

Xây dựng và khai thác BAĐT: CON ĐƯỜNG MỚI TRONG KHÁM CHỮA BỆNH VÀ NGHIÊN CỨU Y HỌC

GS HỒ TÚ BẢO

Viện Khoa học và Công nghệ Tiên tiến Nhật Bản
Viện John von Neumann, Đại học Quốc gia TP Hồ Chí Minh

Từ vài năm gần đây, bệnh án điện tử (BAĐT) được nói nhiều ở các nước phát triển như con đường dẫn đến sự thay đổi sâu sắc trong khám chữa bệnh và nghiên cứu y học. Câu chuyện BAĐT cũng đã bắt đầu ở nước ta với nỗ lực đặt những viên gạch đầu tiên cho con đường dài này...

BAĐT

BAĐT là gì?

Mỗi bệnh nhân trong thời gian nằm viện đều có một *bệnh án* ghi chép và lưu trữ dữ liệu và thông tin về quá trình khám chữa bệnh. Xưa nay, các bệnh án được ghi chép trên giấy, lưu giữ trong kho, và lấy ra dùng khi cần thiết.

Từ vài chục năm lại đây, các bệnh viện đều có *hệ thông tin bệnh viện* (HTTBV, HIS: hospital information system). Các HTTBV ở thời đầu của công nghệ thông tin (CNTT) thường khá đơn giản, được tạo ra và dùng để quản lý bệnh viện và người bệnh với những dữ liệu cơ bản, và luôn được cải tiến hoặc thay bởi các HTTBV có những chức năng mới theo sự tiến bộ rất nhanh của CNTT. Hiện nay, các HTTBV là những hệ thống thông tin phức tạp với ba nhóm chức năng mới được quan tâm nhiều:

- Kết nối 3 yếu tố: *viện phí, thuốc điều trị* và *bảo hiểm y tế* cho người bệnh. Một HTTBV tốt sẽ nâng cao được tính minh bạch và xác đáng trong việc tính toán ba loại chi phí trên, hỗ trợ việc quản lý bệnh viện, bảo hiểm y tế, giảm thất thoát...
- Sử dụng hiệu quả *dữ liệu cận lâm sàng* (dữ liệu xét nghiệm, siêu âm, nội soi, X-quang, CT scan, MRI...). Máy móc y tế ngày càng được dùng phổ biến đang tạo ra rất nhiều dữ liệu dạng *ảnh* và

số, và các phương pháp phân tích dữ liệu với máy tính có thể hỗ trợ rất nhiều cho cách phân tích định tính hiện nay ở bệnh viện.

- *Xây dựng và khai thác* BAĐT (Electronic Medical Record - EMR) là nhóm chức năng mới đang được theo đuổi ở nhiều quốc gia trong những năm gần đây. BAĐT chính là bệnh án của người bệnh ở dạng điện tử tạo ra trên các HTTBV. BAĐT gồm dữ liệu cơ bản cho quản lý, dữ liệu cận lâm sàng và dữ liệu lâm sàng của người bệnh trong một lần nằm viện. Trong BAĐT, dữ liệu lâm sàng là các *văn bản lâm sàng* (clinical text) do bác sĩ và y tá ghi chép hàng ngày về việc khám chữa và tình trạng người bệnh. Chính phần nội dung này tạo ra giá trị và ý nghĩa đặc biệt của BAĐT trong chăm sóc sức khỏe và nghiên cứu y học.

Văn bản lâm sàng trong BAĐT chủ yếu gồm 3 loại: (a) *Phiếu điều trị* (doctor daily notes), ghi chép các chẩn đoán, nhận định và y lệnh hàng ngày của bác sĩ về bệnh nhân; (b) *Phiếu chăm sóc* (nurse narratives), là ghi chép mô tả thường xuyên trong ngày của y tá về tình trạng bệnh nhân khi chăm sóc và thực hiện y lệnh của bác sĩ; (c) *Hồ sơ xuất viện* (discharge summary), tổng hợp các dữ liệu và thông tin cơ bản của bệnh nhân trong toàn bộ lần điều trị.

Bên cạnh khái niệm “BAĐT”, khái niệm “hồ sơ

sức khoẻ điện tử” (electronic health record - EHR) cũng rất phổ biến. Đại thể, EMR là dạng điện tử của thông tin khám chữa bệnh của một người bệnh trong một lần nằm viện, còn EHR là dạng điện tử của thông tin cơ bản của một người bệnh trong nhiều lần khám chữa bệnh, chia sẻ giữa nhiều tổ chức.

Có thể làm gì với BAĐT?

Ngoài việc có thể dễ dàng khai thác BAĐT trên máy tính để làm các việc như với bệnh án trên giấy, BAĐT còn có thể dùng làm nhiều việc khác, là nền tảng cho một sự thay đổi sâu sắc trong ngành y. Sự khác biệt chủ yếu ở chỗ con người chỉ có thể xem từng bệnh án trên giấy, nhưng máy tính có thể cùng một lúc “xem” rất nhiều BAĐT để phân tích, so sánh chúng, tìm ra những hiểu biết mới của y học. Có thể thấy rõ điều này nếu nhìn lại quá trình hình thành tri thức y học của con người. Đây là việc mọi hiểu biết về bệnh tật xưa nay đều do các thầy thuốc phát hiện khi phân tích, đúc rút từ những gì thấy được trong quá trình khám và chữa bệnh cho rất nhiều người. Các ghi chép lâm sàng trong BAĐT chính là “những gì thấy được” của rất nhiều thầy thuốc trong lúc khám chữa bệnh, và chúng nhiều hơn rất nhiều lần so với những gì mỗi thầy thuốc đã thấy trong suốt đời mình. Khi hàng vạn, hàng triệu và nhiều hơn nữa các BAĐT được tích lũy trong nhiều năm, được kết nối với nhau và phân tích trên máy tính bằng những phương pháp thích hợp và hiệu quả, rất nhiều tri thức y học mới và quý báu sẽ được phát hiện, cũng như những tri thức y học đã biết sẽ được hiểu rõ hơn.

Trong những năm gần đây, việc khám chữa bệnh dựa nhiều hơn vào sự kết hợp các tri thức lâm sàng và cận lâm sàng. Khám chữa lâm sàng dựa trên tri thức và kinh nghiệm tích lũy của từng bác sĩ sau nhiều năm học tập và hành nghề, còn khám chữa cận lâm sàng dựa vào phân tích hình ảnh và dữ liệu số tạo ra từ máy móc y tế. Các tri thức lâm sàng của thầy thuốc, nhiều khi là trực cảm không thể diễn giải, ẩn chứa trong các mô tả khám chữa bệnh (tacit knowledge), và do đó BAĐT có thể là phương thức duy nhất cho phép tự động hoá việc kết hợp hai nguồn tri thức lâm sàng và cận lâm sàng trong khám chữa bệnh. Đây là một giá trị rất lớn của BAĐT trong tương lai.

Xây dựng và khai thác BAĐT

Hai vấn đề chính của BAĐT là xây dựng và khai thác chúng.

Xây dựng BAĐT

Năm 2009, sau khi trở thành Tổng thống và trong kế hoạch hồi sinh nền kinh tế, Tổng thống Mỹ Barack Obama đã yêu cầu nước Mỹ hiện đại hoá việc chăm sóc sức khoẻ dựa vào việc *chuẩn hoá và số hoá mọi bệnh án trong bệnh viện*, và chủ trương “trong vòng 5 năm, mọi bệnh án của nước Mỹ phải được số hoá”.

Ở Nhật Bản, Chính phủ có chính sách ưu tiên kinh phí để xây dựng BAĐT tại các bệnh viện cỡ lớn và vừa. Tính đến năm 2011, 51,5% trong số 822 bệnh viện cỡ lớn (hơn 400 giường bệnh) ở Nhật Bản đã xây dựng BAĐT; 27,3% trong số 1832 bệnh viện cỡ vừa (200-399 giường bệnh) đã có BAĐT; và 13,5% trong số 5951 bệnh viện cỡ nhỏ (dưới 200 giường bệnh) có BAĐT.

Việc xây dựng BAĐT trước hết dựa trên nhận thức của các cấp quản lý ngành y về tính tất yếu và vai trò của BAĐT trong việc hiện đại hoá nền y tế, chủ trương của các bệnh viện về xây dựng HTTBV hiện đại và BAĐT, và nhất là sự nỗ lực của đội ngũ y tá, bác sĩ trong việc thay đổi thói quen ghi chép hàng ngày về người bệnh.

Khai thác BAĐT

Đương nhiên, sau khi tạo ra được các BAĐT, câu hỏi sẽ là sử dụng và khai thác chúng thế nào?

Trước hết, BAĐT có thể được dùng một cách tiện lợi đối với các chức năng của các bệnh án thông thường: khi một bệnh nhân quay lại bệnh viện, các bác sĩ dễ dàng tham khảo BAĐT đã có của bệnh nhân này từ những lần nằm viện trước, và nhiều bác sĩ có thể cùng hội chẩn từ xa dựa trên BAĐT. Nhưng hơn thế, sự vô giá của BAĐT chính ở việc chúng cho ta những cơ hội mới để phân tích, so sánh đồng thời rất nhiều BAĐT để hiểu hơn về bệnh tật, về chữa trị, về tác dụng của thuốc, về quản lý khám chữa bệnh, thuốc men...

Nền tảng của khai thác BAĐT là việc *khai thác các văn bản lâm sàng*. Các ghi chép của bác sĩ và y tá trong các BAĐT, xét về phương diện ngôn ngữ, là một loại văn bản đặc biệt. Chúng thường ngắn, mang tính trần thuật, thường không tuân theo các quy ước ngữ pháp và ngôn ngữ một cách chặt chẽ, và đặc biệt hơn cả là chúng chứa rất nhiều thuật ngữ y học, tải theo nhiều tri thức của ngành y.

Xử lý ngôn ngữ tự nhiên (natural language

processing) là một lĩnh vực của CNTT, nhằm làm cho máy tính có thể tự động phân tích và hiểu được ngôn ngữ của con người. Để làm việc này, máy tính phải tuân tự thực hiện 4 công đoạn phân tích về hình thái (morphology), cú pháp (syntax), ngữ nghĩa (semantics) và diễn ngôn (discourse), với nhiều vấn đề phải giải quyết trong mỗi công đoạn. *Làm cho máy tính “hiểu” được văn bản lâm sàng là bước quyết định để tìm ra các tri thức y học mới.*

Việc phân tích văn bản lâm sàng ở hai công đoạn đầu của xử lý ngôn ngữ tự nhiên phụ thuộc vào ngôn ngữ của BÀĐT (tiếng Việt, tiếng Anh, tiếng Nhật...). Do vậy hai công đoạn này phải làm riêng cho từng thứ tiếng, và có thể thấy điều này qua hoạt động của các cộng đồng đi trước đối với BÀĐT tiếng Anh và tiếng Nhật.

Tại Mỹ, từ năm 2006, I2B2 - Trung tâm Quốc gia về Tính toán Y - Sinh (<https://www.i2b2.org>) - hàng năm tổ chức các cuộc thi tìm các phương pháp phân tích văn bản lâm sàng của BÀĐT tiếng Anh, gọi là các Thách thức (Challenges). Mỗi Thách thức đưa ra một số BÀĐT và một vấn đề phân tích cụ thể. Mỗi năm thường xuyên có trên dưới 100 nhóm nghiên cứu từ khắp nơi trên thế giới đăng ký tham gia đề xuất giải pháp và gửi kết quả phân tích, và những giải pháp tốt được chọn lọc để công bố ở một hội thảo quốc tế, cũng như phổ biến rộng rãi. Một vài Thách thức những năm qua là:

- Thách thức 2008: làm sao dùng BÀĐT để phân loại các nhóm bệnh nhân béo phì và các bệnh phát sinh do béo phì.
- Thách thức 2009: trích rút thông tin từ BÀĐT về việc dùng thuốc, liều lượng, hiệu quả, quản lý thuốc...
- Thách thức 2012: phát hiện từ BÀĐT các sự kiện về bệnh và điều trị gắn với thời gian và quan hệ phụ thuộc thời gian giữa chúng.
- Thách thức 2014: phát hiện, lọc bỏ các thông tin cá nhân từ BÀĐT và các yếu tố rủi ro của bệnh tim.

Tại Nhật Bản, Viện Tin học Quốc gia (NII) gần đây cũng chủ trì các hội thảo MedNLP về xử lý ngôn ngữ y học nhằm tạo ra nền tảng cho nghiên cứu và khai thác BÀĐT tiếng Nhật, với bài toán chính là dựa vào các BÀĐT để mở rộng các công cụ xử lý văn bản tiếng Nhật Bản đã có cho xử lý ngôn ngữ y học:

- MedNLP1 (2013): từ BÀĐT, phát hiện và lọc bỏ các thông tin riêng tư của người bệnh, như tuổi tác, giới tính, quê quán, tên, thời gian...
- MedNLP2 (2014): từ BÀĐT, phân tích tự động mô tả bệnh của bệnh nhân và nhận định của bác sĩ, chuẩn hoá chúng theo mã bệnh ICD-10.

Xây dựng BÀĐT ở Việt Nam

Có thể chia ra hai loại BÀĐT: *loại cơ bản* (có thông tin quản lý cơ bản) và *loại hoàn chỉnh* (có thông tin quản lý, văn bản lâm sàng và các thông tin khám chữa bệnh khác). Theo quan sát của chúng tôi, hầu hết các bệnh viện của ta nếu có BÀĐT cũng mới là các BÀĐT cơ bản. Tuy nhiên đã có những bước đi tiên phong. Có thể nói, Bệnh viện đa khoa Vân Đồn của tỉnh Quảng Ninh là cơ quan y tế đầu tiên ở ta có một HTTBV hiện đại với các BÀĐT hoàn chỉnh. Từ đầu năm 2013 đến nay, tất cả bệnh án của Bệnh viện đa khoa Vân Đồn đều đã được tạo ra trên máy tính, với ghi chép lâm sàng chi tiết hàng ngày của các bác sĩ và y tá về từng người bệnh. Bệnh viện đa khoa Vân Đồn và một số bệnh viện tại Quảng Ninh đã xây dựng HTTBV hiện đại với hệ MEDI SOLUTIONS, là phần mềm quản lý tổng thể bệnh viện, xây dựng bởi Công ty Công nghệ Phần mềm Hoa Sen (<http://hoasensoft.com/>), có trụ sở tại TP Hồ Chí Minh.



Xây dựng BÀĐT tại một bệnh viện ở Quảng Ninh

Hệ MEDI SOLUTIONS có độ chuẩn hoá cao do được xây dựng dựa theo các chuẩn quốc tế như công nghệ điện toán đám mây, giao thức HL7 (Health Level-7, tập các chuẩn quốc tế để truyền dữ liệu giữa các HTTBV), các hệ thuật ngữ tiêu chuẩn về bệnh ICD (International Classification of Diseases), từ

The screenshot displays the 'To điều trị' (Treatment) window in the MEDI SOLUTIONS software. It contains patient details such as name, date of birth, gender, and address. Below this, there's a section for 'Chẩn đoán' (Diagnosis) and 'Y lệnh' (Prescriptions). The 'Y lệnh' section lists several prescriptions with columns for 'STT' (Serial Number), 'Y lệnh' (Prescription), and 'Ghi chú' (Remarks). The list includes items like 'Atropin sulfate 0.25mg 2 ống', 'Seduxen 5mg (Hưng) 2 viên', and 'TĐ: M, HA, NT, T con đau bụng 3h/lần'. At the bottom, there are buttons for 'Thêm' (Add), 'Sửa' (Edit), 'Xóa' (Delete), and 'Mật khẩu' (Password).

Văn bản lâm sàng của bác sĩ trên hệ MEDI SOLUTIONS

diễn thuốc của WHO... hay hệ quản trị cơ sở dữ liệu Oracle. Ngoài ra, là một hệ thống để giải quyết những vấn đề phức tạp, MEDI SOLUTIONS có cấu trúc và thiết kế mở và hợp lý, được viết theo các quy trình công nghiệp nên có tính ổn định cao, có giao diện thân thiện với người dùng, giúp y tá, bác sĩ, nhà quản lý... sử dụng dễ dàng.

Hệ MEDI SOLUTIONS đang được triển khai sử dụng tại Bệnh viện phụ sản - nhi Đà Nẵng (9.2014), Bệnh viện đa khoa Cẩm Phả (3.2014), Bệnh viện sản - nhi Hưng Yên (4.2014), Bệnh viện y - dược cổ truyền tỉnh Quảng Ninh (7.2014)...

Khai thác BADT tiếng Việt

Cùng lúc các BADT hoàn chỉnh tiếng Việt được xây dựng ở một số bệnh viện, từ giữa năm 2013, tập thể nghiên cứu về "Học máy và ứng dụng" của Viện John von Neumann (JVN) thuộc Đại học Quốc gia TP Hồ Chí Minh - kết hợp với Viện Nghiên cứu Cao cấp về Toán (VIASM) tại Hà Nội, và các cộng tác viên từ một số bệnh viện, viện nghiên cứu và đại học - đã xây dựng một kế hoạch nghiên cứu nhằm phát triển các phương pháp và phần mềm công cụ cho việc khai thác BADT tiếng Việt.

Việc xây dựng kế hoạch nghiên cứu này không chỉ bắt nguồn từ nhận thức về vai trò thiết yếu của BADT trong sự phát triển của y tế và y học của đất nước, mà còn từ việc đánh giá các điều kiện để có thể thực hiện thành công việc khai thác BADT ở nước ta:

- Có nhận thức và tầm nhìn ở nhiều cấp quản lý về tính tất yếu và tầm quan trọng của BADT, như

từ năm 2011 Bộ Y tế đã triển khai dự án nghiên cứu khả thi về "BADT và quản lý hệ thống khám chữa bệnh" (dù BADT mới ở mức cơ bản), hay Bệnh viện đa khoa Vân Đồn đã tiên phong trong xây dựng BADT hoàn chỉnh dù còn phải chờ các phương pháp và công cụ khai thác được phát triển.

- Có sự quan tâm, đồng ý hoặc mong muốn tham gia xây dựng và khai thác BADT của nhiều bác sĩ và dược sĩ giỏi trên cả nước, của Bệnh viện đa khoa Vân Đồn, Bệnh viện phụ sản - nhi Đà Nẵng, Bệnh viện Thống Nhất, Bệnh viện bệnh nhiệt đới TP Hồ Chí Minh...

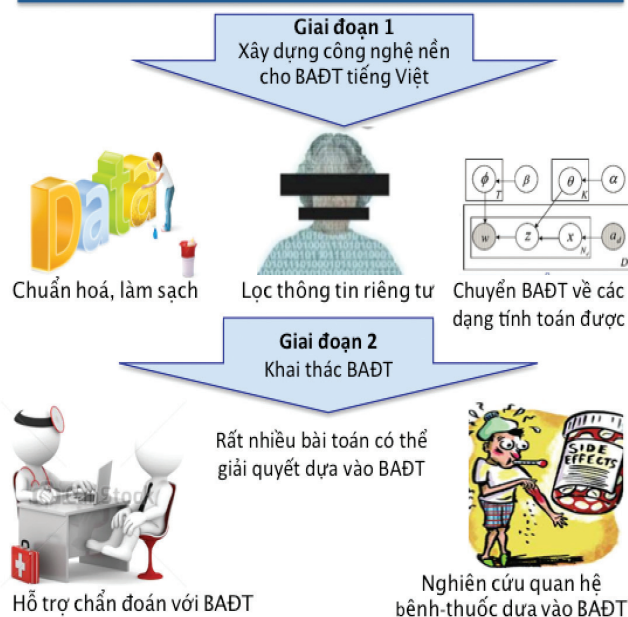
- Đã bước đầu có các cơ sở dữ liệu BADT hoàn chỉnh đủ lớn để tiến hành nghiên cứu, chuẩn bị cho việc khai thác trong một vài năm tới khi nhiều bệnh viện có BADT hoàn chỉnh.

- Có lực lượng nghiên cứu về CNTT và toán học cần cho 3 nội dung của việc khai thác BADT tiếng Việt: (1) Học máy và phân tích dữ liệu, (2) Phân tích tự động tiếng Việt, và (3) Tin y - sinh học. Đây là kết quả của việc đào tạo và nghiên cứu trong nhiều năm, như những đề tài về học máy và tin sinh học được Quỹ NAFOSTED tài trợ, như đề tài cấp nhà nước KC.01/06-10 về xây dựng công cụ và tài nguyên cho xử lý tiếng nói và văn bản tiếng Việt (2007-2010) với hợp tác của 9 trường đại học và viện nghiên cứu trên cả nước, như Trường hệ quốc tế về học máy do Viện Toán VIASM tổ chức năm 2012, hay Hội nghị châu Á về học máy ACML tại Nha Trang vào tháng 11.2014 và Hội nghị châu Á - Thái Bình Dương về khai phá dữ liệu PAKDD tại TP Hồ Chí Minh sẽ tổ chức vào tháng 5.2015.

Kế hoạch nghiên cứu dài hạn nêu trên hướng đến việc xây dựng các phương pháp và phần mềm công cụ cơ bản và thiết yếu cho việc khai thác BADT tiếng Việt trong khám chữa bệnh và nghiên cứu y học tại Việt Nam (gọi tắt là kế hoạch BADT). Kế hoạch BADT gồm 2 giai đoạn. Giai đoạn 1 nhằm xác định và xây dựng công nghệ nền để khai thác BADT tiếng Việt và ứng dụng thử nghiệm. Giai đoạn 2 nhằm sử dụng các công nghệ nền vào khai thác BADT để giải quyết một số bài toán quan trọng trong nghiên cứu y học và chăm sóc sức khỏe ở Việt Nam. Mỗi giai đoạn dự kiến trong quãng thời gian 3-5 năm với những nội dung cụ thể đã xác định. Kế hoạch BADT đã được Đại học Quốc gia TP Hồ Chí Minh tài trợ bước đầu, nhằm tạo ra những kết quả nền tảng và từ đó tìm những nguồn kinh phí khác để thực hiện kế hoạch.



BỆNH ÁN ĐIỆN TỬ



2 vấn đề được chọn lọc để nghiên cứu thử nghiệm gồm hệ thống hỗ trợ chẩn đoán bệnh và nghiên cứu mối quan hệ bệnh - thuốc dựa vào BAĐT.

Hỗ trợ chẩn đoán bệnh bằng máy tính là một bài toán truyền thống của tin - y học (Medical Informatics) từ nhiều chục năm. Các hệ chuyên gia (expert systems) với kỹ thuật suy diễn (deduction) dựa vào tri thức của con người trong y học được kỳ vọng rất nhiều nhưng đã không thành công, chủ yếu vì không biểu diễn được tri thức y học phức tạp của con người dưới dạng máy có thể dùng được. Theo một cách ngược lại, khi dùng các kỹ thuật học máy theo quy nạp (induction) với BAĐT, máy tính hoàn toàn có thể hỗ trợ hiệu quả cho con người trong công việc phức tạp này. Hình dung khi ta có rất nhiều BAĐT và cần chẩn đoán bệnh cho một bệnh nhân mới, máy tính sẽ tìm ra một vài BAĐT của những người đã được khám chữa có triệu chứng giống nhất so với người bệnh. Những kết luận chẩn đoán (sai hoặc đúng), những thành bại trong phương pháp

điều trị từ các BAĐT nêu trên sẽ giúp rất nhiều cho việc khám chữa cho bệnh nhân mới.

BAĐT có triển vọng to lớn trong nghiên cứu quan hệ bệnh - thuốc (disease-drug relationship). BAĐT cho phép phân tích hiệu quả sử dụng thuốc để đánh giá phương pháp điều trị, mức độ tác dụng của thuốc trên các nhóm người bệnh, tương tác giữa các loại thuốc, phát hiện các hiệu ứng phụ khi dùng đồng thời nhiều loại thuốc, phát hiện các tác dụng mới từ các loại thuốc sẵn có... Với một thử nghiệm trong năm 2014, nhóm nghiên cứu đã xây dựng một kỹ thuật phân tích BAĐT để tự động tạo các bảng tần suất về tác dụng (hiệu quả, hiệu quả thấp, không hiệu quả...) của các nhóm thuốc dùng điều trị bệnh (những chỉ định thuốc bất kỳ) lên các nhóm bệnh nhân (như theo giới tính, tuổi, mức độ bệnh nặng nhẹ...). Thông thường, để tạo ra một bảng tần suất này - một nhiệm vụ cơ bản trong nghiên cứu y dược - cần tới vài tháng với kinh phí lớn. Tuy nhiên, với BAĐT ta có thể tạo ra tức khắc bảng các tần suất cho bất kỳ nhóm thuốc và nhóm bệnh nhân nào.

Rất nhiều vấn đề khác của y học hiện đại có thể nghiên cứu với dữ liệu lâm sàng từ BAĐT, như phân tích thống kê được hiệu quả của các phương pháp can thiệp (intervention) trên các nhóm bệnh nhân, mối liên hệ của các phương pháp khám chữa bệnh hiện đại dựa trên các yếu tố di truyền, trao đổi chất... Nói cách khác, nếu trước kia các thống kê y học chỉ làm được trên các con số thì từ nay trở đi, thống kê y học có thể làm trên các văn bản lâm sàng của BAĐT, là nơi chứa những thông tin quan trọng hơn cả về sức khỏe và bệnh tật của con người.

Kết luận

BAĐT đang trở thành một phần tất yếu của ngành y. Nó có thể thay đổi hiệu quả việc khám chữa bệnh và nghiên cứu y học, nếu ta xây dựng và khai thác chúng. Nỗ lực và kết quả ban đầu của những tập thể khoa học và những chính sách đúng đắn cho phép chúng ta hy vọng các BAĐT tiếng Việt sẽ góp phần vào sự phát triển của ngành y nước nhà, cũng như cho thấy việc nghiên cứu khoa học và công nghệ cao rất cần thiết cho những vấn đề thiết thực của cuộc sống.