на 10 000чел. |                         |       | Qs – количество госпитализированных на 1 000 кв.км |                         |       |
| Характеристики    | Значение критерия                                 | Величина P <sub>v</sub> | Вывод | Значение критерия                                  | Величина P <sub>v</sub> | Вывод |
| Краскелла-Уоллеса | 29,441  | $4,04 \cdot 10^{-7}$    | Да    | 35,389   | $2,066 \cdot 10^{-8}$   | Да    |
| Муда для медиан   | 15,5  | 0,0004                  | Да    | 32,0   | $1,125 \cdot 10^{-7}$   | Да    |

Таблица 6

Результаты проверки  
 попарного сравнения медиан

| Сравниваемые пары                 | Значение критерия | Величина P <sub>v</sub> | Вывод | Сравниваемые пары                 | Значение критерия | Величина P <sub>v</sub> | Вывод |
|-----------------------------------|-------------------|-------------------------|-------|-----------------------------------|-------------------|-------------------------|-------|
| N <sub>pu</sub> ↔ N <sub>ps</sub> | 4,992             | $4,72 \cdot 10^{-6}$    | Да    | Q <sub>su</sub> ↔ Q <sub>ss</sub> | 3,918             | 0,33315                 | Да    |
| N <sub>pu</sub> ↔ N <sub>pg</sub> | 3,658             | 0,0003                  | Да    | Q <sub>su</sub> ↔ Q <sub>sg</sub> | 11,483            | $3,89 \cdot 10^{-15}$   | Да    |
| N <sub>ps</sub> ↔ N <sub>pg</sub> | 8,651             | $1,98 \cdot 10^{-11}$   | Да    | Q <sub>ss</sub> ↔ Q <sub>sg</sub> | 7,564             | $7,47 \cdot 10^{-10}$   | Да    |

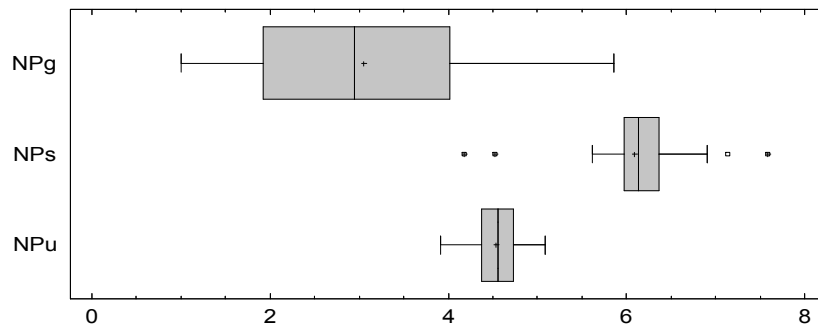


Рис. 1. Усатые ящики Тьюки показателя Np для различных уровней АД

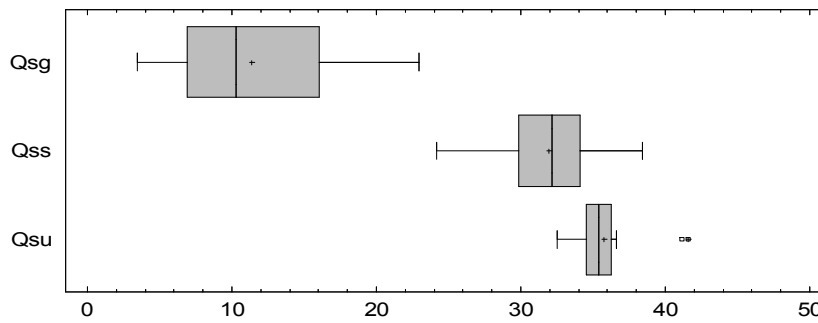


Рис. 2. Усатые ящики Тьюки показателя Qs для различных уровней АД

Для изучения статистической связи между одинаковыми показателями разных уровней АД вычислены коэффициенты их взаимной корреляции Спирмена. Графы связей между этими показателями приведены на рис. 3, 4. На этих рисунках принято следующее правило записи результатов. Прямым шрифтом приведено численное значение коэффициента корреляции, курсивом в скобках значение величины  $P_v$ . Если величина  $P_v < 0,05$ , то коэффициент корреляции считают отличным от нуля на уровне доверительной вероятности 0,95.

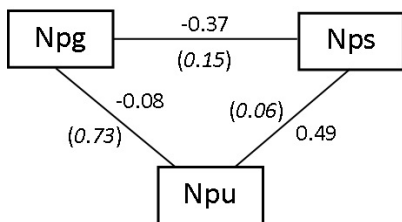


Рис. 3. Граф взаимнокорреляционных связей для показателя Np

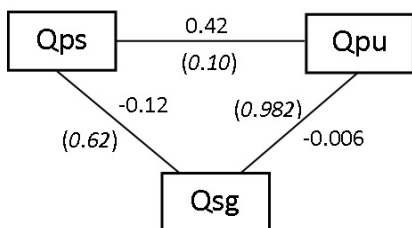


Рис. 4. Граф взаимнокорреляционных связей для показателя Qs

Из приведенных рисунков следует, что для показателей Np и Qs статистически значимая связь между разными уровнями АД практически отсутствует. Для изучения статистической связи показате-

телей Np и Qs для одинаковых уровней АД использован коэффициент парной ранговой корреляции Спирмена. Результаты приведены в табл. 7. Из этих данных видно, что показатели госпитализации, характеризующие количественную и территориальную составляющую изучаемого процесса, связаны между собой, и следует рассматривать их совместное действие на процесс заболеваемости.

Обработка информации наших исходных данных содержит методические погрешности, обусловленные тем, что действующей система кодирования диагноза «язвенная болезнь», к сожалению, не отображает многие особенности течения заболевания, на что указано в работе [14].

### Выводы

1. Для различных уровней административно-территориального деления Украины показано, что количество ежегодных госпитализаций больных с язвенными кровотечениями для Украины не может быть описано нормальным законом распределения. Это установлено после использования критериев Колмогорова-Смирнова, Крамера-Мизеса, Андерсона-Дарлинга, Шапиро-Уилка, Шапиро-Френсиса.

2. Используя критерий Краскелла-Уоллеса и критерий Муда для медиан установлено, что медианы изученных рядов распределений существенно различаются между собой.

3. Используя коэффициент ранговой корреляции Спирмена, установлено, что статистически значимая связь между количеством госпитализаций для разных уровней административно-территориального деления связь между количеством госпитализаций для разных уровней административно-территориального деления

Оценка коэффициент ранговой корреляции Спирмена между показателями Np и Qs для равных уровней административно-территориального деления

| Сравниваемые пары | Величина коэффициента корреляции | Величина Pv | Вывод |
|-------------------|----------------------------------|-------------|-------|
| Npu↔Qsu           | 0,929                            | 0,0003      | Да    |
| Nps↔Qss           | 0,859                            | 0,0009      | Да    |
| Npg↔Qsg           | 0,988                            | 0,0001      | Да    |

отсутствует. Это верно как для показателя количества госпитализаций на 10 000 чел., так и для показателя количества госпитализаций на 1 000 кв. км.

4. Показано, что население местностей различных уровней АД имеет разную склонность к желудочно-кишечным кровотечениям, что обусловлено территориальной принадлежностью госпитализованного больного.

### Список литературы

1. Chiu P.W. Acute nonvariceal upper gastrointestinal bleeding [Текст] / P.W. Chiu, J.J. Sung // *Curr Opin Gastroenterol.* – 2010 Sep; 26(5): 425-8.
2. Lauret E. Epidemiological Clinical Features and Evolution of Gastroduodenal Ulcer Bleeding in a Tertiary Care Hospital in Spain, during the Last Seven Years [Текст] / E. Lauret, J. Herrero, L. Blanco et al. // *Gastroenterol Res Pract.* – 2013; 2013: 584540.
3. Lanás A. The changing face of hospitalisation due to gastrointestinal bleeding and perforation [Текст] / A. Lanás, L.A. García-Rodríguez, M. Polo-Tomás et al. // *Aliment Pharmacol Ther.* – 2011 Mar; 33(5): 585-91.
4. Irwin J. Incidence of upper gastrointestinal haemorrhage in Maori and New Zealand European ethnic groups, 2001-2010 [Текст] / J. Irwin, R. Ferguson, F. Weilert, A. Smith // *Intern Med J.* – 2014 Aug; 44(8): 735-41.
5. Ljubičić N. North vs south differences in acute peptic ulcer hemorrhage in Croatia: hospitalization incidence trends, clinical features, and 30-day case fatality [Текст] / N. Ljubičić, T. Pačić, I. Budimir et al. // *Croat Med J.* – 2014 Dec; 55(6): 647-54.
6. Yamanaka K. Hemorrhagic gastric and duodenal ulcers after the Great East Japan Earthquake Disaster [Текст] / K. Yamanaka, H. Miyatani, Y. Yoshida et al. // *World J Gastroenterol.* – 2013. – Nov 14; 19(42): 7426-32.
7. Paspatis G.A. An epidemiological study of acute upper gastrointestinal bleeding in Crete, Greece [Текст] / G.A. Paspatis, E. Matrella, A. Kapsoritakis et al. // *Eur J Gastroenterol Hepatol.* – 2000 Nov; 12(11): 1215-20.
8. Cavallaro L.G. Time trends and outcome of gastrointestinal bleeding in the Veneto region: a retrospective population based study from 2001 to 2010 / L.G. Cavallaro, F. Monica et al. // *Dig Liver Dis.* – 2014 Apr; 46(4): 313-7.
9. Svoboda P. [The benefits of centralization of care for patients with acute upper gastrointestinal bleeding] [Текст] / P. Svoboda, M. Konečný, V. Hrabovský et al. // *Vnitř Lek.* 2012 Mar; 58(3): 191-5. [Article in Czech].
10. Теорія статистики [Текст] / П.Г. Ваукив, П.І. Пастер, В.П. Сторожук, Є.І. Кач. – К.: Либідь, 2001. – 320 с.
11. Шеффе Г. Дисперсионный анализ [Текст] / Г. Шеффе. – М.: Наука, 1980. – 512 с.
12. Холлендер М. Непараметрические методы статистики [Текст] / М. Холлендер, Д.А. Вульф. – М.: Финансы и статистика, 1983. – 518 с.
13. Гайдышев И.П. Моделирование стохастических и детерминированных систем: Руководство пользователя программы AtteStat [Текст] / И.П. Гайдышев. – Курган, БИ.: – 2013. – 496 с.
14. Langner I. Regional and temporal variations in coding of hospital diagnoses referring to upper gastrointestinal and oesophageal bleeding in Germany [Текст] / I. Langner, R. Mikolajczyk, E. Garbe // *BMC Health Serv Res.* – 2011 Aug 17; 11: 193.

Поступила в редколлегию 8.12.2015

Рецензент: д-р техн. наук, проф. Б.М. Конорев, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков.

### ОЦІНКА СТАТИСТИЧНОГО ЗВ'ЯЗКУ МІЖ ПОКАЗНИКАМИ ЧАСТОТИ ГОСПІТАЛІЗАЦІЇ І ТЕРИТОРІАЛЬНИМ РОЗПОВСЮДЖЕННЯМ ВИРАЗКОВОЇ ХВОРОБИ

В.Ю. Дубницький, І.Д. Дужий, С.В. Харченко, О.І. Ходирев

Визначено наявність статистичного зв'язку між показниками частоти госпіталізації і територіального розповсюдження виразкової хвороби. Джерелом даних для аналізу були відомості державної статистики про госпіталізацію хворих з діагнозом виразкової хвороби, ускладненої кровотечею, для різних рівнів адміністративно-територіального поділу України. Рівень статистичного зв'язку оцінювали методами непараметричної статистики. У роботі використані критерії Колмогорова-Смирнова, Крамера-Мізеса, Андерсона-Дарлінга, Шаніро-Уїлка, Шаніро-Френсіса, Краскелла-Уоллеса, критерії для медіан Муда, коефіцієнт рангової кореляції Спірмена.

**Ключові слова:** виразкова хвороба, непараметрична статистика, критерії Колмогорова-Смирнова, Крамера-Мізеса, Андерсона-Дарлінга, Шаніро-Уїлка, Шаніро-Френсіса, Краскелла-Уоллеса, критерії для медіан Муда, коефіцієнт рангової кореляції Спірмена.

### ASSESSMENT OF STATISTICAL RELATIONSHIP OF ADMISSION INCIDENCE AND TERRITORIAL PREVALENCE FOR ULCER DISEASE

V.Yu. Dubnitskiy, I.D. Duzhiy, S.V. Kharchenko, A.I. Khodyrev

The study identified a statistical relationship of admission incidence and territorial prevalence for ulcer disease. The source for analysis was the official statistical data for admissions of patients with ulcer disease complicated with bleeding on different administrative and territorial levels in Ukraine. The power of the statistical relationship was calculated with nonparametric statistical methods. The work used Kolmogorov-Smirnov test, Cramer-von Mises test, Anderson-Darling test, Shapiro-Wilks test, Shapiro-Francia test, Kruskal-Wallis test, Mood's median tests and Spearman rank correlations.

**Keywords:** ulcer disease, nonparametric statistics, Kolmogorov-Smirnov test, Cramer-von Mises test, Anderson-Darling test, Shapiro-Wilks test, Shapiro-Francia test, Kruskal-Wallis test, Mood's median tests, Spearman rank correlations.