Tuberculosis in the 21st century: an emerging pandemic?

Gunnar Bjune

Institute for General Practice and Community Medicine, University of Oslo, P.O. Box 1130 Blindern, NO-0317 Oslo, Norway

Telephone: +47 22850640    telefax: +47 22850672    e-mail: g.a.bjune@medisin.uio.no

ABSTRACT

Since the mid-1980ies the world has witnessed a dramatic increase in tuberculosis. Our knowledge of tuberculosis epidemiology was established during the previous epidemic that, at least in Europe, took two hundred years from rise to fall. It is difficult to understand the new epidemic if we assume that the genetics of the host-parasite relationship is unchanged from the previous epidemic.

The paper discusses how both host and parasite genetics may have changed. Molecular epidemiology done in Archangel, Russia, where most of the classical reasons for increase in tuberculosis were absent, indicated that strains of *Mycobacterium tuberculosis* with changed biological properties could be responsible. Strains belonging to the so-called Beijing family were strongly associated with primary drug resistance and caused clusters ten times as big as “traditional” strains.

The paper describes how the research consortium takes these observations further to explore the importance of “new” strains of tubercle bacilli in the on-going pandemic. What are the changes in biological properties and what genetic changes do they reflect? The previous epidemic may have changed the genetic susceptibility of the human host by selection, but what has constituted a selective pressure for the bacillary population? The two most dramatic changes in the environment of *M. tuberculosis* are mass introduction of chemotherapy and BCG vaccination. We explore further the nature of drug resistance in these strains and the possibility that BCG may fail to protect against bacilli of the new pandemic.

Lastly the paper points at some action that can be taken instantly and that may have a major impact on transmission, even before the questions mentioned above are answered. The clue here is to shorten the time of transmission by a rapid test to secure early diagnosis and treatment. What is needed, however, is not a diagnostic test for tuberculosis, but a simple screening test with high sensitivity that could tell us whom among the numerous people who have a chronic cough and systemic symptoms who are the true “tuberculosis suspects” eligible for rapid examination of sputum by smear microscopy.

THE NEW PANDEMIC

Since the mid-1980ies the world has witnessed a dramatic increase in tuberculosis (TB)\(^1\)^\(^-\)\(^4\). Based on the assumption that the genetic relationship between host and parasite was unchanged, the causes for this comeback of the White Plague have been listed as an underlying increase in:

1. risk of reactivation of latent TB by increase in life expectancy\(^5\)
2. host susceptibility due to HIV co-infection\(^6\)
3. risk of exposure through urbanization, migration and destitution\(^7\)
4. the size of the infectious reservoir through sub-standard drug treatment\(^8\)

This may all be true, but could it really explain an epidemic that seem to have started in so many sites at about the same time?

THE PREVIOUS EPIDEMIC

It is important to notice that even though there seems to be a solid agreement that there is an epidemic increase in global TB, the data are not rock solid. From many countries with fairly new or revitalized TB control programs, the reported increases in new cases could be an operational effect of more efficient or more decentralized TB programs. Also the introduction of an international target for case detection rate has been a strong drive in many countries to do cosmetics on the figures to please superiors. But let us, for this discussion, assume that there is a global epidemic trend for TB.

Our knowledge of TB epidemiology was established during the previous epidemic that, at least in Europe, took two hundred years from rise to fall. Even the dynamics of that epidemic is only partially understood. The fall could surely not be due to development of immunity in individuals through exposure to the pathogen like in most other infectious diseases. Neither can it be fully explained by a slow spread from one focus in central places to more and more remote parts of Europe. Each and every community experienced the effects of the epidemic in at least fifty to hundred years. Thus changes in the genetics of the parasite or the host, or both could have had a significant impact.

The rise could have been fuelled by early industrialization and urbanization\(1\). This would certainly facilitate transmission. But TB took high death tolls also where there was no industry and no cities. The fall
could have been assisted by improvements in living conditions, particularly regarding nutrition and housing. But it is hard to find evidence that this was the case everywhere. We remain with a suspicion that there must have been genetic changes in the host parasite relationship.

**GENETICS IN THE COURSE OF AN EPIDEMIC**

During an epidemic natural selection of mutant pathogens will favour low virulence to keep the host alive and contributing to spread as long as possible. The end result of this genetic adaptation in the previous epidemic (and thus the only model we have for host-parasite balance in TB), was that *Mycobacterium tuberculosis* (*Mtb*) established latent infection in more than ninety percent of the infected individuals and achieved, mainly through reactivation, a dissemination that could go over long time and long distance. The reproductive rate was 1.0; the perfect host-parasite balance!

This balance was based on a certain life expectancy of the host. When life expectancy increased, as it has in most countries since the 1960ies, and the rate of reactivation per year stays constant, the reproductive rate and the incidence of TB increased.

With a reproductive rate of 1.0, we are also faced with an epidemiological situation where a large number of factors that contributes to reduced spread or higher host resistance would inevitably result in a long-term decline in TB incidence.

A slight reduction in annual rate of infection (ARI) due to better housing, or isolation of a certain percentage of infectious TB patients, or more varied nutrition reducing the risk of malnutrition induced immunodeficiency would result in a long term decline in TB. The decline throughout most of the last century can be seen as a fairly constant secular trend. The introduction of tuberculin skin testing (TST) making it possible to target social reforms in the fight against TB, or BCG vaccination, and even the introduction of effective chemotherapy are not reflected in significant changes in the inclination of the downward trend in TB incidence. Even more surprisingly, the trend was the same all over Europe and North America in spite of substantial differences between economic development and nutritional transitions. Only very remote regions like the Arctic were seeing the decline later, and small epidemics occurred in virgin populations like the enuits.

Was there a component of genetic change in the human host contributing to the European and North American decline in TB? About 50\% of patients with pulmonary TB will die within three years from the start of their disease. With annual incidence figures of 200-600/100,000 population, as may be representative for most of Europe, death by itself could hardly explain a significant weeding out of susceptibility genes in the population. However, in most societies TB was soon established as a stigmatizing disease. Not only were patients themselves shunned, which could reduce exposure of the healthy population, but even their families were not regarded suitable for matrimony. The effect of this reducing the transfer of potential susceptibility genes to the following generations has, to my knowledge not been mathematically modelled. That the impact would be much higher than from disease specific mortality is in any case obvious.

The pathogen’s genome is also subject to changes. The last ten years experience with restriction fragment length polymorphism (RFLP) has taught us that it takes about 1.5 years for an actively growing strain to change its genetic make-up sufficiently to produce a visible change in the RFLP pattern. When bacilli go into dormancy, we can assume that this evolutionary watch stops until the disease is reactivated. It explains why the isolates characterized by RFLP in low endemic settings today show so much diversity. The “time window” we look at in terms of period when patients picked up their infections is 30-50 years. The low degree of clustering in such populations does not necessarily prove that the cases we see are all caused by reactivation. It could equally well mean that many sources of infection go undetected in the community. Each visible change in the RFLP pattern may represent another 5-10 point mutations in the genome, based on the average length of the restriction fragments, all of them compatible with survival of the pathogen. It is perceivable that this genetic drift could have caused phenotypic changes of importance for host-parasite relationship.

**IS THE NEW TB PANDEMIC AN “EMERGING” EPIDEMIC OR A “COME-BACK”?**

From the discussion above we must conclude that it is quite likely that resistance due to host genetics changed during the previous epidemic. It is also clear that *Mtb* is able to change its genetic make-up and adapt to a changing host.

To look for evidence for such changes we have analysed the epidemic in Archangel oblast, Russia. The data from this area are particularly complete and HIV has not yet had any impact on the TB situation. The downward trend in TB shifted to an increase nearly in the year of the break-down of the Sovjet Union. Even though some very difficult years regarding food and housing followed, it is unlikely that this alone should explain such a swift change in trends. TB services had been accessible for all in the Sovjet period and remained so after the political change. Drug regimens had been a weak point contributing both to an increase in drug resistant TB and the reservoir of chronic transmitters in the community. Some more prisoners were released, but at the same time transfer of prisoners over long distances stopped. These changes were all gradual and can hardly explain the sudden shift in TB trends.

What we found, however, was that strains of *Mtb*
new to this area were introduced at the time of the shift and had increased parallel to the incidence in TB in general. We can provisionally call these strains “successful strains” and describe them as strains that rapidly expand their share among clinical isolates in a given population. In Archangel oblast they belonged to the Beijing family. Since this family has a well known history, emerging in China some 50 years back and now representing a high percentage of clinical isolates in South East Asia, South Africa and several locations in The Russian Federation\textsuperscript{15,16}, it is reasonable to ask whether the Beijing family represent “successful strains” fuelling an emerging pandemic.

**WHAT MAKES A NEW PHENOTYPE OF \textit{Mtb} A “SUCCESSFUL STRAIN”?**

The Beijing family has in most studies been associated with drug resistance\textsuperscript{17}. We know from Archangel that acquired resistance in the form of multi drug resistance (MDR) is not significantly more frequent among Beijing than among non-Beijing strains. However, in newly diagnosed patients Beijing was strongly associated with MDR.

The general rule is that a mutation that causes drug resistance will result in a certain loss of fitness of the microbe. A high frequency of primary resistance speaks against a significantly lower fitness in resistant strains versus fully susceptible ones. When fitness was measured as growth rate in vitro, a number of Beijing strains from Archangel did not show reduced fitness after they required resistance to rifampicin. Non-Beijing strains behaved as predicted and had reduced growth rate after acquisition of resistance\textsuperscript{17}. Presumably this makes some Beijing strains successful in spreading in an environment without antituberculous drugs as in newly infected and not yet diagnosed TB patients.

Another observation was that clusters caused by Beijing strains were common and most patients with such strains were also members of a cluster contrary to the situation among non-Beijing patients\textsuperscript{14}. A “cluster” is defined as a group of clinical isolates that are identical or show very little differences in their RFLP pattern. The Beijing clusters were also about ten times as big in average\textsuperscript{18}.

The most tempting hypothesis to extract from these observations would be that Beijing strains are more infective than non-Beijing strains. But there are two precautions to bear in mind. The Beijing strains have a very short evolutionary history as an identifiable entity. Diversity has thus had much shorter time to develop than for other \textit{Mtb} strains. Strains from different Beijing isolates could easily be misclassified as identical and erroneously included in a cluster without being part of a true epidemiological cluster. Secondly, non-Beijing strains reflect a time window of 30-50 years of acquisition of the strains that cause active disease in the relatively short period of the RFLP survey. If Beijing strains on the other side have lost the ability to establish stable latency and rather progress directly to primary disease in susceptible individuals, the time window is only 2-3 years. Their whole potential for spreading is picked up during a short RFLP survey. We find the last hypothesis most attractive on the background of the rapid expansion of Beijing strains in this area.

The list of phenotypic changes that could lead to the emergence of successful strains does not end here, but the two mentioned ones are born out of the observations from Archangel oblast.

**WHAT WOULD IT IMPLY THAT SUCCESSFUL STRAINS HAVE LOST THE ABILITY TO ESTABLISH STABLE LATENCY?**

Firstly, all TB caused by such strains would in principle be primary TB.

Another effect of \textit{Mtb} strains being unable to establish stable latency is that in a high HIV prevalence population, non-Beijing strains would have less competitive advantage as the majority of \textit{Mtb} infections would progress to primary TB anyhow.

In high endemic settings, with a high risk of exposure to \textit{Mtb} in the community, primary TB is mainly found among children and present as extra pulmonary or sputum smear negative pulmonary TB. This may create the impression that this is the clinical presentation of primary TB versus reactivated sputum smear positive TB in reactivation disease. In fact, this is more an illustration of the difference between the clinical picture of TB in children and adults and is a result of individual and age related differences in resistance against haematogenous spread of \textit{Mtb} infection in the body.

**BCG AND THE EMERGENCE OF SUCCESSFUL \textit{Mtb} STRAINS**

BCG seems to be an effective vaccine against primary TB in all settings where the efficacy is not blurred by concomitant immunization of unvaccinated controls by environmental mycobacteria. A preventive effect on infection, establishment of stable latency and reactivation of the disease is, however, highly debatable. Let us assume that the effect is nil. Consequently, BCG should provide a higher degree of protection against the spread of Beijing and other successful strains, if these strains only cause primary TB. The ratio between the ARI based expected incidence of cases with sputum smear positive TB and the observed number of new cases should thus be one or less in countries with a low contribution of Beijing strains among the isolates, and higher than one where Beijing strains are expanding, given the same BCG coverage and TB service quality. However, the opposite seems to be the case!

Does BCG protect against the new successful
strains of *Mtb*? This remains an important question for TB prevention. Preliminary studies from Vietnam indicated that BCG did not provide protection against TB caused by Beijing strains. This case-control study was, however, not designed to answer this question. Patients from a period with low and high BCG coverage were mixed. After correction for age of the TB affected individuals, the formerly significant association between harbouring a Beijing strain and being BCG vaccinated disappeared. An experimental study in mice with chronic progressive infection (analogous to primary TB in humans) revealed no protection against a few Beijing strains tested, while there was significant protection against H37Rv and some non-Beijing strains. One should still be cautious not to extrapolate this result from mice to men.

**SELECTION FACTORS FOR THE EVOLUTION OF NEW SUCCESSFUL Mtb PHENOTYPES**

If such strains really represent new phenotypes with a substantial spread, there must have been certain (new?) factors that have selected for them. We have already mentioned that BCG may turn out not to protect against, at least, some of the emerging strains. BCG is the oldest vaccine in current use and reaches today more than 80% of the world’s children. But as a selection factor it has only played a role in the last three to four decades. Age specific TB incidence in cohorts exposed to *Mtb* before and after BCG was introduced at mass scale will reflect what strains have become successful in the years following full vaccine coverage.

Another selective factor already mentioned is drug resistance. Again, such drugs have been available for more than 50 years. National TB control programmes providing access for the majority population in high endemic countries have only been in full operation for the last three decades. If poorly managed, mass provision of anti TB drugs will cause an increase in the infectious reservoir and induce acquired drug resistance. A competitive advantage is established if drug resistance can be acquired without the loss of fitness as was found in Archangel. But even there, with more than 10% of isolates from the new cases showing primary MDR, drug resistance can only partly explain the increasing trend in TB notification.

TB control programs with a high coverage could possibly function as a selective factor through another mechanism. People with different strains of *Mtb* are exposed to each other, and thus to foreign strains of *Mtb*, in health facilities where they come for diagnosis and treatment of their disease. TB patients, and even TB suspects, will share host related risk factors for reactivation and superinfection. If the newly acquired strain is more likely to progress to disease, or carries drug resistance genes, such a strain will slowly increase its share of the re-treatment market.

A similar mechanism would be effective in a population made more susceptible through acquired immune deficiency, whether HIV induced or related to malnutrition. The strains that take the shortest time to fully manifest themselves in new cases will soon dominate the market for new TB cases. HIV has been an effective amplifier for the TB epidemic in sub-Saharan Africa for the last 25 years, and severe malnutrition is increasingly common among refugees and the poorest of the poor.

**DOES A PREVIOUS Mtb INFECTION INFLUENCE SUSCEPTIBILITY TO SUPERINFECTION?**

There is ample indication for a protective role against disease following *Mtb* re-exposure by established latent infection with *Mtb*. Recently a surprisingly high fraction of newly diagnosed TB patients were, however, found to harbour two strains of *Mtb* at the same time. The methodology applied would be particularly sensitive to double infections where a Beijing strain is one of the two, so the frequency of this type of pairing could have a technical explanation, but it could also be due to an ability of Beijing strains to manifest themselves shortly after superinfection.

Of significance is also the finding that a high number of “new” TB patients in fact are found to have had previous chest X-ray findings compatible with fibrotic TB lesions and probably have had a slowly progressive infection for a long time. The finding of Beijing strains in isolates from such patients are not uncommon in spite of their lesions indicating an infection long before Beijing was introduced in the actual population.

In mice and men we found that slowly progressive tuberculous foci contained a subpopulation of heavily infected macrophages that evaded the host’s immunity by overexposure of FAS ligand. At this stage of infection in a host cell, T cytotoxic cells and natural killer cells should have induced apoptosis in the infected cell and thereby got the bacilli transferred to fresh macrophages that could have been activated to kill the infective agent. Instead such cells seem to be able to keep cytotoxic cells at bay and thus constitute a *locus minoris resistentia* in the individual host. Together people with slowly progressive *Mtb* infection will represent a *locus minoris resistentia* at population level. Strains currently spreading in the population will feed on these weak spots and, in competition, strains that are more likely to progress rapidly to disease will have an advantage and increasingly dominate the epidemic.

This implies that *Mtb* strains of the recent past, those we can suppose are dominating the huge population (one third of the world’s population) that are latently infected today, in fact constitute a selective pressure for new phenotypes competing for their share of the market.

In this situation it would be surprising if NOT different successful strains emerged in different locations and populations at the same time.
HOW COULD A PHENOTYPIC CHANGE TAKE PLACE?

We have discussed above what could be the phenotypic change behind Beijing strains’ large tendency to occur in clusters and to form big clusters. It is assumable that infectivity is a complex process depending upon the action of a number of genes. An increase in infectivity is thus not something that would come about by a single random mutation or two. The second hypothesis, that successful strains could have lost the ability to establish a stable latent infection, is more attractive. The metabolic shift to a dormant stage is probably also due to the function of several genes, but in this case a compromising mutation in an essential gene or deletion of genetic material involved in dormant metabolism would result in such a phenotype.

Mycobacteria are prone to loss of genetic material through deletions. BCG is itself a new phenotype of \textit{M. bovis} that lost virulence through three major deletions during repeat passages in vitro\textsuperscript{24}. \textit{M. leprae} became an obligate human parasite through successive deletions of nearly 50% of the usual mycobacterial genome size\textsuperscript{25}. Probably, the tendency of mycobacteria to contain numerous insertion elements indicates that they are highly susceptible to insertions and deletions. Little is known about the mechanism of this susceptibility, but mutator genes and defects in gene repair mechanisms are possible avenues to explore.

HOW COULD WE RESEARCH IF TB IN THE 21\textsuperscript{ST} CENTURY IS AN EMERGING OR RECURRENT EPIDEMIC?

This paper proposes several hypotheses, even some “tandem hypotheses”. Can they be tested? A key question is whether Beijing strains, as the most established example of successful strains currently, are more infective than non-Beijing strains. This question is addressed in an ongoing contact investigation where the source cases have spread strains with an identified genotype. The number of infected contacts is measured by tuberculin skin testing of contacts with a standard definition. The contact investigators are blinded regarding the identity of the infecting strains.

The alternative hypothesis, that the high clustering of Beijing strains is due to a loss of ability to establish stable latency, is not feasible to test in a human population and is addressed in experimental animals.

The second important challenge is to collect high quality epidemiological data and molecular strain patterns from as many relevant sites as possible. The current situation and dynamics of selective pressures are taken into account as well as performance indicators and coverage of the TB control program. Several countries in Asia, Africa, Europe and Latin America are joining efforts to achieve this goal. Presumably this data collection will also give us indications of successful strains other than the Beijing family. But only longitudinal data will give clear answers to this question.

The problem regarding protective efficacy of BCG against successful strains is both of extreme strategic importance and scientifically mind boggling. A clinical placebo controlled trial is out of question for medical ethics reasons. Age distribution of clinical isolates analyzed against changes in BCG coverage and individual vaccination histories could provide valuable indications. Such data exist already and more are to be collected. However, much stronger evidence would be available if we had a surrogate marker for future course shortly after the initial infection has taken place. Mouse experiments with models for slowly progressive and latent TB showed distinct differences in cytokine pattern from very early in infection\textsuperscript{26}. A similar distinction seems to predict clinical disease or not in adults exposed to \textit{M. leprae} in childhood. If this is the case also in people infected with \textit{Mtb}, testing contacts exposed to Beijing versus non-Beijing strains will provide the essential information.

Lastly, the genetic mechanisms behind new phenotypes need to be studied. This is going to be a tedious process, but is of fundamental importance in providing the type of understanding that we need to update or revise the current global TB control strategy.

WHAT MUST BE DONE INSTANTLY?

The main hypothesis this paper describes is not only an intellectual challenge. It is an existential question for TB control and thus for the fight against poverty that the Millennium Goals have committed us to. The TB situation is already serious and has been declared a “global emergency” by WHO. If there is an emerging TB pandemic, and this is due to \textit{Mtb} strains with a changed biology, the situation is both serious and urgent!

The actions to be taken can be defined at three levels:

Firstly, the classical factors fuelling a TB epidemic are at work and can be adequately addressed with our current knowledge. Poverty, particularly extreme poverty, has to be reduced through all the political, economical and humanitarian strategies we know. HIV control must be intensified. TB control programs must be strengthened to use drugs correctly and cure all diagnosed patients to prevent chronic excretors and drug resistance. They must also be decentralized to offer people access to the services. Health facilities must review their routines to minimize transmission of \textit{Mtb} between patients.

The second level is to run all the research activities indicated in this paper to be up-front in developing new tools and strategies adequate to meet the challenges of a new \textit{Mtb} phenotype, emerged through selection of some of the current tools and strategies we apply in TB control.
The third level refers to the underutilized possibilities in the present TB control strategy. From numerous studies we know that diagnostic delay, from TB patients’ first symptoms till they are put on adequate chemotherapy is long; mostly two to three months\textsuperscript{37}. A recent study has indicated that sputum smear grading is highest in patients with the shortest delay\textsuperscript{38}. The potential for reducing spread of \textit{Mtb} in the community seems thus very large – and mostly ignored in the present passive case finding strategy. By more detailed analysis of the delay period usually attributed to the patients, we found that patients with TB symptoms contacted an educated health worker on average after just 25% of the total delay period\textsuperscript{39}.

However, the symptoms, and particularly chronic cough with sputum, are so prevalent in most societies that the early contact with health services causes delay rather than suspicion of TB. If we could have a simple diagnostic tool to rapidly tell the health worker at the most peripheral post in the health system that this is a “true TB suspect” and send him instantly for sputum smear examination, we could theoretically reduce transmission in the community with 75%!

We are currently revisiting previous published and unpublished tests to see if we can find one with a very high sensitivity and feasible specificity (50% or better) that could be made durable, affordable and simple enough to be used at the very frontline of primary health.

**THE TB IN THE 21\textsuperscript{ST} CENTURY CONSORTIUM**

University of Oslo (Norway): Gunnar Bjune (scientific coordinator), Tone Tonjum, Olga Toungoussova, Birgitte Freiesleben de Blasio, Fekadu Abebe
Norwegian Institute of Public Health: Carol Holm-Hansen (administrative coordinator), Ulf Dahle, Turid Mannsåker, Dominique Caugant, Einar Heldal
Northern Medical University and Archangel Tuberculosis Control (Russia): Andrey Mariandyshiev, Olga Toungoussova
Epilab and Sudan National Tuberculosis Control (Sudan): Asma Elsomy
National Institute of Medical Sciences and Nutrition “Salvador Zubrin” (Mexico): Rogelio Emández-Pando
National Institute of Hygiene and Epidemiology and National Tuberculosis Hospital (Vietnam): Vu Tân Trao, Bui Duc Duong
National Institute of Public Health and Environment (The Netherlands): Dick van Soolingen
Stellenbosch University (Republic of South Africa): Nulda Beyers, Paul van Helden, Rob Warren, Ben Marais, Anneke Hesseling
Federal University of Rio de Janeiro (Brazil): Afranio Kritski
Kibong’oto National Tuberculosis Referral Hospital and Kilimanjaro Christian Medical College, Tumaini University, and Uhumbili University, Dur es Salaam (Tanzania): Leonard Uiso, Mrama Nyindo, Mecky Matee
University of Pennsylvania (USA): Daniel Malamud

**ACKNOWLEDGEMENTS**

This project is made possible by a grant from the Norwegian Research Council (project no: 164053/S30). It has also bee a great help for us that the University of Oslo, Medical Faculty has targeted global infections as one of their priority areas and established a structure to coordinate and support joint research in this area (Globinf). We have received several “seed money grants” from Globinf in preparation of this project and this has proved absolutely essential to bring the consortium together and get activities started. A special thanks to Dr. Fekadu Abebe who has helped with the references and Dr. Carol Holm-Hansen who has improved the language of this paper.

**REFERENCES**

8. Grzybowski S, Enarson D. Results in pulmonary tuberculosis patients under various treatment program conditions. \textit{Bull Int Union Tuberc} 1978; \textbf{53}: 70.